Externalities of Marijuana Legalization: Marijuana Use in Non-Legalizing States*

Marit Hinnosaar[†] Elaine M. Liu[‡] Eva Loaeza-Albino[§]

September 14, 2023

Abstract

We study the impact of distant connections on marijuana use. Leveraging the Facebook Social Connectedness Index, which measures the strength of connectedness between geographic areas based on Facebook friendship ties, we explore the impact of connections to states where recreational marijuana use is legalized on marijuana use and workplace drug testing positivity rates in areas where marijuana remains illegal. The findings reveal that areas which are more connected to legalized states exhibit higher rates of marijuana use and workplace drug testing positivity even after controlling for geographic proximity to the legalized states. The results suggest that even connections beyond closed proximity can play a significant role in shaping individuals' behaviors. Our findings of the externality of legalization in one state to other more connected out-of-state areas imply that studies that estimate the impact of legalization using a standard difference-in-differences approach without taking into account the network underestimate the direct effect on the state that legalizes.

*We thank Manuela Angelucci, Sonia Bhalotra, Laura Dague, Clément de Chaisemartin, Mary Evans, Toomas Hinnosaar, Richard Murphy, Rosalie Pacula, Maria Petrova, Kevin Schnepel, and participants at ASHEcon and Princeton Industrial Relations Section Centenary Symposium for helpful comments and suggestions. We also thank Riley Wilson and Michael Bailey for generously sharing the data.

[†]University of Nottingham and CEPR, marit.hinnosaar@gmail.com.

[‡]University of Houston, NBER, IZA, and HCEO, emliu@uh.edu.

[§]University of Houston, edloaeza@cougarnet.uh.edu.

1 Introduction

Many studies have demonstrated the influence of peers on individuals' behaviors and outcomes (Sacerdote, 2001; Kremer and Levy, 2008; Carrell, Fullerton and West, 2009; Sacerdote, 2011; Card and Giuliano, 2013). Typically, these studies focused on how peers physically close to someone, like roommates or classmates, affect them. However, the landscape of peer dynamics has undergone a dramatic transformation in the past two decades due to the widespread adoption of smartphones and the pervasive presence of social media. In the era preceding the advent of smartphones, interactions with acquaintances met only sporadically throughout the year held limited sway over one's decisions and behaviors. The rise of smartphones has significantly altered this dynamic, enabling geographically distant peers to exert their influence through updates and feedback shared via social media platforms. In the context of this paper, we aim to examine the extent to which these geographically dispersed peers can shape our behaviors.

Nevertheless, akin to research on close-proximity peers, the identification of peer effects poses a challenge, given that the selection of peers is inherently endogenous. Individuals who share similarities tend to naturally gravitate toward each other as peers, making it difficult to distinguish between peer effects and the influence of other factors. In our analysis, we employ a strategy akin to the one proposed by Wilson (2022). We operate under the assumption that social networks remain fixed and leverage temporal variations in the enactment of state-level marijuana legalization laws. First, we explore changes in one's own marijuana use resulting from marijuana legalization changes within their own state. Subsequently, we examine whether such changes extend to the marijuana use patterns of distant peers in an aggregated fashion, which will be described in more detail later.

The study of marijuana use presents an interesting case, given its varied legal status across the United States. Federally, marijuana use remains illegal; however, as of August 2023, 23 states and Washington D.C. have legalized recreational marijuana. Remarkably, approximately 20% of American adults aged 18 and older reported marijuana use in the past year. This figure surges to 38% among individuals aged 21 to 25, and even within the 41-45 age group, 21% acknowledged marijuana use within the past year.[1] Many of these individuals reside in states where marijuana use remains illegal under federal and state law. Previous research on marijuana legalization has indicated that state-level legalization leads to increased marijuana use within the legalized states (Hollingsworth, Wing and Bradford, 2022). Additionally, other research has explored the cross-border spillover effects of mari-

[1]Source: 2021 NSDUH Detailed Tables, Table 1.7B, published January 4, 2023, accessed August 31, 2023, `https://www.samhsa.gov/data/sites/default/files/reports/rpt39441/NSDUHDetailedTabs2021/NSDUHDetailedTabs2021/NSDUHDetailedTabs2021.htm#tab1.7b`

juana legalization (Hansen, Miller and Weber, 2020). Our study delves into the intriguing question of how one state's marijuana legalization can impact even geographically distant areas, extending beyond the immediate bordering localities.

Our analysis is conducted at the aggregated level, utilizing the Social Connectedness Index (SCI) created by Facebook as our measure of social connections. The SCI reflects the likelihood of Facebook connections between residents of any two US counties and any two zip codes. While Facebook users do not represent a fully representative sample of the population, a PEW Research Center Survey (2021) reported that 69% of American adults have used Facebook at some point, and this index has been employed in other economic research papers (including Bailey et al. (2018*c*, 2019); Wilson (2022)). Furthermore, we use this index as a proxy for social connections. To the extent that the SCI may introduce noise in measuring true social connections, our analysis might be conservative, providing a lower estimate of the true impact. Using this index, we construct a measure of exposure to legalized marijuana, which varies across localities and time based on the strength of connections to legalized states and the timing of legalization. We gauge marijuana consumption using two aggregate-level datasets: self-reported marijuana usage data from the National Survey of Drug Use and Health (NSDUH) and workplace marijuana testing positivity rates provided by Quest Analytics. One concern with the NSDUH is that marijuana legalization might affect respondents' truthfulness in reporting marijuana use. While the Quest dataset is not a representative sample of the population, it helps alleviate concerns related to potential divergence between self-reported marijuana usage and actual occurrence.

We regress on the self-reported marijuana use or workplace marijuana positivity rates, the calculated measure of exposure to legalized marijuana, and include both locality fixed effects and state-year pair fixed effects. As the state-year pair fixed effects are included, we effectively compare more or less exposed localities within the same state in a given year. We also estimate the impact of exposure to legalization in the event study framework and we conduct robustness checks across various dimensions, including providing heterogeneity-robust estimates and permutation tests.

Our findings reveal that both self-reported marijuana use and workplace marijuana positivity rates exhibit greater increases in areas more strongly connected to states where recreational marijuana use is legal. Specifically, a one standard deviation increase in exposure to states with legalized use corresponds to roughly a one-sixth of a standard deviation increase in self-reported marijuana use and one-quarter of a standard deviation increase in workplace marijuana positivity rates. These coefficients remain similar in magnitude even when the geographic distance between two localities exceeds 500 miles, underscoring the fact that our distant peers indeed exert a notable influence on our behaviors.

Contribution to the literature. Our study adds to the research on spatial externalities of policies, in particular, the cross-border effects of substance use regulations. Prior studies have mainly explored cross-border effects for alcohol (Lovenheim and Slemrod, 2010; Johansson, Pekkarinen and Verho, 2014; Jacks, Pendakur and Shigeoka, 2021) and cigarettes (Merriman, 2010; Harding, Leibtag and Lovenheim, 2012; DeCicca, Kenkel and Liu, 2013). In the context of marijuana, Hansen, Miller and Weber (2020) analyzed the impact of cross-border shopping on sellers' profits. The above research has used geographic proximity, typically, neighboring status, to quantify spatial externalities. Our work focuses on a different measure of connectedness. The measure we are using can include strong connections between even distant regions. We show that the measure facilitates the spread of marijuana use beyond the borders of the state that legalized. While we don't rule out that the impact of Facebook connectedness is driven by some form of geographic connections, we believe the Facebook connectedness provides another approach to model spatial externalities in this setting.

In addition, our results highlight the importance of spatial externalities of policies. If legalization in one state has externalities to other more connected out-of-state areas, then studies that estimate the impact of legalization using a standard difference-in-differences approach comparing states that legalized to those that did not, without taking into account the externalities, underestimate the direct effect of legalization on the state that legalizes.

Our paper also contributes to the literature on the impact of marijuana legalization on marijuana consumption (including Williams and Bretteville-Jensen (2014); Jacobi and Sovinsky (2016).[2] Hollingsworth, Wing and Bradford (2022) have analyzed the effects of recent changes, including recreational marijuana legalization within the United States, focusing on the impact within the same state. In contrast, our paper examines the externalities in areas beyond the legalized states, broadening the scope of analysis.

In the literature on peer effects in risky behaviors, existing economic studies predominantly rely on a quasi-random exposure to fellow students within the same school (Lundborg, 2006; Clark and Lohéac, 2007) or random roommate assignments (Eisenberg, Golberstein and Whitlock, 2014). Our approach differs from previous studies in that we consider networks as given, albeit potentially endogenous, and utilize a policy shock for identification to

[2] Literature on marijuana legalization has also studied effects on consumption of other substances including alcohol, tobacco, opioids, prescription painkillers, and illegal drugs (Kelly and Rasul, 2014; Wen, Hockenberry and Cummings, 2015; Bradford and Bradford, 2018; Powell, Pacula and Jacobson, 2018; Carrieri, Madio and Principe, 2020; McMichael, Van Horn and Viscusi, 2020), traffic fatalities (Anderson, Hansen and Rees, 2013), crime (Adda, McConnell and Rasul, 2014; Gavrilova, Kamada and Zoutman, 2018; Dragone et al., 2019), education (Marie and Zölitz, 2017). Other topics the literature has analyzed include the regulatory policy of the legal marijuana market (Hollenbeck and Uetake, 2021) and cross-border trade (Hansen, Miller and Weber, 2020).

assess changes in behavior within existing social networks.

