Is Marijuana a "Gateway" Drug Among Youth? Evidence From the National Longitudinal Survey of Youth 1997

Jin Xing*

September 2022

Abstract

The "gateway" hypothesis contends that marijuana use increases people's risk of progressing to use illicit hard drugs (e.g., cocaine, heroin). As some United States (U.S.) states have legalized recreational marijuana in recent years and many are considering decriminalizing marijuana, it is crucial to investigate whether such legislation will subsequently fuel hard drug use. Although a large body of medical and economic literature has examined the "gateway" hypothesis, existing studies tend not to differentiate between correlation and causation. This study contributes to the literature by using a bivariate survival panel model that controls for confounding variables. Using the National Longitudinal Survey of Youth 1997, I found strong evidence of "gateway" effects among youth in the U.S. Furthermore, the effects are more pronounced among those who first used marijuana before the age of 18, as well as those who used marijuana more frequently. Moreover, the effects are lower in African Americans and become less potent as people age. These results inform the current debate over the potential of marijuana use during adolescence to further hard drug involvement and highlight the importance of postponing the onset and reducing the frequency of marijuana use.

Keywords: *marijuana, hard drug, the "gateway" hypothesis, NLSY97, bivariate random-effects probit model*

JEL Classification: *C10, C30, C41, I10*

*Tulane University, Department of Economics, 6823 St. Charles Ave., New Orleans, LA 70118, USA. Email: jxing@tulane.edu.

1 Introduction

Marijuana is the most widely used federally illicit drug in the United States (U.S.).[1] Studies have consistently found that marijuana usage commonly precedes the use of hard drugs, such as cocaine and heroin (Kandel 1975). Hard drugs are substances that can seriously harm the user's health, whereas soft drugs, such as alcohol and marijuana, pose fewer health risks. This sequence of drug use can be explained by the "gateway" hypothesis, which posits that the experimentation with or use of marijuana increases the likelihood that an individual will progress to and engage in the use of hard drugs. The use of any hard drugs, even occasionally, can significantly increase the risk of adverse health effects, including abuse, dependence, addiction and overdose, which raises a number of public health concerns.

"Gateway" effects may arise via the following three channels. First, marijuana use may provide instant pleasure to users, causing them to crave more potent hard drugs in the hopes of experiencing heightened gratification. If this is the case, legalizing marijuana use, which is likely to increase marijuana usage, would increase the consumption of hard drugs. Second, interactions with drug dealers and peer users may increase the chance for marijuana users to interact with suppliers of hard drugs.[2] To remedy this, drug policies aimed at separating the markets for marijuana and hard drugs may be effective in reducing the number of individuals who progress to the use of hard drugs.[3] In addition, by legalizing marijuana, the government can oversee and track the sale of marijuana, thereby improving the transparency, safety, and quality of the product being sold. Furthermore, the legalization of marijuana would introduce competition to the marijuana market, thus deterring drug cartels that only exist in underground markets and possibly curbing hard drug use. Third, since marijuana is relatively safe to use, those who started using marijuana first may have formed a false perception that other illicit drugs are also

[1] According to a 2020 report by the National Institutes of Health (NIH), among people aged 12 years or older in the U.S. in 2020, 17.9% (or about 49.6 million people) reported using marijuana in the past 12 months (NIH 2020).

[2] Bretteville-Jensen and Jacobi (2011) reported that 23% and 19% of non-marijuana users claimed to be able to obtain heroin and cocaine respectively within three days, whereas the corresponding figures for marijuana users were 57% and 33% respectively. In addition, more marijuana users than non-users claimed that they had been offered hard drugs (80% versus 26%).

[3] Such policies exist in the Netherlands, where the sale of a limited quantity of marijuana is permitted in "coffee shops," attempting to keep marijuana experimenters and users away from hard drugs. Meanwhile, the sale of hard drugs is subject to severe punishment. "The idea is that strict prohibition of soft drugs would stimulate the black market and lead soft drugs users into hard drug use" (Van Ours 2003).

safe, leading them to experiment with hard drugs.[4] If so, creating additional and enhancing current health warnings relating to the adverse effects of hard drugs may help correct users' false perceptions and reduce hard drug usage.[5]

However, the observed escalation from marijuana to hard drug use does not necessarily imply that usage of marijuana causes hard drug use. Certain characteristics that could cause some individuals to be more susceptible to soft and hard drug usage include genetic predisposition, psychological problems (e.g., stress, depression, and childhood trauma), family background, peer environment, and accessibility.[6] Regardless of their history of marijuana use, individuals with these characteristics could start using marijuana because it is cheaper and more readily available than hard drugs (Pudney 2003). Pudney (2003) suggested that the relative costs may also be attributed to the observed sequence of drug use. Since hard drugs are more expensive than marijuana and people tend to have little resources early in life, they may only be able to afford marijuana at first and then progress to hard drugs as their earnings increase.

Ideally, randomized human trials should be conducted to investigate the "gateway" hypothesis, but such trials have been deemed to be unethical and illegal worldwide. The existing medical and economic literature on the "gateway" hypothesis relies mainly on observational data (e.g., self-reported survey data) and researchers disagree on whether marijuana is a "gateway" drug.[7] Currently, the mechanism underlying the observed sequence of drug use is not well understood, with many studies not distinguishing correlation from causation. This limitation leads to results that can be explained by confounding unobserved heterogeneity (e.g., the propensity to problem behavior, genetic predisposition). However, despite limited empirical evidence, opponents of the legalization of marijuana often cite the "gateway" hypothesis in their argument. As the U.S. is considering decriminalizing marijuana at the federal level, which we submit would increase marijuana consumption, it is of great significance that the validity of the "gateway" hypothesis is investigated in order to predict whether such legislation will subsequently lead to the increased use of hard drugs.[8]

[4] "Light-to-moderate marijuana consumption generates few adverse health effects" (Sabia et al. 2021).

[5] Heroin and cocaine score higher than marijuana in terms of physical harm, dependence, and social harm in a scoring proposed by Nutt et al. (2007).