Finally, our paper adds to the recent list of papers that use the Facebook Social Connectedness Index (including Bailey et al. (2018*c*, 2019); Wilson (2022)). We examine the impact of Facebook connections on risky behavior, specifically marijuana smoking. Marijuana smoking is an intriguing case due to its varying legality based on one's state of residence. Our findings reveal that when friends' states legalize marijuana, it influences the marijuana consumption of individuals residing in areas where marijuana is still illegal.[3] Our paper complements the recent work of Mäckle and Ruenzi (2022) which studies opioid usage and is also using the Facebook Social Connectedness Index.

The paper is organized as follows. Section 2 provides background on marijuana policy. Sections 3 and 4 describe the data and the empirical methodology. Section 5 reports the results. Section 6 concludes.

2 Marijuana legalization

In the US, the legalization of recreational marijuana followed the earlier legalization of medical marijuana, which began as early as 1996 in California. Our analysis specifically focuses on the legalization of recreational marijuana, which began in Washington and Colorado in 2012. As of August 2023, recreational marijuana has been legalized in 23 states and Washington, DC.[4] For a comprehensive list of recreational marijuana legalization events, please refer to Table A.1 and Figure A.1 in the online appendix.

The legalization of recreational marijuana can be viewed as a two-step process. The first step involves legalizing the use of marijuana, while the second step entails licensing the sale of marijuana. The latter step ensures that recreational marijuana is readily available to consumers. In our primary analysis, we will focus on the impact of marijuana use legalization. The analysis of licensed sales is presented as a robustness check (Figure 3) and the results are similar to our main estimates.

3 Data

The analysis utilizes three main data sources. To measure marijuana use, we utilize National Survey on Drug Use and Health (NSDUH) and the Drug Testing Index from Quest

[3]Vannucci et al. (2020) provides an overview of papers that examine the relationship between social media use and risky behaviors. All of these papers point to a correlation between the use of social media and risky behaviors.

[4]Source: `https://en.wikipedia.org/wiki/Legality_of_cannabis_by_U.S._jurisdiction`, accessed August 31, 2023.

Diagnostic. To measure connectedness, we use Facebook's Social Connectedness Index (SCI).

National Survey on Drug Use and Health. NSDUH is a nationally representative survey of the US population aged 12 and above. Specifically, we rely on the NSDUH Substate report public data, which provides the prevalence of marijuana use in the previous year and the previous month. Previous studies in the marijuana literature have also used the NSDUH aggregated public data (Hollingsworth, Wing and Bradford, 2022; Choi, Dave and Sabia, 2019). The NSDUH Substate report combines data from multiple years (2002-2004, 2004-2006, 2006-2008, 2008-2010, 2010-2012, 2012-2014, 2014-2016, 2016-2018).[5] In the paper, to simplify the exposition, we refer to each 3-year interval by the interval end year.[6] In the main analysis, we restrict the sample to 250 substates (out of 406 substates) where boundaries have remained constant over time, representing approximately 85% of the US population. But the results remain similar when we relax the restriction (see Online Appendix). See Figure A.2 for substate sample coverage. Our main outcome variable is the percentage of populations aged 18 and above that report using marijuana at least once in the past year in a given substate. Later, we also investigate marijuana use in narrower age groups.

Marijuana Positivity Rates from Quest Diagnostic. To address concerns about self-reporting bias, we supplement our analysis with nonsurvey data sources. Specifically, we use the data from Quest Diagnostic, which provides annual workplace positivity rates for marijuana at the 3-digit zip code level from 2007 to 2020. Quest Diagnostics is a commercial lab that conducts workplace drug tests on behalf of employers. These data is a non-representative sample of the US workers since only those employers who contract with Quest Diagnostics for workplace drug testing would be present in the dataset. Given the sample selection issue of this dataset, we are viewing the coefficients from the analysis conducted with this dataset as the average treatment effect for this particular group, which might differ from the average treatment effect of the entire population. For all the years 2007-2020, the marijuana positivity rate is available for 84% of the zip codes with residential addresses.[7] Our main sample includes 759 3-digit zip codes. The marijuana positivity rate in about three-quarters of the time is provided not as a single number, but instead, as an interval. For our main analysis, we use the lower bound of the interval. In the robustness analysis, we use the median and the upper bound of the interval. To make it more comparable to outcomes from NSDUH

[5]The NSDUH substate 2018-2020 report was originally available, but as of July 2023, the NSDUH website has removed the 2018-2020 data citing issues with inconsistent modes of data collection during Covid.

[6]See 2016-2018 NSDUH Overview and Summary Substate Region Estimation Methodology.

[7]In the US, there are more than 900 3-digit zip codes in use, but some of these don't are not residential addresses. We exclude from the analysis Washington government-related 3-digit zip codes, military and Guam.

dataset, we calculated a 3-year running average of the positivity rate (e.g., 2007-2009, 2010-2012, etc.). In the robustness analysis, we also provide estimates without using the 3-year averaging. This dataset has also been used by several other papers that examine the impact of marijuana legalization (Abouk, Mansouri and Powell, 2023; Dong, 2021; Hollingsworth, Wing and Bradford, 2022).

Facebook's Social Connectedness Index. The SCI is based on friendship links on Facebook and measures the relative probability of two individuals from different locations being friends on Facebook (Bailey et al., 2018*a*). The SCI data, provided to us from Facebook, is available at two levels and two time periods: county-to-county (for both 2016 and 2021) and 5-digit zip code-to-zip code (only available for the year 2021). To match the level of the main outcome variables, we aggregate the SCI based on population weights at the substate level (using county-level SCI) and the 3-digit zip code level (using 5-digit zip code-level SCI). More details about the construction of population-weighted SCI are presented in online appendix D.

The main variable of interest: exposure to legalization. We would like to measure how exposure via the social network to out-of -state areas that have legalized marijuana affects marijuana use. To do that, first, we construct for each locality a measure of exposure as a population-weighted average of the Social Connectedness Index across all the localities where marijuana is legal in a given period i except those in the current state. The network exposure measure in locality i in state s in period t equals:

$$Exposure_{it} = \sum_{j \notin s}^{J} \frac{1[LegalUseIn\ j]_t \times Population_j}{\sum_{k \notin s}^{J} Population_k} \times SCI_{ij} \tag{1}$$

where $1[LegalUseIn\ j]_t$ is an indicator for legal use in locality j in period t, $Population_j$ is the total population in locality j, and SCI_{ij} is the Social Connectedness Index between locality i and locality j. The measure captures the probability of having friends in areas where marijuana use is legal. We construct this network exposure variable at both locality levels: substate level and 3-digit zip code level.[8] In the analysis, we use the logarithm of the exposure measure.[9]

[8]3-digit zip code exposure is constructed using 2021 zip code-zip code SCI while the substate exposure is constructed using 2021 county-county SCI. In the robustness checks, we have also used 2016 county-county SCI to construct alternative substate exposure, and the results are robust. 2016 SCI is not available at the zip code level to us.

[9]The previous literature using the index, also used it in the logarithm form (Bailey et al., 2020, 2021*a*).

Geographic measures and economic conditions. To investigate the role of physical connectedness in addition to the online social network, we calculate the minimum distance from each locality to any state that has legalized marijuana using the centroid-coordinates of each locality. The minimum distance variable for years prior to 2012 (first state marijuana legalization) was set to a constant of two times the maximum post-2012. For the econometric analysis, the exact specification of the distance prior to 2012 does not matter as state-year fixed effects would have absorbed it. When a state has legalized marijuana, the minimum distance would be 0. In our analysis, we include a standardized version of this variable as a control.

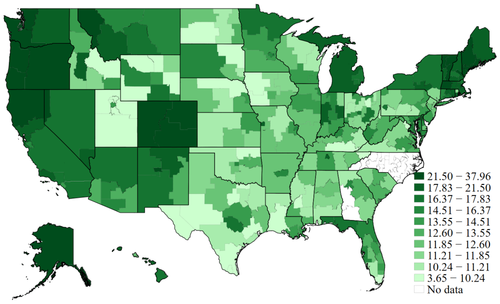
To control for current economic conditions, we use per capita income from the Bureau of Economic Analysis and the unemployment rate from the Bureau of Labor Statistics. We calculate the income to 2018 dollars using the consumer price index for all urban consumers. More details of the data are presented in online appendix D.4.

Summary statistics. Table 1 presents summary statistics for our main variables as a snapshot in the latest period available. On average, 14.6 percent of the population aged 18 and above report using marijuana in 2016-2018. The workplace marijuana positivity rate is about 2 percent in 2020. Importantly, in many states, at the locality level, there is substantial variation in both marijuana use and exposure to legalization (Figure 1).

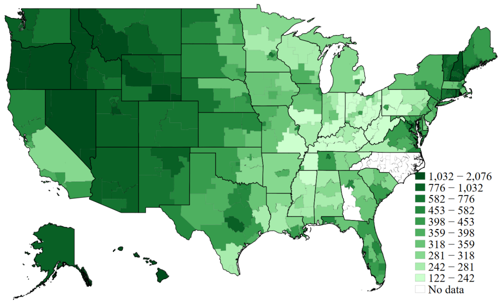
Table 1: Summary statistics

	mean	sd	min	max
Panel A: Substate Level Data				
Marijuana use in the past year	14.65	4.42	6.35	39.37
Log of network exposure	6.09	0.57	4.81	7.64
Standardized min. distance to a legalized state	0.00	1.00	-1.23	5.80
Unemployment Rate	4.33	1.07	2.17	8.23
Per capita income in thousands of dollars	50.28	14.30	26.79	179.74
	mean	sd	min	max
Panel B: 3 Digit Zip Code Level Data				
Workplace drug test marijuana positivity rate	2.00	0.93	0.00	6.50
Log of Network Exposure	7.08	0.45	6.15	8.89
Standardized min. distance to a legalized state	-0.00	1.00	-1.07	9.67
Unemployment rate	5.12	1.25	2.53	11.08
Per Capita Income in thousands of dollars	52.75	16.32	28.27	216.98

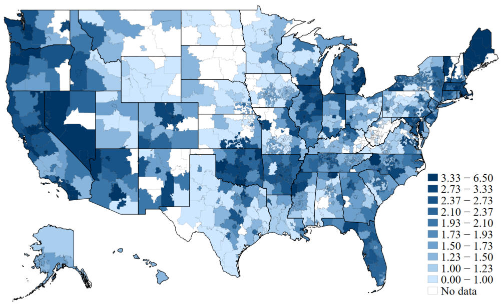
Notes: In Panel A, observation is a substate in NSDUH. There are 250 substates in our sample. In Panel B, observation is a 3-digit zip code. There are 759 3-digit zip codes in our sample. The table presents statistics from the latest available time period (2018 for panel A and 2020 for panel B; except the reported marijuana use, which is the three-year average from the 2016-2018 waves of the NSDUH Public Substate Report).



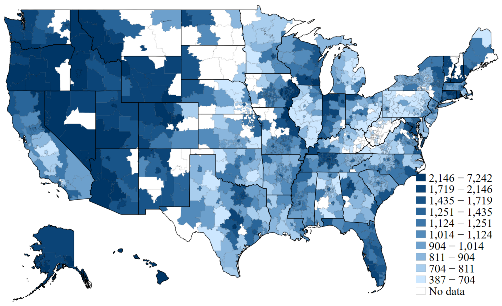
(a) Incidence of Marijuana Use in the Past Year at substate level, 2016-2018



(b) Network Exposure to states with legal use at substate level, 2018



(c) Workplace marijuana positivity rate at 3 digit zip code level, 2018-2020



(d) Network Exposure to states with legal use at 3 digit zip code level, 2020

Figure 1: Geographic Distribution of our Outcome Variables and Network Exposure Measure

Notes: Two substates in Georgia were excluded, and all substates from North Carolina were excluded from the sample because they changed the substate borders and are not comparable across NSDUH waves. The 3 digit zip code analysis includes localities in these states.

4 Empirical strategy

Fixed-effects regression. Using the network exposure measures we estimate the regression where the probability of marijuana use in locality i in state s in period t equals:

$$\begin{aligned} MarijuanaUse_{it} &= \beta \log(Exposure_{it}) + X'_{it}\alpha + LocalityFE_i \\ &+ StateYearFE_{st} + \varepsilon_{it} \end{aligned} \tag{2}$$

where the outcome variable $MarijuanaUse_{it}$ is either the percentage of the population that reported using marijuana in the past year or the workplace drug test marijuana positivity rate. $Exposure_{it}$ is the main variable of interest and is defined by equation (1). X'_{it} is a vector of time-varying covariates, such as the unemployment rate and the logarithm of per capita income. The regression includes fixed effects for each locality, $LocalityFE_i$. It also includes state-year pair fixed effects, $StateYearFE_{st}$. The state-year fixed effects control for state-level changes, for example, when a state itself legalizes marijuana. As the state-year fixed effects are included, we are effectively comparing localities within the same state-year. We cluster standard errors at the state level.

Some might worry that the level of SCI reflects the distance between two localities, and we are capturing the behavior of two close localities which might experience similar shocks. For example, a substate that is neighboring Colorado, which legalized recreational marijuana in 2012, may experience an increase in marijuana use in 2012 because its physical proximity and an increase in easy access to marijuana. The same substate could also have a high SCI with substates in Colorado because of its physical proximity. We might attribute the importance of geographic proximity onto the SCI due to the omitted variable bias. Therefore, we include the shortest distance to any state that legalized marijuana.

Figure A.3 in the online appendix illustrates our identification strategy. Prior to 2012, when the first state legalized marijuana, the network exposure was uniformly zero across all substates. We display the network exposure at the substate level in Florida for the year 2018 in Panel a. This network exposure was imputed based on the population-weighted network exposure to states that legalized marijuana between 2012 and 2017. Panel b presents the last year's marijuana usage from NSDUH between 2016 and 2018. Our regression would show us whether there is a relationship between panel a and panel b. In one of the robustness checks, we also impute the network exposure (panel a) as a 3-year average, and the results remain robust.

In our main analysis, we present two sets of results. First, we present results from all the time periods (the standard difference-in-difference specifications), then we estimate regression (2) using data from only three periods: two before the legalization and one after some states have legalized. Specifically, the three periods are 2002–2004, 2010–2012 and 2016-2018 for NSDUH and 2007–2009, 2010–2012 and 2018-2020 for Quest data. The benefit of using less granular data is that now we have only a single period with treatment. If instead, we were to use data from all time periods, then we would have a staggered roll-out of the treatment, where states legalize marijuana in different time periods. Recent econometrics literature has shown that standard difference-in-differences regressions often do not provide valid estimates of the average treatment effect when treatment timing is staggered (for an overview of the literature, see a recent survey by de Chaisemartin and D'Haultfœuille (2022)). Our approach avoids analyzing staggered treatment over time and is likely to give more robust estimates. In the online appendix, we provide heterogeneity-robust estimates developed by de Chaisemartin and D'Haultfœuille (2020). We discuss this in more detail in the next section.

To further simplify the specification and interpretation of the results, we estimate a modified version of regression (2), where instead of the continuous measure of network exposure to legalization, we explicitly compare the high and low exposure substates. To do that, we construct an indicator variable $AboveMedianExposure_{it}$ that takes the value of 1 after legalization if, in a given state, the locality has the above median exposure to out-of-state areas that have legalized marijuana. Using the variable we estimate the following regression:

$$\begin{aligned} MarijuanaUse_{it} = \beta AboveMedianExposure_{it} + X'_{it}\alpha + LocalityFE_i \\ + StateYearFE_{st} + \varepsilon_{it} \end{aligned} \tag{3}$$

The coefficient of interest β measures the percentage point increase in the probability of marijuana use in a locality with above-median exposure compared to the locality in the same state with below-median exposure.

Event study. To analyze changes over time and evaluate whether high-exposure localities had similar trends before legalization as the low-exposure localities, we estimate an event study. To estimate an event study, we consider each state's legalization as a separate event and estimate its impact on all localities in other states. The following specification measures the effect of network exposure from the legalization in state k on localities i in state s at time t:

$$\begin{aligned} MarijuanaUse_{it} = \sum_{\tau=-1}^{1} \beta_\tau \log\left(SCI_{ik} \times 1[\tau\ PeriodFromLegalizationInState\ k]\right) \\ + LocalityFE_i + X'_{st} + StateYearFE_{st} + \varepsilon_{it} \end{aligned} \tag{4}$$

where SCI_{ik} is the Social Connectedness Index measuring the relative probability in localities i having friends in state k. The Facebook dataset does not include substate-to-state Social Connectedness Index. Therefore, we calculate the locality i to state k Social Connectedness Index from the county-to-county or zipcode-to-zipcode indexes using the population-weighted average. More details of the imputation are in the online appendix.

In our main specification, instead of estimating regression (4) separately for each state that legalized, we stack the data and estimate the average effect across legalization events. Event study estimates separately for each state that legalized are presented in the online appendix. The stacked event study framework has been used for example by Autor, Donohue and Schwab (2006); Cengiz et al. (2019); Deshpande and Li (2019) and in a more similar setting by Wilson (2022). An observation then is locality i, time period t, and state legalization event k triplet. We estimate the following event study regression:

$$\begin{aligned} MarijuanaUse_{itk} = & \sum_{\tau=-1}^{1} \beta_\tau \log\left(SCI_{ik} \times 1[\tau\ PeriodFromLegalizationInState\ k]\right) \\ & + LocalityEventFE_{ik} + StateYearEventFE_{stk} + \varepsilon_{itk} \end{aligned} \tag{5}$$

where we include locality-event pair fixed effects $LocalityEventFE_{ik}$, and state-year-event triplet fixed effects. To be able to have a balanced panel, it is not possible to estimate the regression across legalization events that are either at the beginning or end of the time period. Therefore, our analysis focus on the analyzing the impact of those states that implemented the laws between 2012 and 2016. Similar to the (2) we also conduct the analysis both using all years or restrict the analysis to only 3 periods: 2002-2004, 2010-2012, and 2016-2018 (for NSDUH) and 2007-2009, 2010-2012, 2018-2020 (for Quest).

Identification. As we include state-year fixed effects, we compare marijuana usage in two localities within the same state and year, where one locality experiences a larger increase in exposure to legalized marijuana via more connections to states that have legalized. Additionally, By including locality fixed effects, we account for potential differences in the level of consumption between localities. The identifying assumption is that, in the absence of legalization, localities with more or fewer connections to legalized states would have exhibited the same behavior over time.