[6] Bretteville-Jensen and Jacobi (2011) refer to accessibility as economic, cultural, and physical factors that affect an individual's access to drugs, such as monetary and non-monetary costs (e.g., transaction costs, social stigma).

[7] See Nöel and Wang (2018) for a review of this literature.

[8] At the state level, as of May 2022, medical marijuana is legal in 38 states and D.C., and recreational marijuana is legal in 19 states and D.C.

This study aims to disentangle the "gateway" effects from the results of unobserved heterogeneity using data collected in the National Longitudinal Survey of Youth 1997 (NLSY97). The research question is as follows: does marijuana use hastens the onset of hard drug use among adolescents and young adults.[9] While studies that model the intertemporal dependence between marijuana and hard drug use (e.g., DeSimone 1998; Deza 2015) answer how current marijuana use affects future hard drug use, the results of this study interpret how marijuana use affects the timing of hard drug initiation, which is more in line with the conventional notion of the "gateway" effect. In addition, previous studies with a similar interpretation to ours (e.g., Beenstock and Rahav 2002; Van Ours 2003; Pudney 2003; Melberg, Jones and Bretteville-Jensen 2010; Bretteville-Jensen and Jacobi 2011) all used cross-sectional data. Our method of analysis is different as we apply a bivariate survival panel model to the longitudinal survey data, which allows for time-varying observables, while also accounting for confounding fixed and time-varying unobservables. Moreover, the proposed model is a more flexible method for heterogeneity analysis, which informs about the underlying mechanisms that cause the "gateway" effects.

2 Literature Review

Random animal trials generally support the "gateway" hypothesis (e.g., Pistis et al. 2004; Ellgren, Spano and Hurd 2007; Panlilio et al. 2013). Results from animal studies have revealed potential biological mechanisms that give rise to the "gateway" effect. For example, studies have shown that exposing rats to THC (a component of marijuana that is responsible for most of the psychological effects of the drug) during a critical phase of brain development (e.g., adolescence) is associated with an altered brain reward system that increases the likelihood that the rats will self-administer other drugs, such as heroin (NIH 2020). To the extent that the findings of these studies are applicable to humans, this could help to explain why individuals who began marijuana use in youth were found to be more susceptible to hard drug usage than those who began later in life.

Another aspect of research on "gateway" effects examines how marijuana laws, either medical or recreational, affect hard drug use (e.g., Wen, Hockenberry and Cummings 2014; Chu 2015; DeAngelo and Redford 2015; Sabia et al. 2021). Studies related to this component generally found that laws regarding the regulation of marijuana usage in-

[9] Here, young adults refer to individuals who are between 18 and 35 years old.

creased marijuana use, but had no impact on hard drug use, thus providing no support for the "gateway" hypothesis. However, as Shover et al. (2019) cautioned one must refrain from drawing a causal connection at the individual-level from population-level data.[10] Moreover, since marijuana laws may only affect a small subpopulation, insignificant estimates based on aggregate data may mask the actual effect of policies in the affected subpopulation. In contrast, longitudinal survey data have the advantage of following a sample of individuals over a period of time, observing their demographics along with the evolution of their self-reported drug use behavior and other variables. In the circumstances, longitudinal survey data facilitates the analysis of heterogeneous effects among subgroups.

This study is similar to those that rely on econometric techniques to distinguish the "gateway" effects from observed heterogeneity using individual-level survey data. This class of research has generally found "gateway" effects to be statistically significant and has placed importance on controlling for unobservables (e.g., DeSimone 1998; Pudney 2003; Van Ours 2003; Melberg, Jones and Bretteville-Jensen 2010; Bretteville-Jensen and Jacobi 2011; Deza 2015), except for Beenstock and Rahav (2002). Using the NLSY from the U.S., DeSimone (1998) estimated the effect of past marijuana use on current cocaine use with instrumental variable (IV) estimation, and concluded that there was strong evidence for the "gateway" hypothesis.[11] Van Ours (2003) made use of a bivariate survival model to investigate the dynamics of marijuana and cocaine use in the Netherlands. He found limited, but significant "gateway" effects from first-time marijuana to first-time cocaine use, as well as that some individuals initiated marijuana use before cocaine use, and this was mostly driven by unobserved heterogeneity. Pudney (2003) modeled the interdependency between multiple types of drug use in the UK and after considering unobserved heterogeneity, he found that the "gateway" effects from soft to hard drugs appeared to be modest, but statistically significant. Melberg, Jones and Bretteville-Jensen (2010) used a bivariate survival model with shared frailty (i.e., unobserved characteristics) to examine how past marijuana initiation affected hard drug initiation in Norway.

[10] An earlier study (Bachhuber et al. 2014) found that from 1999 to 2010, states with medical marijuana laws saw slower rates of increase in opioid overdose death rates than other states. However, extending Bachhuber et al.'s analysis through 2017 using the same method, Shover et al. (2019) show that the findings not only do not hold for the extended period, but the relationship reversed in sign, from −21% to 23%. From this, the authors deduced that the earlier found relationship were spurious, likely due to unobserved factors to researchers.

[11] The author used two measures of state-level penalties for marijuana possession (i.e., maximum jail time and an indicator that no fines for first-time offenders) and two alcohol-related variables (i.e., beer tax and an indicator of parental alcohol problems) as instrumental variables.