To further support the identifying assumption, in online appendix we analyze whether pre-legalization trends in marijuana use are correlated with changes in exposure. Figure A.4a presents scatter plots of residualized changes in marijuana use between 2004 and 2012 (prior to the first marijuana legalization in 2012) against changes in exposure between 2012 and 2018 (during the period of incremental state-level marijuana legalization). However, this correlation is not statistically significant at the 10 percent level. In contrast, Figure A.4b shows a significant positive correlation between changes in marijuana use and changes in network exposure between 2012 and 2018. Similar figures are presented in Figure A.4c and A.4d using workplace marijuana testing data. Although Figure A.4c shows a positive and significant correlation, the slope is only one-third the size compared to the post-legalization period in Figure A.4d. In addition, in this set of figures, we only control for state-fixed effects. Later, when we examine the parallel trend assumption using a event-study framework, we would see that there is no differential trend between places with high and low connections to legalized states.

A potential concern arises regarding the construction of the network exposure measure using the 2021 Social Connectedness Index (SCI). Most recreational cannabis legalization occurred between 2015 and 2020, raising the possibility of legalization influencing the network. However, it is important to note that Facebook claims the network measure to be highly stable over time. The correlation between the county-to-county SCI for 2016 and 2021 in our sample is 0.9. Furthermore, in the robustness check section, we find that the results for marijuana use remain positive significant at 5% when utilizing either the 2016 or 2021 county-to-county SCI to construct the exposure measure. This further confirms the stability of the SCI over time and strengthens our confidence in the results.

5 Main results

Evidence of the direct effect of marijuana legalization on the state that legalized. Before delving into our main analysis of the network's role, it is crucial to confirm the direct impact of marijuana legalization on marijuana use within the states that have legalized it. While this relationship has been demonstrated in the study by Hollingsworth, Wing and Bradford (2022), it is necessary to validate it ourselves as we utilize aggregated marijuana data at the substate level instead of the state level, as done in their study.

To accomplish this, we employ a panel data fixed effects regression model, where the outcome variable is the percentage of the population aged 12 and above who reported marijuana use in the past year. The key variable of interest indicates the legalization status of marijuana use within a state. Each regression incorporates substate fixed effects and state-year fixed effects to account for the variability across different substates and over time.

In columns 1 and 4 of Table 2, we observe that the legalization of marijuana use does indeed lead to a significant increase of 3.7 percentage points in marijuana use within the states that have legalized it. This represents a substantial change, considering that prior to legalization, only 15 percent of individuals aged 18 and above reported using marijuana in the past year. Hence, the legalization resulted in an approximate 25% increase in yearly usage. In Panel B, we present the results using Quest dataset. Legalization of marijuana use increases the marijuana positivity rate by 0.65 percentage points, which is about a 32% increase.

Panel data fixed effects results. Table 2 Columns 2 and 5 present the estimated effect of network exposure to the states that have legalized marijuana. It shows that marijuana consumption, both in reported use and marijuana positivity rates, increases more in an area that is more connected to the states that legalized compared to an area in the same state that is less connected to the state that legalized. A one standard deviation increase in exposure to

Table 2: Direct and indirect impact of exposure to states that have legalized marijuana on marijuana outcomes

	Panel A: Marijuana Use in Past Year					
	All Years: 2004-2018			3 Periods: 2004, 2012, 2018		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized	3.651***			3.771***		
recreational cannabis	(1.104)			(1.276)		
Log. exposure to		1.010***	1.033***		1.275***	1.293***
states with legal use		(0.257)	(0.290)		(0.420)	(0.443)
Standardized min.			0.092			0.080
distance to a legalized state			(0.315)			(0.335)
Year-state FE	No	Yes	Yes	No	Yes	Yes
Substate FE	Yes	Yes	Yes	Yes	Yes	Yes
Num of Substates	250	250	250	250	250	250
Observations	2000	2000	2000	750	750	750
	Panel B: Workplace Positivity Rate for Marijuana					
	All Years: 2007-2020			3 Periods: 2009, 2012, 2020		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized	0.657***			0.772***		
recreational cannabis	(0.117)			(0.128)		
Log. exposure to		0.270***	0.276***		0.477***	0.487***
states with legal use		(0.071)	(0.071)		(0.096)	(0.102)
Standardized min.			0.027			0.031
distance to a legalized state			(0.057)			(0.078)
Year-state FE	No	Yes	Yes	No	Yes	Yes
3 digit zip code FE	Yes	Yes	Yes	Yes	Yes	Yes
Num of 3 digit zip code	759	759	759	759	759	759
Observations	10626	10626	10626	2277	2277	2277

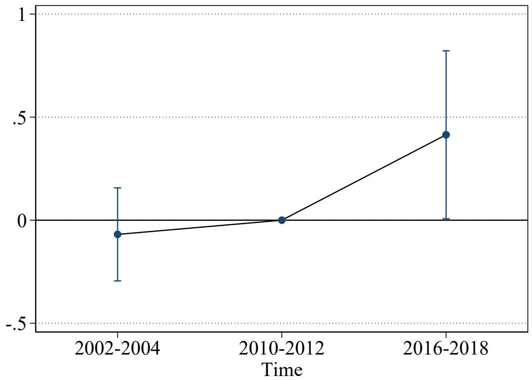
Notes: Columns 1 and 4 are the direct impact of own state's marijuana legalization. Columns 2, 3, 5, and 6 are the impact of exposure to other states that have legalized marijuana. Columns 1 - 3 use data from all periods while Columns 4-6 only use data from only 3 periods: two periods prior to any marijuana legalization and one period post marijuana legalization.

states with legalized use corresponds to a roughly one-sixth of a standard deviation increase in marijuana self-reported use and a quarter of a standard deviation increase in workplace marijuana positivity rates. In Columns 3 and 6, we include a standardized measure of the distance between a given locality to the nearest marijuana legalized locality. There are two noteworthy points to consider here. Firstly, it's worth mentioning that the coefficients for network exposure remain relatively stable when comparing Columns 2 and 3 in this regression analysis. In this particular regression, we've utilized the standardized minimum distance to a legalized locality as a control. Secondly, it's important to highlight that the coefficient associated with the minimum distance measurement lacks statistical significance.

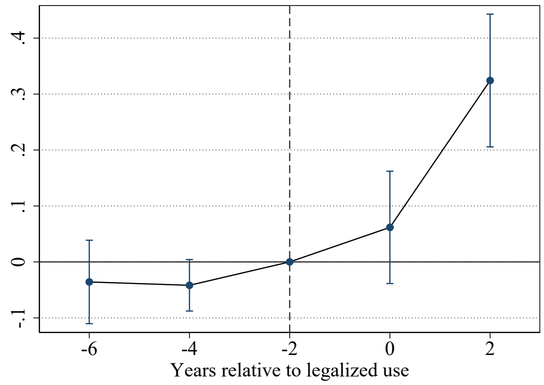
To elaborate further, when we introduce the distance measurement without factoring in network exposure in Panel A, we observe a negative and statistically significant coefficient for the years 2014 and 2016. However, this significance diminishes when we examine the year 2018. This implies that in the earlier years of marijuana legalization, individuals residing further from legalized states were less likely to use marijuana. One plausible explanation is that in the early years when marijuana was less prevalent, proximity to legalized states indeed played a significant role.

Event study results. To analyze changes over time and assess whether substates with varying levels of exposure to legalized states had similar trends before legalization, we conduct an event study. Figure 2 presents the event study estimates. As discussed in (5), we present the event study results in two ways. Panels a and c utilize all legalized states in the analysis and we only examine the 3 periods (two periods prior to any legalization, and one period post legalization). In Panel b and d, these are typical event study where the legalized year for a given state is set as year 0. The referenced year is the period prior to the legalization event. In order to keep a balanced panel, we only examine the events where states legalized marijuana use between 2012 and 2016, In these panels, the trends in marijuana consumption prior to out-of-state marijuana legalization were comparable in areas with both high and low social network exposure. This finding suggests that the larger impact observed in high exposure areas is not due to divergent trends before legalization. Moreover, Figures A.5 and A.6 in the online appendix demonstrates that the event study estimates are similar when each event (state passing marijuana legalization law) is examined separately.