Utilizing a latent class approach (which assumes that each individual belongs to one of the subgroups or latent classes in the population), they identified two distinct groups of youth: a smaller group of youths who manifested "gateway" effects and a larger group of youths who demonstrated no "gateway" effects. Bretteville-Jensen and Jacobi (2011) employed a bivariate probit model and Bayesian estimation framework to analyze the effects and relative importance of previous marijuana use, proneness to drug use, and accessibility factors on hard drug initiation using 2006 Norwegian survey data. They found that these factors contributed to the observed higher hard drug use among marijuana users, with previous marijuana use having the largest effect. Deza (2015) modeled the intertemporal dependence of drug use (within and between drugs) using a dynamic discrete choice model of alcohol, marijuana, and hard drug use. She found statistically significant "stepping-stone" effects from soft to hard drugs and concluded that alcohol, marijuana, and hard drugs were complements in utility.[12] However, Beenstock and Rahav (2002), who applied various methods (e.g., two-stage logit, bivariate probit, survival analysis) to survey data from Israel, found "gateway" effects from cigarettes to marijuana, but not from marijuana to hard drugs.[13]

3 Methodology

In this section, I constructed an empirical model that relates hard drug initiation to marijuana use, while accounting for unobserved heterogeneity to assess "gateway" effects.

For individual i in year t, D_{it} denotes an indicator of hard drug initiation, M_{it} denotes an indicator of marijuana use.[14] I first specified a probit model associating latent hard drug initiation to marijuana use, as follows:

$$D_{it}^{*} = \alpha \cdot M_{it} + X_{it}'\beta + T_t + \eta_i^1 + \upsilon_{it}^1 \tag{1}$$

where D_{it}^{*} is latent hard drug initiation, that is, $D_{it} = \mathbb{1}[D_{it}^{*} > 0]$, which measures the

[12] The author referred to the "stepping stone" effect from drug i to drug j as the causal effect of the use of drug i in period t on the use of drug j in period $t+1$. Also, she referred to complements in utility in the sense that utility derived from consuming drug i and j together was higher than the summation of the utilities derived from consuming either drug alone.

[13] The authors acknowledged that their approach was "somewhat *ad hoc*" due to the absence of price data for marijuana and hard drugs and that their sample size may have insufficient power to detect "gateway" effects under their estimations.

[14] An indicator of an event equals 1 if the event occurs and 0 otherwise.

propensity to initiate hard drugs.[15] The coefficient of interest, α, determines the "gateway" effects from marijuana to hard drugs. The sign of α indicates whether marijuana and hard drugs are complements ($\alpha > 0$), substitutes ($\alpha < 0$), or bear no relationship ($\alpha = 0$). X_{it} is a vector of controls that consists of pre-determined and time-varying observables (variables that can be measured). β is a vector of the coefficients. T_t is the year fixed effect. η_i^1 represents fixed unobservables (e.g., the propensity to use hard drugs) and υ_{it}^1, time-varying unobservables (e.g., hard drug prices).

Since marijuana use (M_{it}) could have a mutual relationship with hard drug initiation via unobservables (η_i^1, υ_{it}^1) separately from any causal link, the estimated "gateway" effect ($\hat{\alpha}$) could be biased if Equation (1) was estimated in isolation. To address this endogeneity issue, I specified an auxiliary probit model for latent marijuana use, M_{it}^*, as follows:

$$M_{it}^* = X_{it}'\gamma + T_t + \eta_i^2 + \upsilon_{it}^2 \tag{2}$$

where the variables and coefficients are defined in a manner similar to Equation (1). To control for endogeneity in Equation (1), I allowed η_i^1 and η_i^2 to be jointly normally distributed with mean 0 and covariance matrix Ω_η,

$$\begin{pmatrix} \eta_i^1 \\ \eta_i^2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega_\eta = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 \cdot \rho_\eta \\ \sigma_1\sigma_2 \cdot \rho_\eta & \sigma_2^2 \end{pmatrix} \right) \tag{3}$$

where σ_1 and σ_2 are the standard deviations of η_i^1 and η_i^2. ρ_η denotes the correlation between η_i^1 and η_i^2. Additionally, I allowed υ_{it}^1 and υ_{it}^2 to be jointly normally distributed with a mean of 0 and covariance matrix Ω_υ,

$$\begin{pmatrix} \upsilon_{it}^1 \\ \upsilon_{it}^2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega_\upsilon = \begin{pmatrix} 1 & \rho_\upsilon \\ \rho_\upsilon & 1 \end{pmatrix} \right) \tag{4}$$

where υ_{it}^1 and υ_{it}^2 have a standard deviation of 1. ρ_υ is the correlation between υ_{it}^1 and υ_{it}^2. Collectively, Equations (1)–(4) are known as a bivariate random-effects probit model.

[15] $\mathbb{1}[\cdot]$ denotes the indicator function.

4 Data and Variables

4.1 Sample Selection

The analyses were based on the NLSY97, an individual-level panel survey where information was collected from a nationally representative sample of 8,984 individuals born between 1980 and 1984 in the U.S. Surveys were conducted annually from 1997 to 2011 and biennially from 2011 onwards. NLSY97 started collecting information relating to individual's use of marijuana in 1997 and in respect of hard drug use in 1998.[16]

This study ran from 1999 to 2015, which was the last year that the NLSY97 asked questions on marijuana and hard drug use.[17] To fully investigate the self-reported drug use history of the participants, I selected a sample of 5,821 individuals who had never reported using marijuana or hard drugs before 1999.[18] The key and control variables along with the corresponding NLSY97 questions are listed in Table 1. Table 2 provides the summary statistics for the individuals in the sample. Note that, in Table 2, the average starting age for marijuana usage (19.68 years) was lower than that for hard drugs (20.66 years).

Following Rabe-Hesketh and Skrondal (2008) and Melberg, Jones and Bretteville-Jensen (2010), I constructed an indicator of hard drug initiation, M_{it}, which is used in estimations as the dependent variable. For those who did not initiate hard drugs during the study period, M_{it} was equal to zero throughout the study period. In contrast, for those who started using hard drugs during the study period, M_{it} equals 1 for the year of initiation and 0 before that year. After the first instance of hard drug use, the individuals were no longer at risk of initiating hard drugs and were thus removed from the sample for subsequent years. As a result, the analysis panel data consisted of 61,799 observations at the individual and year levels.[19]

[16] For example, NLSY97 started asking, "R ever use marijuana?" in 1997, "# days use marijuana in last 30 days?" in 1997, "Has R used marijuana since DLI?" in 1998, "R ever use cocaine/hard drugs?" in 1998, and "Has R used cocaine/hard drugs since DLI?" in 1999. Here, R stands for respondent and DLI stands for date of last interview.

[17] The study period begins in 1999 as it was the first year the NLSY97 posed the question, "Has R used cocaine/hard drugs since DLI?", which is used to derive the dependent variable (hard drug initiation) in later analyses.

[18] The sample selection is based on three NLSY97 questions: "R ever use marijuana?" in 1997, "Has R used marijuana since DLI?" in 1998, and "R ever use cocaine/hard drugs?" in 1998.

[19] The NLSY97 indicated missing data with five negative values: (−1) refusal, (−2) don't know, (−3) invalid skip, (−4) valid skip, and (−5) noninterview. This study treats data with negative values as missing for all variables except for marital status. For marital status, teenagers with a value of −4 (valid skip) are assumed to be never married.

4.2 Descriptive Survival Analysis

Malone et al. (2010) described the "gateway" hypothesis in necessary conditions regarding risks, that is, the conditional probability of an event, given that the event has not yet occurred.[20] A key condition posits that compared to individuals who have never used marijuana and hard drugs, individuals who have used marijuana are more at risk of starting to use hard drugs.

To examine whether this condition is consistent with the empirical hazards of hard drug initiation, following the method of Rabe-Hesketh and Skrondal (2008), I assessed the discrete-time hazards as predicted probabilities by estimating a logistic regression:

$$\text{logit}[Pr(D_{it} = 1|d)] = \sum_{s=0}^{T_i - T_i^0} \alpha_s \cdot d_s \tag{5}$$

where D_{it} is an indicator of hard drug initiation in year t for individual i, which was constructed in the previous section. For those who started using hard drugs during the study period, T_i is the year of initiation. For those who did not initiate hard drugs during the study period, T_i was the last year before being right-censored.[21] T_i^0 is the first year at risk of initiating hard drugs. s indexes the number of years at risk in year t, that is, $s = t - T_i^0$. d is a vector containing all dummy variables d_s, where s ranges from 0 to $T_i - T_i^0$. α's are coefficients.

Figure 1 shows the estimated hazards of hard drug initiation against years at risk of initiating hard drugs versus those conditional on having used marijuana.[22] In this figure, the x-axis represents the number of years since the first year at risk and the y-axis represents the hazard value.[23] Except for one year at risk, the risks for individuals who had used marijuana before were higher than the risks for all individuals. This observation raises the question: was this pattern driven by the causal effect of marijuana use on increasing the risk of initiating hard drugs or other factors? The main objective of this study is to determine the extent to which this pattern is due to "gateway" effects.

[20] Mathematically, hazard at year t can be expressed as $Pr[T = t|T \geq t]$, where $Pr[\cdot]$ is the probability function, and T is the year that the event under investigation occurs.

[21] Right-censoring occurred in either of the following two scenarios. First, hard drug initiation did not occur before the end of the study period. Second, the subject ceased to be at risk of initiating hard drugs before the end of the study period due to dropping out or death.

[22] In estimating the hazards of hard drug initiation conditional on having used marijuana, I drop data before the year of initiating marijuana for those who used marijuana.

[23] For those who never used marijuana and hard drugs, the first year at risk of initiating hard drugs is set at 1999, the first year of the study period. For those who had used marijuana, the first year at risk is set as the year they began using marijuana.

5 Results

5.1 Baseline Estimates

In this section, as our primary interest is the "gateway" effect from marijuana to hard drugs, only the estimates for the model of hard drug initiation, i.e., Equation (1) in Section 3, are presented and discussed.[24]

Table 3 shows the estimates from various specifications in which standard errors are clustered at the individual-level.[25] In columns (1)–(3), controlling for pre-determined controls (listed in Table 1), I estimated Equation (1) in Section 3 alone with random-effects logit, random-effects complementary log-log, and random-effects probit, respectively. The semi-elasticity of hazards of hard drug initiation with respect to marijuana use was estimated at 2.69, 2.67, and 2.84, respectively.[26] Thus, the semi-elasticity estimate is robust across different model types. In column (4), I jointly estimated Equations (1) and (2) in Section 3, allowing the correlation of time-varying but not fixed unobservables in both equations. In column (5), I further allowed fixed unobservables in both equations to correlate. The column (5) estimate (0.49) was statistically significant and roughly half of the column (4) estimate (0.96), suggesting that fixed unobservables partly explain why marijuana use tends to precede hard drug use. Based on column (5), I excluded the pre-determined controls in column (6). The column (6) estimates were similar to the column (5) estimates, but only marginally significant, suggesting that the bivariate random-effect model is fairly robust with different sets of pre-determined controls. Nevertheless, controlling for pre-determined controls is preferred as it is more efficient. Finally, based on column (5), column (7) includes time-varying controls (listed in Table 1). Column (7) estimates were similar to those in column (5) and were statistically significant, suggesting that the bivariate random-effect model is robust with different sets of time-varying controls as well. However, missing values in time-varying controls reduce the size of the analysis sample. In addition, certain time-varying controls (e.g., employment status, marital status, educational attainment, and health status) could be impacted by marijuana use and thus serve as channels by which marijuana use affects hard drug initiation. Therefore,

[24] The full set of estimates is available upon request.

[25] Estimations are implemented in Stata. Estimations in columns (1)–(4) use Stata command xtlogit, xtcloglog, xtprobit, and biprobit, respectively. Estimations in columns (5)–(7) use Stata module CMP (Roodman 2022).

[26] The semi-elasticity of y with respect to x is evaluated as the average of $\frac{\partial \hat{y}}{\partial x} \cdot \frac{1}{\hat{y}}$, where $\hat{y}$ is the expected value of y.

it is preferable not to control for time-varying controls. Accordingly, the specification in column (5) is preferred and is hereafter referred to as the preferred specification.