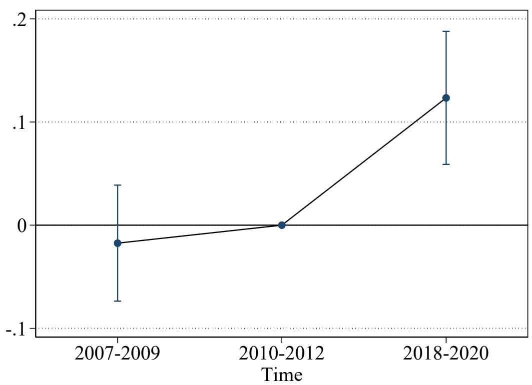
Robustness and Permutation Test Below, we summarize the analyses conducted to assess the sensitivity of our results to alternative functional forms, controls, clustering, and sample. These findings are presented in Figure 3. Each bar represents a point estimate and the 95% confidence interval derived from a separate regression, analogous to the specification in column 3 of Table 2. In Panel A of the NSDUH dataset analysis, we incorporate bootstrapped standard errors (Estimation 1). In Estimation 2, we deliberately omit all states that have previously legalized marijuana. In Estimation 3, we extend this exclusion to areas within a 500-mile radius of any legalized states. This selective exclusion of localities near legalized states allows us to investigate whether geographically distant counterparts exert a comparable level of influence compared to those in close proximity. Notably, the similarity in coefficients between Estimation 3 and the original specification (highlighted in red line) suggests that these distant peers also play a role in shaping our behaviors. We also explore



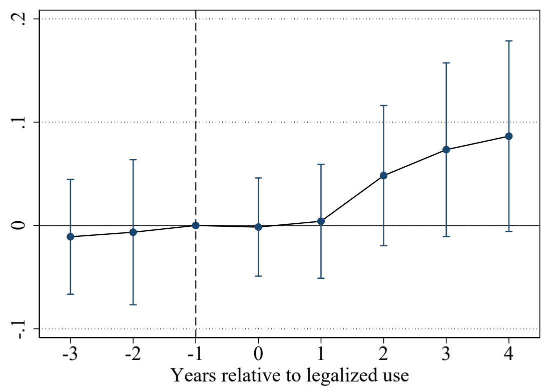
(a) Marijuana use in past year



(b) Marijuana use in past year



(c) Workplace drug test



(d) Workplace drug test

Figure 2: Event study: the impact of the probability of having out-of-state friends exposed to a new state marijuana legalization on the marijuana use in the past year (figures 2a and 2b), and on positivity rate for marijuana in workplace drug tests (figures 2c and 2d)

Notes: Figure 2a uses data of three time periods: 2002-2004, 2010-2012, and 2016-2018; similarly, Figure 2c uses data of 2007-2009, 2010-2012, and 2018-2020. The data is stacked, each area N times where N is the number of states that legalized from December 2012 until December 2017. (Results for each state legalization separately are presented on Figures A.5 and A.6.) Instead of substate fixed effects or county fixed effects, there are now substate-event fixed effects (Figure 2a and Figure 2b) or county-event fixed effects (Figure 2c and Figure 2d). Similarly, instead of state-year fixed effects there are now state-year-event fixed effects. Standard errors are clustered at either substate level for marijuana use or 3-digit-zip-code level for workplace drug test results. In Figure 2b and Figure 2d , we only use 5 states that have legalized before 2016. For example, Oregon past marijuana legalization law in Jul 2015, year 0 in Figure 2b would correspond to 2014-2016, and 2016-2018 will be marked as year +2; whereas in Figure 2d, 2016 will be marked as year 1, so on and so forth.

alternative specifications for the key variable of interest: using 3-year average of exposure (Est 4), namely raw variable of *Network exposure* (Est 5), as well as standardizing *Network exposure* (Est 6). In our main analysis, we use the legalization of marijuana use as the primary law change to estimate exposure. Here, we provide estimates using the legalization of marijuana licensed sales (Est 7) to determine exposure.

In our main specification, the social connectedness index used was constructed based on Facebook connections in 2021. However, considering the possibility of friendship formation as a result of marijuana legalization and potential reverse causality, we obtained the 2016 version's county-to-county SCI. Consequently, we constructed substate-level SCI in 2016 and used it with the NSDUH data. The results from using the 2016 SCI are reported in Panel a, Estimation 8.[10] Furthermore, in the NSDUH dataset, each respondent reports whether they used marijuana in the previous month or not. We use last month's marijuana use as a dependent variable (Est 9). Finally, we conduct the analysis by self-reported marijuana above age 12 and above and 26 and above (Est 10 and 11).[11] In all of the regression in Panel a, other than Est 9 which has p-value = 0.09, all the other coefficients have p-value below 0.05.

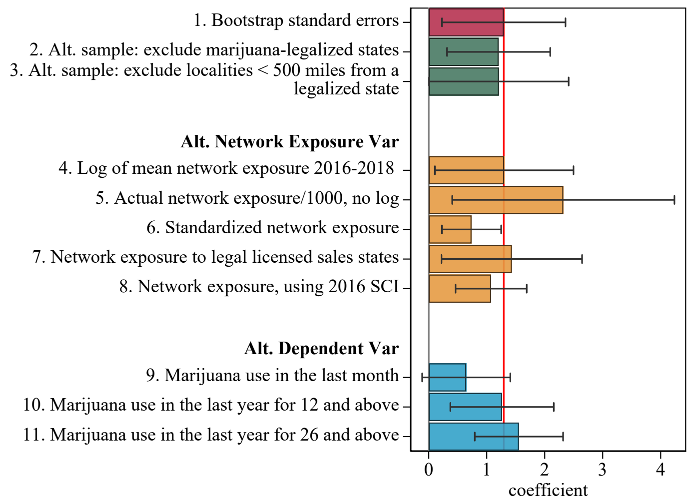
Regarding the workplace drug test data, for each 3-digit zip code in each year, the raw data reports both the lower and upper bounds of the positivity rate for three-quarters of the observations. However, for one-quarter of the time, the raw data only provides a single number for the positivity rate. In our current analysis, we consider this single number as the lower bound. In Table 2, we present the results using the lower bound of the positivity rates as the dependent variable. In Figure 3 Panel b, Est 8 and 9, we display the coefficients obtained when we treat the single number as the median or the upper bound, and use it as the dependent variable, respectively. These coefficients are statistically significant at 5% level.

Our results are also robust to relaxing the restrictions on the NSDUH sample by including those substates that have changed borders over time (Table B.1).

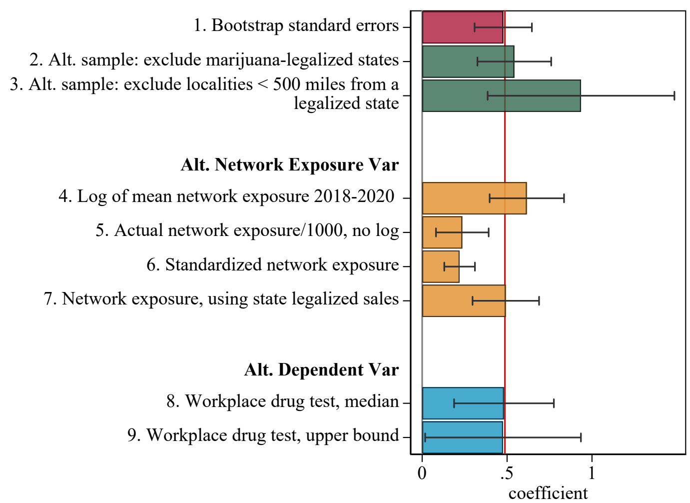
Recent literature has highlighted that the standard two-way fixed effects estimator may be biased when treatment effects exhibit heterogeneity and treatment is staggered (for an overview, refer to a recent survey by de Chaisemartin and D'Haultfœuille (2022)). First, to assess the robustness of the estimates in Table 2, in the appendix (Table A.2), we also present the heterogeneity-robust estimates proposed by de Chaisemartin and D'Haultfœuille (2020) using the data where treatment is a continuous variable. The point estimates remain similar to our main results in Table 2. Second, in the appendix (Table A.3), we also present the estimates from a regression where the treatment variable is binary (below or above median exposure) and treatment is not staggered (only a single period after treatment). When the treatment is not staggered and is binary, the estimates are not subject to the concerns

[10]Since we do not have 2016 zip code to zip code SCI, we cannot generate analogous SCI for Quest data. Additionally, Facebook has changed the SCI scale between 2016 and 2021, making the coefficients on 2016 and 2021 SCI not directly comparable. Therefore, we should not compare the magnitude of the coefficients between 2016 and 2021 SCI.

[11]See Table A.4 in Appendix for more details.



(a) Marijuana Use



(b) Workplace Drug Test

Figure 3: Robustness Checks

Notes: Each bar presents a coefficient and its 95% confidence interval from a separate regression. All regressions are based on the specification in Equation (2), including state-year FE, locality FE, geographic proximity controls, unemployment and local income per capita. Est 1 reports standard errors from bootstrapping 500 times. Est2 uses an alternative sample with only localities where marijuana is still illegal. Est 3 exclude those localities that are within 500 miles from a legalized state. Est 4 uses the natural log of 3 years average of network exposure (e.g. 2016–2018) instead of just the last year (e.g. 2018). Est 5 uses the actual network exposure. Est 6 uses standardized network exposure. Est 7 uses marijuana licensed sales legalization to impute the network exposure. Est 8 in Panel (a) uses 2016 SCI to impute network exposure. Est 9, 10 and 11 in Panel (a) use alternative dependent variable: last month's use, marijuana use for age 12 and above, and marijuana use for age 26 and above, respectively. Est 8 and 9 in Panel (b) use the lower bound and upper bound from workplace marijuana positivity rates as the dependent variable. The red bar indicates the coefficient estimates from Col 3 in Table 2.

raised above, because then the standard difference-in-differences estimator is equivalent to the heterogeneity robust estimator proposed by de Chaisemartin and D'Haultfœuille (2020).

Last, we perform permutation tests. Since our key variable of interest is network exposure to marijuana legalization, we shuffle $LogNetworkExposure$ within a given state and year 500 times and run the regression specification same as Equation 2.[12] We present the distribution of coefficient estimates and t-tests results in Figure A.7. The red bar in each panel indicates the corresponding statistics from Column 3 from Table 2. One can see that our estimates are far from the estimates generated from randomly assigned network exposure. It confirms the importance of the network exposure variable we use in our analysis.

Heterogeneity by age group. In the results not presented here, we also examined marijuana use across different age groups, as reported by NSDUH. Specifically, NSDUH provides data on marijuana use among individuals aged 12 to 17 and those aged 18 to 25, separately. While the findings for these subgroups still indicate positive associations and larger effect sizes compared to the original specification, their estimates are less precise, with one subgroup yielding a p-value of 0.6 and the other 0.11.

Several potential reasons may account for these coefficients not achieving statistical significance. Firstly, we encountered a limitation in terms of sample size for these subgroups. Whenever the number of observations from a specific substate is insufficient, NSDUH does not report the corresponding mean. For instance, we lost nearly one-third of our samples (470 observations compared to 750) in the age group 18 to 25.

Another possibility pertains to the fact that the younger age group, those aged 12 to 18, may not represent the primary users of Facebook. Consequently, the network exposure employed in this study may not be as relevant for this particular age bracket. To illustrate this, consider an extreme scenario where none of the individuals aged 12 to 18 have access to smartphones. In such a case, we would not expect these social networks to be as influential as they might be among smartphone users.

Mechanism. To learn about the mechanism of how exposure to legalization via social network affects marijuana consumption, we isolate the variation in the social network that is due to geographic characteristics. Our goal is to understand the remaining importance of social network when we flexibly control for geographic characteristics. The exercise is motivated by the literature that has shown that geographic variables are strongly correlated with social network (Bailey et al., 2018*b*).

[12] In this exercise, we exclude those states that have fewer than 4 substates. Otherwise, often it would shuffle back to the actual assignment.

Our approach uses the following steps and is described in more detail in Online Appendix C. In the first step, we flexibly model the social connectedness index as a function of geographic characteristics (physical distance). The estimates from the first step are used to calculate the predicted (based on geographic characteristics) exposure according to equation (1). We then use the control function approach and the predicted exposure is used as an instrument for the actual exposure variable.

When controlling for geographic characteristics in a flexible way, then in the NSDUH data, there is not much relevant variation left in the social network measure (Table C.3). In our main regressions in Table 2, we controlled for a minimum distance to states that had legalized and found that the exposure via social network still had a significant effect. However, it was a more restrictive specification, dependent on specific functional forms, compared to the more flexible specification in Appendix C.

We conclude that the social network measure is a nuanced measure of connectedness that captures more than the minimum distance alone. The social network measure is partly explained by geographic variables. Therefore, attributing exposure effects solely to online social networks would be misleading. While we cannot rule out that the impact of exposure to legalization via Facebook connectedness is driven by some form of geographic connections, Facebook connectedness provides another approach to model spatial externalities.

Discussion. Social networks' impact is notoriously difficult to identify. Social connections are likely to form with areas that are more geographically connected either via road or flight networks. When a geographically connected state legalizes marijuana, does marijuana use increase because it is easy to go to buy from there or because of social connections? In this paper, we think of the impact of social networks in a broader sense. In the main regressions, we control for the shortest distance to states that have legalized and interpret the remaining effect as the impact of social connections.

We showed that marijuana legalization in some states has externalities to other more connected areas in other states, which implies that studies estimating the direct impact of policies using a standard difference-in-differences approach without taking into account the externalities underestimate the direct effect of the policy. For example, according to column 4 of panel A of Table 2, the direct effect of the policy is 3.8 percentage points without taking into account the externalities. Indeed, if there were no externalities then 3.8 percentage points increase would be the total direct effect. But column 5 Table A.3 shows that the externality to more connected areas (compared to less connected areas) is about 0.4 percentage points. Suppose for simplicity that in less connected areas there were no externalities. Then back-of-the-envelope calculations suggest that the direct effect of the policy is instead about 4 percentage points. Hence, without taking into account the externality, we underestimated the direct effect by about five percent.

6 Conclusion

In conclusion, this paper investigates the spatial externalities of marijuana legalization on marijuana use. We find that connections to states where recreational marijuana use is legalized have a significant impact on marijuana use and workplace drug testing positivity rates in areas where marijuana use remains illegal. Our analysis utilizes the Facebook Social Connectedness Index, which measures the strength of connectedness between different geographic areas based on Facebook friendship ties, as a measure of exposure to legalized marijuana among Facebook friends. Specifically, we compare within a given state, a one standard deviation increase in exposure to states with legalized use corresponds to a roughly 1/6 of a standard deviation increase in reported marijuana use and 1/4 of a standard deviation increase in workplace marijuana positivity rates. Overall, our findings highlight the importance of spatial externalities of policies. The connections can lead to increased adoption of risky behaviors, such as marijuana smoking, even in areas where marijuana use is still illegal.

References

Abouk, Rahi, Mohammad Mansouri, and David Powell. 2023. "Can Access to Non-Opioid Pain Management Therapies Reduce Overdose Rates? Evidence from Medical Cannabidiol Laws." *Evidence from Medical Cannabidiol Laws.*

Adda, Jérôme, Brendon McConnell, and Imran Rasul. 2014. "Crime and the Depenalization of Cannabis Possession: Evidence from a Policing Experiment." *Journal of Political Economy*, 122(5): 1130–1202.

Anderson, D. Mark, Benjamin Hansen, and Daniel I. Rees. 2013. "Medical Marijuana Laws, Traffic Fatalities, and Alcohol Consumption." *The Journal of Law and Economics*, 56(2): 333–369.

Autor, David H, John J Donohue, III, and Stewart J Schwab. 2006. "The Costs of Wrongful-Discharge Laws." *The Review of Economics and Statistics*, 88(2): 211–231.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel. 2021*a*. “International trade and social connectedness.” *Journal of International Economics*, 129: 103418.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel. 2021*b*. “International trade and social connectedness.” *Journal of international economics*, 129: 103418–.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel. 2019. “House Price Beliefs And Mortgage Leverage Choice.” *The Review of Economic Studies*, 86(6): 2403–2452.

Bailey, Michael, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel. 2020. “Social connectedness in urban areas.” *Journal of Urban Economics*, 118: 103264.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018*a*. “Social Connectedness: Measurement, Determinants, and Effects.” *Journal of Economic Perspectives*, 32(3): 259–280.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018*b*. “Social Connectedness: Measurement, Determinants, and Effects.” *Journal of Economic Perspectives*, 32(3): 259–280. and the Facebook Data for Good Program, Social Connectedness Index (SCI). https://dataforgood.facebook.com/, accessed June 7, 2022.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel. 2018*c*. “The Economic Effects of Social Networks: Evidence from the Housing Market.” *Journal of Political Economy*, 126(6): 2224–2276.

Bradford, Ashley C., and W. David Bradford. 2018. “The Impact of Medical Cannabis Legalization on Prescription Medication Use and Costs under Medicare Part D.” *The Journal of Law and Economics*, 61(3): 461–487.

Card, David, and Laura Giuliano. 2013. “Peer Effects and Multiple Equilibria in the Risky Behavior of Friends.” *The Review of Economics and Statistics*, 95(4): 1130–1149.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. “Does Your Cohort Matter? Measuring Peer Effects in College Achievement.” *Journal of Labor Economics*, 27(3): 439–464.

Carrieri, Vincenzo, Leonardo Madio, and Francesco Principe. 2020. "Do-It-Yourself medicine? The impact of light cannabis liberalization on prescription drugs." *Journal of Health Economics*, 74: 102371.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs*." *The Quarterly Journal of Economics*, 134(3): 1405–1454.

Choi, Anna, Dhaval Dave, and Joseph J Sabia. 2019. "Smoke gets in your eyes: medical marijuana laws and tobacco cigarette use." *American Journal of Health Economics*, 5(3): 303–333.

Clark, Andrew E., and Youenn Lohéac. 2007. ""It wasn't me, it was them!" Social influence in risky behavior by adolescents." *Journal of Health Economics*, 26(4): 763–784.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–2996.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2022. "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey." *Econometrics Journal*, forthcoming.

DeCicca, Philip, Donald Kenkel, and Feng Liu. 2013. "Excise tax avoidance: The case of state cigarette taxes." *Journal of Health Economics*, 32(6): 1130–1141.

Deshpande, Manasi, and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy*, 11(4): 213–248.

Dong, Xiuming. 2021. "Recreational marijuana sales legalization and monday work injury claims." *The BE Journal of Economic Analysis & Policy*, 22(1): 99–121.

Dragone, Davide, Giovanni Prarolo, Paolo Vanin, and Giulio Zanella. 2019. "Crime and the legalization of recreational marijuana." *Journal of Economic Behavior & Organization*, 159: 488–501.

Eisenberg, Daniel, Ezra Golberstein, and Janis L. Whitlock. 2014. "Peer effects on risky behaviors: New evidence from college roommate assignments." *Journal of Health Economics*, 33: 126–138.

Gavrilova, Evelina, Takuma Kamada, and Floris Zoutman. 2018. "Is Legal Pot Crippling Mexican Drug Trafficking Organisations? The Effect of Medical Marijuana Laws on US Crime." *The Economic Journal*, forthcoming.

Hansen, Benjamin, Keaton Miller, and Caroline Weber. 2020. "Federalism, partial prohibition, and cross-border sales: Evidence from recreational marijuana." *Journal of Public Economics*, 187: 104159.

Harding, Matthew, Ephraim Leibtag, and Michael F. Lovenheim. 2012. "The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data." *American Economic Journal: Economic Policy*, 4(4): 169–98.

Hollenbeck, Brett, and Kosuke Uetake. 2021. "Taxation and market power in the legal marijuana industry." *The RAND Journal of Economics*, 52(3): 559–595.