Taken together, the unobserved heterogeneity is partly attributable to the observed sequence of drug use from marijuana to hard drugs. However, after accounting for unobserved heterogeneity, the estimate of the "gateway" effect from marijuana to hard drugs is still statistically significant and robust with various sets of controls.[27] The semi-elasticity estimate in the preferred specification indicates that marijuana use increases the risks of starting to use hard drugs by 1.39 times, which is comparable to the estimate indicated by Melberg, Jones and Bretteville-Jensen (2010), who found that the risk of hard drug initiation more than doubled after marijuana initiation.

5.2 Heterogeneous Effects

Marijuana use does not affect all individuals equally, offering scope for heterogeneity among subgroups.[28]

To examine the heterogeneous effects, I included interaction terms in addition to the preferred specification. The results can be viewed in Table 4. In column (1), the estimate of the interaction between marijuana use and the indicator of being female is not statistically significant, suggesting that the "gateway" effects are similar whether the user is male or female. In column (2), while African Americans have significantly lower "gateway" effects compared to the reference group (non-Black/non-Hispanic), the "gateway" effects are similar across Hispanics, mixed race (non-Hispanic), and non-Black/non-Hispanic. Previous studies conducted (e.g., Bretteville-Jensen and Jacobi 2011) have suggested that individuals who started using marijuana during the development of their brain may be more vulnerable to the influence of marijuana. In column (3), to examine whether the "gateway" effects are more pronounced among people who started using marijuana early in life, I included the interaction between marijuana use and an indicator of "early marijuana users," defined as those who first used marijuana before the age of 18 years old.

[27] On the basis of the preferred specification, I conducted two robustness checks as follows: first, the semi-elasticity estimate remains statistically significant when standard errors are clustered at the region level (instead of the individual level). Second, the NLSY97 oversamples African Americans and Hispanics and provides a nationally representative subsample. After I removed individuals not in the nationally representative subsample from the analysis sample, with the remaining 45,016 observations, the estimate was 1.73 and statistically significant, which is comparable to the semi-elasticity estimate in the preferred specification (1.39).

[28] "Pleasant experiences with marijuana are by no means universal. Instead of relaxation and euphoria, some people experience anxiety, fear, distrust, or panic" (NIH 2020).

Consistent with previous literature, this study indicated that "early marijuana users" have greater "gateway" effects. To examine whether the "gateway" effects are more pronounced among people who used marijuana more frequently, column (4) includes the interaction between marijuana use and an indicator of "intense marijuana users," defined as those who used marijuana for more than three days in the last 30 days at the time of the survey.[29] I found that "intense marijuana users" have greater "gateway" effects. Finally, column (5) estimates suggest that "gateway" effects decline with age, that is, earlier marijuana use has a greater impact on hard drug initiation, which is consistent with the findings of Deza (2015). Figure 2 shows semi-elasticity estimates in different groups.

Although the onset age of and frequency of marijuana use may be endogenous (driven by unobserved factors that induce the use of hard drugs), the heterogeneity effects are informative in suggesting that early onset and frequency of marijuana use may be, at least, partly to blame for the rise of "gateway" effects.

6 Limitations

This study has several limitations. First, self-reported drug use may not accurately measure actual drug use as users may misreport and have difficulty remembering. Although this may be concerning with all survey data, it could be more so in this context. NLSY97 interviews were conducted either face-to-face or via telephone. Despite measures to guarantee confidentiality, NLSY participants may have underreported their drug use due to social stigma and the illegality of illicit drug use (DeSimone 1998). The participants were equally likely to have exaggerated their actual drug use (Bretteville-Jensen and Jacobi 2011).

Second, I assumed that the bivariate random-effects probit model accurately describes how marijuana use and hard drug initiation are determined, in which the error terms are bivariate normally distributed. However, if the bivariate normal distributions did not sufficiently account for unobserved heterogeneity, then the estimate could be biased and driven by certain omitted variables. For example, drug prices could potentially be important omitted variables, as previous studies have suggested that prices were important in explaining drug use (e.g., Van Ours and Williams 2007; Melberg, Jones and Bretteville-Jensen 2010).

[29] The indicator of "intense marijuana users" is derived from the NLSY97 question "# days use marijuana in last 30 days?"

Moreover, the analyses in this study assumes that some participants who never used hard drugs stopped partaking in the survey before the end of the study period (i.e., right-censoring) had nothing to do with potential hard drug initiation (i.e., the event). Namely, right-censoring was not informative about the event, given the covariates. If this was not the case, then the estimates would be biased.[30]

Finally, the nature of unobserved heterogeneity, which is found to partly account for the observed drug use sequence from marijuana to hard drugs, is unknown. In addition, this study does not discuss the mechanisms underlying the "gateway" effects, except that the age of onset and frequency of marijuana use may play a role. Understanding these mechanisms is crucial for policy purposes, as different mechanisms have different or even opposite policy implications. For example, as was noted in the Introduction section, if "gateway" effects arise mainly because marijuana "primes" the brains of users, then legalizing marijuana would stimulate hard drug use. However, if "gateway" effects arise mainly through interactions with drug dealers, then legalizing marijuana, which may aid in separating marijuana users from the hard drug market, may prove effective in curbing hard drug use.

7 Conclusion

The use of hard drugs, such as cocaine and heroin, imposes grave health risks to users, thus raising major public health concerns. Empirical studies have consistently found that marijuana use precedes hard drug usage. These findings led to the "gateway" hypothesis, which posits that marijuana use causes people to progress to experiment with and use hard drugs. As federal and state governments are considering legalizing marijuana, which may stimulate marijuana use, it is important to investigate the validity of the "gateway" hypothesis to inform the debate on legalizing marijuana.

The issue addressed in this study is whether marijuana use hastens hard drug initiation, consistent with the conventional notion of "gateway" effects. The major challenge is how to disentangle "gateway" effects from unobserved heterogeneity, which may also lead to the progression from marijuana to hard drugs, thus confounding correlation for causation. This study makes a methodological contribution to the current literature by proposing a bivariate survival panel model to isolate "gateway" effects from unobserved

[30] Among the 5,821 individuals in my analysis sample, 946 (16.3%) initiated hard drugs during the study period, and the rest were right-censored; among the 4,875 right-censored individuals, 1,074 (22.0%) discontinued the survey before the end of the study period.

heterogeneity. The same framework can be applied to examine broader issues, such as how drug use impacts educational attainment, employment status, or crime.