Hollingsworth, Alex, Coady Wing, and Ashley Bradford. 2022. "Comparative Effects of Recreational and Medical Marijuana Laws On Drug Use Among Adults and Adolescents." *The Journal of Law and Economics*, forthcoming.

Jacks, David S, Krishna Pendakur, and Hitoshi Shigeoka. 2021. "Infant Mortality and the Repeal of Federal Prohibition." *The Economic Journal*, 131(639): 2955–2983.

Jacobi, Liana, and Michelle Sovinsky. 2016. "Marijuana on Main Street? Estimating Demand in Markets with Limited Access." *American Economic Review*, 106(8): 2009–2045.

Johansson, Per, Tuomas Pekkarinen, and Jouko Verho. 2014. "Cross-border health and productivity effects of alcohol policies." *Journal of Health Economics*, 36: 125–136.

Kelly, Elaine, and Imran Rasul. 2014. "Policing cannabis and drug related hospital admissions: Evidence from administrative records." *Journal of Public Economics*, 112: 89–114.

Kremer, Michael, and Dan Levy. 2008. "Peer Effects and Alcohol Use among College Students." *Journal of Economic Perspectives*, 22(3): 189–206.

Lovenheim, Michael F., and Joel Slemrod. 2010. "The fatal toll of driving to drink: The effect of minimum legal drinking age evasion on traffic fatalities." *Journal of Health Economics*, 29(1): 62–77.

Lundborg, Petter. 2006. "Having the wrong friends? Peer effects in adolescent substance use." *Journal of Health Economics*, 25(2): 214–233.

Marie, Olivier, and Ulf Zölitz. 2017. ""High" Achievers? Cannabis Access and Academic Performance." *The Review of Economic Studies*, 84(3): 1210–1237.

McMichael, Benjamin J., R. Lawrence Van Horn, and W. Kip Viscusi. 2020. "The impact of cannabis access laws on opioid prescribing." *Journal of Health Economics*, 69: 102273.

Merriman, David. 2010. "The Micro-geography of Tax Avoidance: Evidence from Littered Cigarette Packs in Chicago." *American Economic Journal: Economic Policy*, 2(2): 61–84.

Mäckle, Kai, and Stefan Ruenzi. 2022. "The Social Transmission of Non-Infectious Diseases: Evidence from the Opioid Epidemic."

Powell, David, Rosalie Liccardo Pacula, and Mireille Jacobson. 2018. "Do medical marijuana laws reduce addictions and deaths related to pain killers?" *Journal of Health Economics*, 58: 29–42.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates*." *The Quarterly Journal of Economics*, 116(2): 681–704.

Sacerdote, Bruce. 2011. "Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In . Vol. 3 of *Handbook of the Economics of Education*, , ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 249–277. Elsevier.

Vannucci, Anna, Emily G Simpson, Sonja Gagnon, and Christine McCauley Ohannessian. 2020. "Social media use and risky behaviors in adolescents: A meta-analysis." *Journal of Adolescence*, 79: 258–274.

Wen, Hefei, Jason M. Hockenberry, and Janet R. Cummings. 2015. "The effect of medical marijuana laws on adolescent and adult use of marijuana, alcohol, and other substances." *Journal of Health Economics*, 42: 64–80.

Williams, Jenny, and Anne Line Bretteville-Jensen. 2014. "Does liberalizing cannabis laws increase cannabis use?" *Journal of Health Economics*, 36: 20–32.

Wilson, Riley. 2022. "The Impact of Social Networks on EITC Claiming Behavior." *The Review of Economics and Statistics*, 104(5): 929–945.

A Online appendix: Additional figures and tables

Table A.1: Changes in recreational marijuana laws until the end of 2020

State	Legalized Recreational Cannabis	Licensed Sales Since
Washington	December 6, 2012	July 8, 2014
Colorado	December 10, 2012	January 1, 2014
Alaska	February 24, 2015	October 29, 2016
Washington, D.C.	February 26, 2015	
Oregon	July 1, 2015	October 1, 2015
California	November 9, 2016	January 1, 2018
Massachusetts	December 15, 2016	November 20, 2018
Nevada	January 1, 2017	July 1, 2017
Maine	January 30, 2017	October 9, 2020
Vermont	July 1, 2018	
Michigan	December 6, 2018	December 1, 2019
Illinois	January 1, 2020	January 1, 2020
Arizona	November 30, 2020	

Source: `https://en.wikipedia.org/wiki/Legality_of_cannabis_by_U.S._jurisdiction`, accessed October 13, 2022.

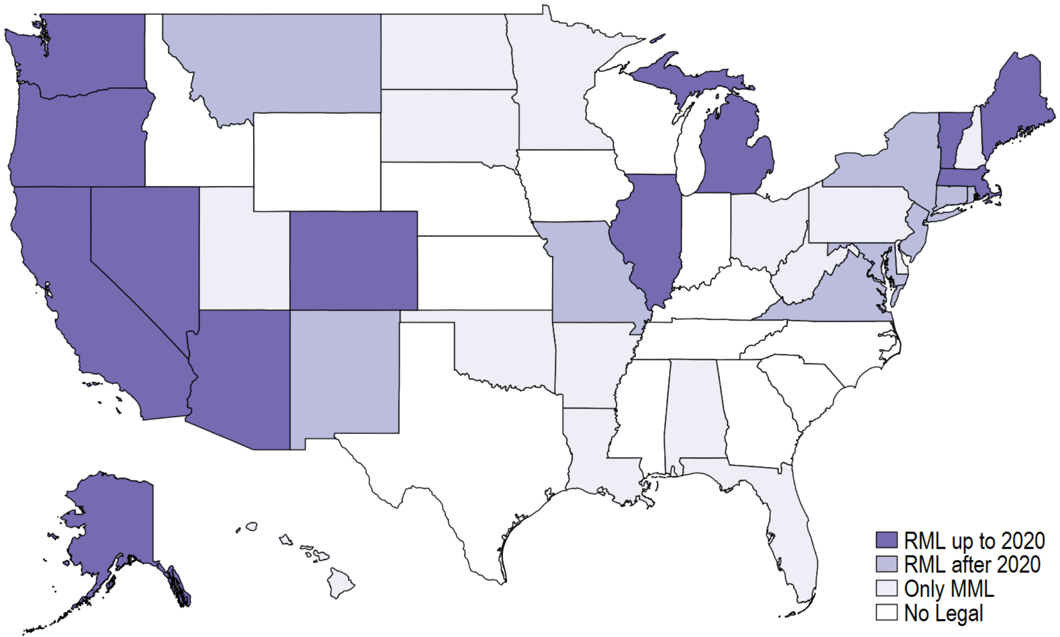


Figure A.1: Medical and Recreational Marijuana Laws

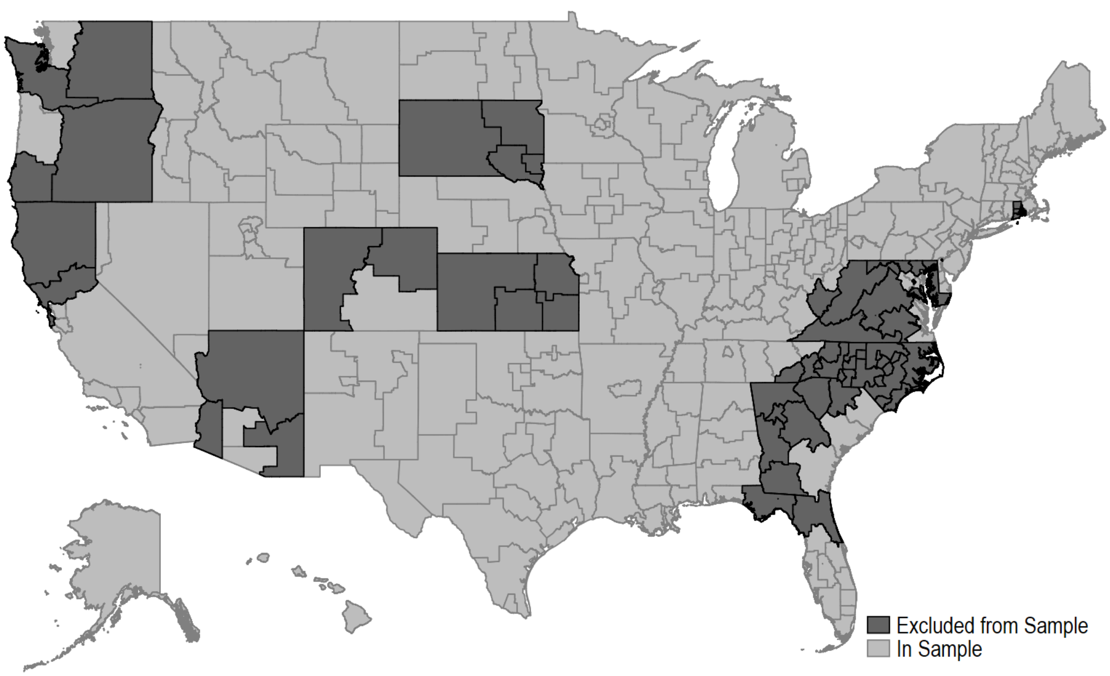
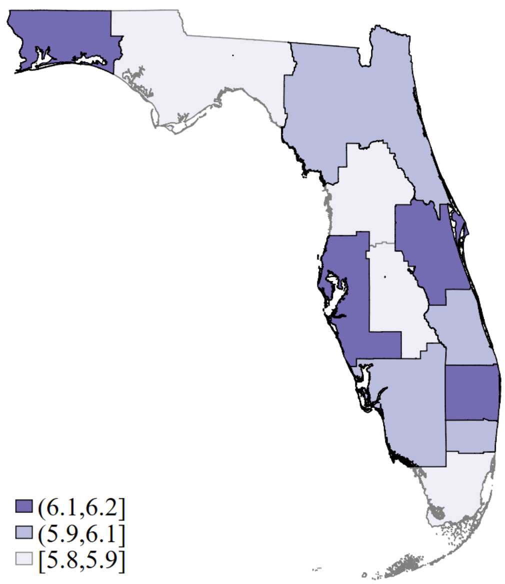
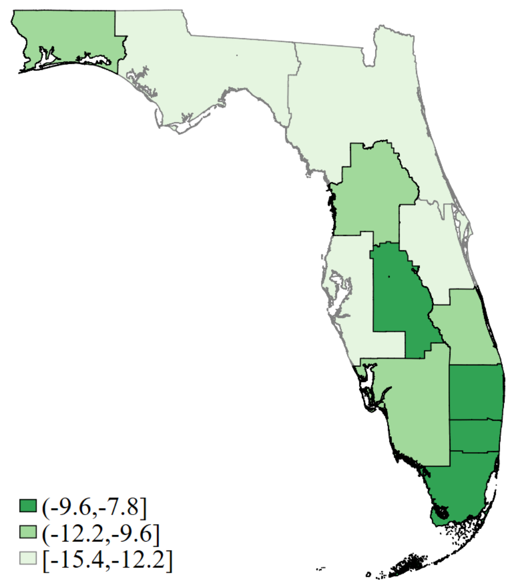