In the analysis sample taken from the NLSY97, I found that unobserved heterogeneity partly explains the sequence of drug use from marijuana to hard drugs. After accounting for unobserved heterogeneity, there were significant "gateway" effects from marijuana to hard drugs among the youth. With the preferred specification, marijuana use significantly increased the risk of hard drug initiation by 1.39 times. The estimate is qualitatively robust with various sets of covariates. Furthermore, "gateway" effects are more pronounced among those who first used marijuana before the age of 18 years, as well as those who used marijuana more frequently. Additionally, the effects were less pronounced in African Americans and were attenuated with age. These findings suggest that the age of onset and frequency of marijuana use may underpin the "gateway" effects. Consequently, it may be effective in curbing hard drug use through policy measures such as (1) launching information campaigns on marijuana's adverse effects, (2) early identification of youths experiencing drug problems and providing help to prevent them from starting to use hard drugs, and (3) increasing the legal age of marijuana use and limiting the potency of marijuana in marijuana-legalized states.

However, this study may suffer from identification problems (e.g., measurement errors and omitted variable bias) and does not speak to the mechanisms underlying "gateway" effects. As different mechanisms yield different, or even opposite, policy implications, future research is needed to better understand them.

References

Bachhuber, Marcus A, Brendan Saloner, Chinazo O Cunningham, and Colleen L Barry. 2014. "Medical cannabis laws and opioid analgesic overdose mortality in the United States, 1999-2010." *JAMA internal medicine*, 174(10): 1668–1673.

Beenstock, Michael, and Giora Rahav. 2002. "Testing Gateway Theory: do cigarette prices affect illicit drug use?" *Journal of Health Economics*, 21(4): 679–698.

Bretteville-Jensen, Anne Line, and Liana Jacobi. 2011. "Climbing the drug staircase: A bayesian analysis of the initiation of hard drug use." *Journal of Applied Econometrics*, 26(7): 1157–1186.

Chu, Yu-Wei Luke. 2015. "Do medical marijuana laws increase hard-drug use?" *The Journal of Law and Economics*, 58(2): 481–517.

DeAngelo, Gregory, and Audrey Redford. 2015. "Is Medical Marijuana a Gateway Drug?: The Effect of Medical Marijuana Legalization on Heroin Use Rates."

DeSimone, Jeffrey. 1998. "Is marijuana a gateway drug?" *Eastern Economic Journal*, 24(2): 149–164.

Deza, Monica. 2015. "Is there a stepping stone effect in drug use? Separating state dependence from unobserved heterogeneity within and between illicit drugs." *Journal of Econometrics*, 184(1): 193–207.

Ellgren, M, Maria Sabrina Spano, and YL Hurd. 2007. "Adolescent cannabis exposure alters opiate reward and opioid limbic neuronal populations in adult rats."

Kandel, Denise. 1975. "Stages in adolescent involvement in drug use." *Science*, 190(4217): 912–914.

Malone, Patrick S, Dorian A Lamis, Katherine E Masyn, and Thomas F Northrup. 2010. "A dual-process discrete-time survival analysis model: Application to the gateway drug hypothesis." *Multivariate behavioral research*, 45(5): 790–805.

Melberg, Hans Olav, Andrew M Jones, and Anne Line Bretteville-Jensen. 2010. "Is cannabis a gateway to hard drugs?" *Empirical Economics*, 38(3): 583–603.

NIH. 2020. "Cannabis (Marijuana) Research Report." https://nida.nih.gov/download/1380/cannabis-marijuana-research-report.pdf?v=7fc7d24c3dc120a03cf26348876bc1e4.

Nöel, Wm, and Judy Wang. 2018. "Is Cannabis a gateway drug?: Key findings and literature review." Library of Congress.

Nutt, David, Leslie A King, William Saulsbury, and Colin Blakemore. 2007. "Development of a rational scale to assess the harm of drugs of potential misuse." *the Lancet*, 369(9566): 1047–1053.

Panlilio, Leigh V, Claudio Zanettini, Chanel Barnes, Marcelo Solinas, and Steven R Goldberg. 2013. "Prior exposure to THC increases the addictive effects of nicotine in rats." *Neuropsychopharmacology*, 38(7): 1198–1208.

Pistis, Marco, Simona Perra, Giuliano Pillolla, Miriam Melis, Anna Lisa Muntoni, and Gian Luigi Gessa. 2004. "Adolescent exposure to cannabinoids induces long-lasting changes in the response to drugs of abuse of rat midbrain dopamine neurons." *Biological psychiatry*, 56(2): 86–94.

Pudney, Stephen. 2003. "The road to ruin? Sequences of initiation to drugs and crime in Britain." *The Economic Journal*, 113(486): C182–C198.

Rabe-Hesketh, Sophia, and Anders Skrondal. 2008. *Multilevel and longitudinal modeling using Stata*. STATA press.

Roodman, David. 2022. "CMP: Stata module to implement conditional (recursive) mixed process estimator."

Sabia, Joseph J, Dhaval M Dave, Fawaz Alotaibi, and Daniel I Rees. 2021. "Is recreational marijuana a gateway to harder drug use and crime?" National Bureau of Economic Research.

Shover, Chelsea L, Corey S Davis, Sanford C Gordon, and Keith Humphreys. 2019. "Association between medical cannabis laws and opioid overdose mortality has reversed over time." *Proceedings of the National Academy of Sciences*, 116(26): 12624–12626.

Van Ours, Jan C. 2003. "Is cannabis a stepping-stone for cocaine?" *Journal of health economics*, 22(4): 539–554.

Van Ours, Jan C, and Jenny Williams. 2007. "Cannabis prices and dynamics of cannabis use." *Journal of health economics*, 26(3): 578–596.

Wen, Hefei, Jason Hockenberry, and Janet R Cummings. 2014. "The effect of medical marijuana laws on marijuana, alcohol, and hard drug use." National Bureau of Economic Research.

Figures and Tables

Figure 1: Discrete-Time Hazards of Hard Drug Initiation

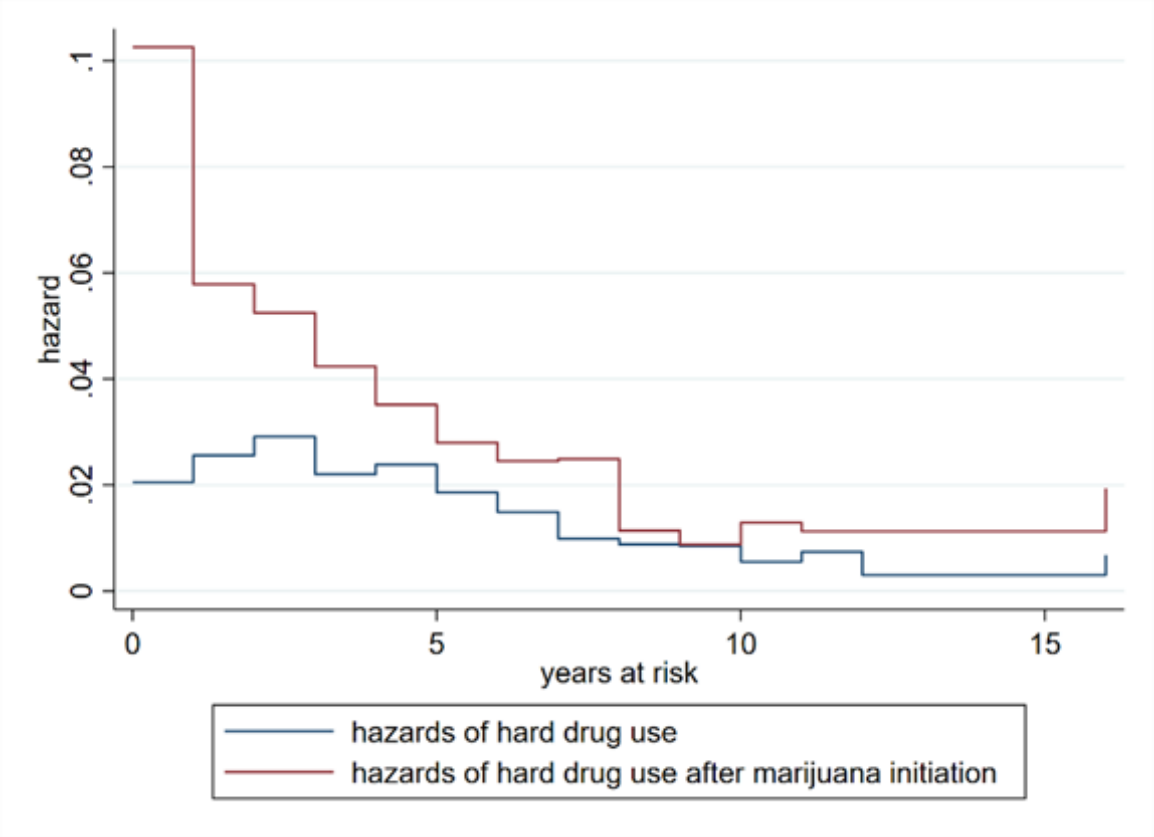
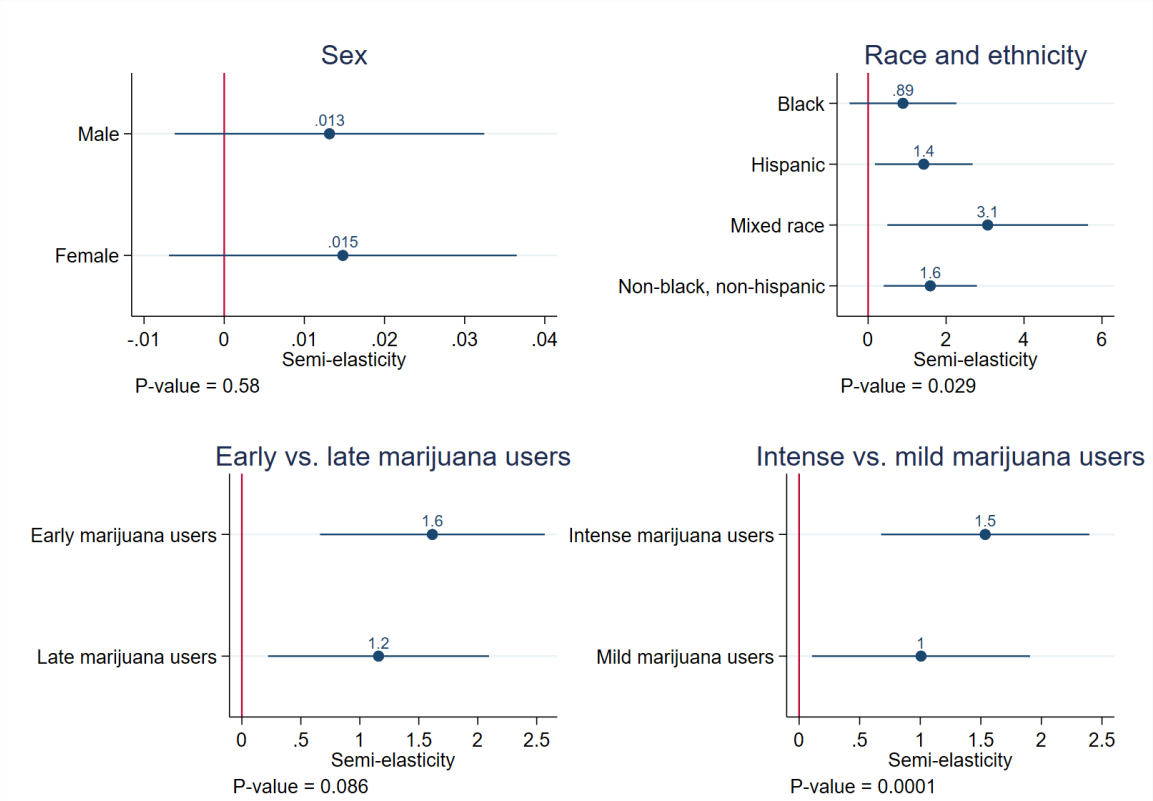


Figure 2: Semi-elasticity Estimates in Different Groups



Notes: Dots show point estimates and horizontal bars show 95% confidence intervals using standard errors clustered at the individual level. In each subplot, the p-value indicates whether the estimates are equal.

Table 1: Key and Control Variables Along With Their Corresponding NLSY97 Questions

Key variables	NLSY97 question
Marijuana use	Has R used marijuana since DLI?
Hard drugs use	Has R used cocaine/hard drugs since DLI?
Pre-determined controls	
Female	R's gender
Race and ethnicity	Race and ethnicity
Black	
Hispanic	
Mixed race (non-Hispanic)	
Ever arrested (in 1997)	Total number of arrests
Ever drank alcohol (in 1997)	R ever drink alcohol?
Ever smoked (in 1997)	R ever smoked?
% Peers using illicit drugs (in 1997)	Percent peers use illegal drugs
Lived with both biological parents (in 1997)	Does R live with both biological parents?
Religious (in 1997)	What is R's current religious preference?
Time-varying controls	
Receive income from job	R receive income from job in past year?
Marital status	R's marital/cohabit status
Single	
Married	
Household size	Household size
Educational attainment	Highest degree received
Less than high school	
High school	
Some college	
Rural	Current residence in urban or rural area
General health	How is R's general health?
Region of residence	Census region of residence
Northeast	
North Central	
South	
Drank alcohol	Has R drank since DLI?
Smoked cigarettes	Has R smoked since DLI?

Notes: R stands for respondent; DLI stands for date of last interview. "% Peers using illicit drugs" is a categorical variable with five categories: (1) less than 10%, (2) about 25%, (3) about 50%, (4) about 75%, and (5) more than 90%. "General health" is a categorical variable with five categories: (1) excellent, (2) very good, (3) good, (4) fair, and (5) poor. "Household size" is the number of family members living together. The other variables are indicators of the indicated event or status.

Table 2: Summary Statistics

	mean	SD
Birth year	1982.22	1.38
% Female	0.51	0.50
% Black	0.28	0.45
% Hispanic	0.22	0.41
% Mixed race (non-Hispanic)	0.01	0.09
% Non-black / non-Hispanic	0.50	0.50
% Arrested (before 1997)	0.26	0.44
% Drank alcohol (before 1997)	0.28	0.45
% Smoked cigarettes (before 1997)	0.23	0.42
% Lived with both biological parents (in 1997)	0.53	0.50
% Peers using illicit drugs (in 1997)	1.97	1.16
% Religious (in 1997)	0.90	0.30
% Used hard drugs (during study period)	0.16	0.37
% Used marijuana (during study period)	0.45	0.50
Age of first hard drug use (in those who used hard drugs)	20.66	3.87
Age of first marijuana use (in those who used marijuana)	19.68	3.73
N	5,821	

Notes: SD represents standard deviation. "% Peers using illicit drugs" is a categorical variable with five categories: (1) less than 10%, (2) about 25%, (3) about 50%, (4) about 75%, and (5) more than 90%.

Table 3: Baseline Estimates

	Hazards of hard drug initiation						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Model	logit	cloglog	probit	probit	probit	probit	probit
Marijuana use	2.76***	2.71***	1.17***	0.96	0.49**	0.45*	0.54***
	(0.082)	(0.078)	(0.047)	(1.03)	(0.29)	(0.25)	(0.17)
$\hat{\rho}_{\upsilon}$				0.11	0.24	0.27	0.19
$\hat{\sigma}_1$					0.55	0.71	0.16
$\hat{\sigma}_2$					1.27	1.39	1.04
$\hat{\rho}_{\eta}$					0.77	0.78	1.00
Semi-elasticity	2.69	2.67	2.84	2.40	1.39	1.29	1.48
Pre-determined controls?	Yes	Yes	Yes	Yes	Yes	No	Yes
Time-varying controls?	No	No	No	No	No	No	Yes
N	61,799	61,799	61,799	61,799	61,799	61,799	57,221

Notes: Pre-determined and time-varying controls are listed in Table 1. The estimates of the covariates are omitted for clarity. Standard errors are clustered at the individual level and reported in parentheses. $^{*}p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$.

Table 4: Heterogeneous Effects

	Hazards of hard drug initiation				
	(1)	(2)	(3)	(4)	(5)
Marijuana use × female	0.04				
	(0.067)				
Marijuana use × Black		-0.29***			
		(0.083)			
Marijuana use × Hispanic		-0.062			
		(0.084)			
Marijuana use × mixed race		0.56			
		(0.41)			
Marijuana use × early user			0.11**		
			(0.052)		
Marijuana use × intense user				0.20***	
				(0.050)	
Marijuana use × age					-0.018**
					(0.0079)
Marijuana use	0.51*	0.59**	0.41**	0.35**	0.89***
	(0.26)	(0.24)	(0.18)	(0.17)	(0.28)
Pre-determined controls?	Yes	Yes	Yes	Yes	Yes
N	61,799	61,799	61,799	61,799	61,799

Notes: All models included pre-determined and time-varying controls, as listed in Table 1. The estimates of the covariates are omitted for clarity. Standard errors are clustered at the individual level and reported in parentheses. "Early user" is an indicator of whether the subject first used marijuana before age 18. "Intense user" is an indicator of whether the subject used marijuana for more than three days in the last 30 days at the time of the survey. $^{*}p < 0.1$ ** $p < 0.05$ *** $p < 0.01$.