Figure A.2: Distribution of Substates

Notes: All substates in District of Columbia are excluded because there is no variation in the social conectedeness index at the substate level. D.C. only has one county. Two substates in Georgia were excluded. All substates from North Carolina were excluded from the sample.



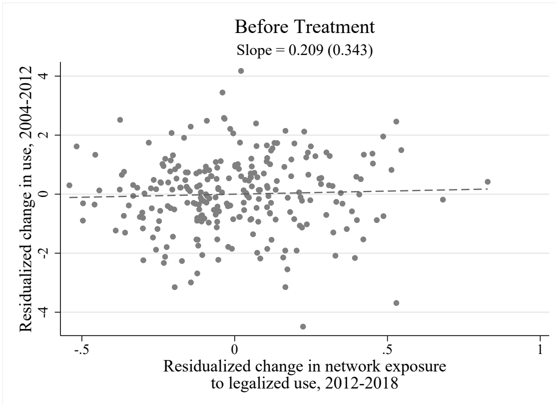
(a) Change in Network Exposure to Legalized States

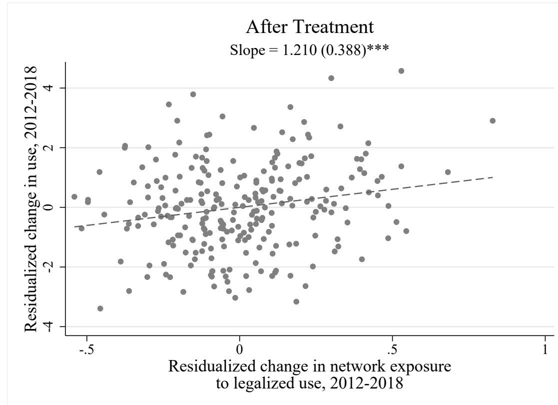


(b) Change in marijuana use during treatment

Figure A.3: **Florida**: Change in the network exposure to legalized use (in 2012–2018) vs change in marijuana use during treatment

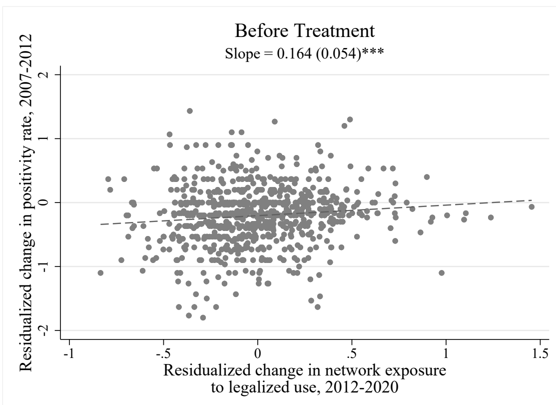
Each block represents a substate in NSDUH. Network exposure was zero in all substates before 2012. Panel a indicates variation in natural log of network exposure at the substate level, and Panel b shows the marijuana use which is available from 2016-2018 wave of NSDUH public data.

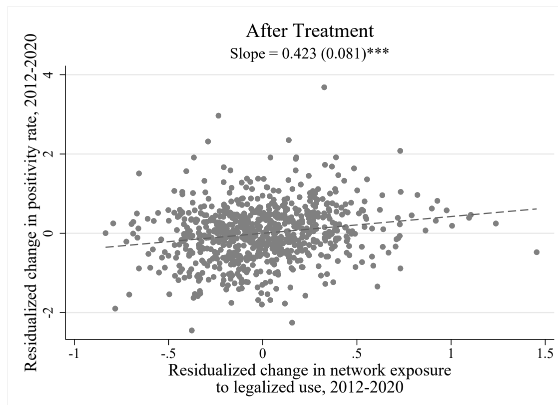




(a) Change in marijuana use before treatment

(b) Change in marijuana use during treatment





(c) Change in employee marijuana positivity rate before treatment

(d) Change in employee marijuana positivity rate during treatment

Figure A.4: Relationship between the exposure to legalization (in 2012–2020) and the change in yearly use before (a and c) and during treatment (b and d)

Notes: Each figure presents a scatter plot of the residualized change in yearly marijuana use or marijuana positivity rates (y-axes) on the residualized changes of the network exposure (x-axes). For each variable, we first take the residuals after controlling for state fixed effects. X-axes in all four figures measure the residualized change in network exposure to legalized use between 2012 and 2018. In Figure A.4a, the y-axis is the change in marijuana use between 2002-2004 and 2010-2012. In Figure A.4b, the y-axis is the residualized change in marijuana use between 2010-2012 and 2016-2018. In Figure A.4c, the y-axis is the residualized change in workplace marijuana positivity rates between 2007-2009 and 2010-2012. In Figure A.4d, the y-axis is the residualized change in workplace positivity rate between 2010-2012 and 2018-2020. The line of best fit is obtained from OLS regression. Its slope coefficient and standard error (in parentheses) are reported on the graph. Figure 1a and 1b utilize NSDUH-substate sample while Figure 1c and 1d utilize Quest data.

Table A.2: Robustness of the impact of Out-of-State Friend Exposure to State Marijuana Legalization on Marijuana Use: heterogeneity robust estimates

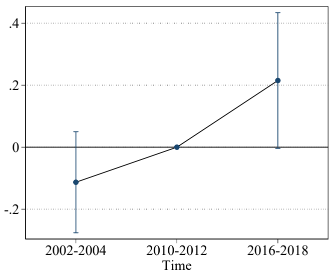
	Marijuana Use in Past Year		Workplace Positivity Rate for Marijuana	
	(1)	(2)	(3)	(4)
Log. exposure to states with legal use	1.144	1.028	0.428	0.501
	(0.552)	(0.665)	(0.120)	(0.160)
State-year FE	No	Yes	No	Yes
Treated localities	125	125	379	379
Localities	250	250	759	759
Observations	750	750	2277	2277

Notes: The coefficients are estimated using the method developed by de Chaisemartin and D'Haultfœuille (2020) and their did_multiplegt Stata package, which is available from the STATA repository. The estimator is the DID_M estimator introduced in de Chaisemartin and D'Haultfœuille (2020). It is a weighted average, across treatment values, of DID estimators comparing the change in the outcome from $t-1$ to t, in groups whose treatment changed, and in groups whose treatment remains the same. DID_M is unbiased even if the treatment effect is heterogeneous across groups. The estimator is obtained assuming that the treatment remained the same if the treatment value in the final period is less than the median. Standard errors are computed using a block bootstrap at the state level with 200 bootstrap replications.

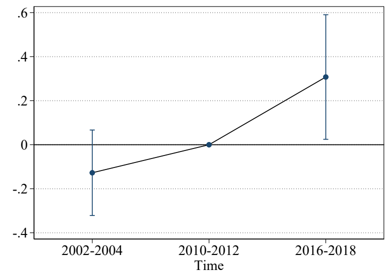
Table A.3: Impact of exposure and distance to states that have legalized marijuana on marijuana outcomes

	Panel A: Marijuana Use in Past Year					
	All Years: 2004-2018			3 Periods: 2004, 2012, 2018		
	(1)	(2)	(3)	(4)	(5)	(6)
Above median	0.466***	0.456***		0.424**	0.419**	
exposure to states with legal use	(0.151)	(0.148)		(0.169)	(0.170)	
Standardized min.		-0.267	-0.321		-0.069	-0.216
distance to a legalized state		(0.289)	(0.304)		(0.345)	(0.360)
Year-state FE	Yes	Yes	Yes	Yes	Yes	Yes
Substate FE	Yes	Yes	Yes	Yes	Yes	Yes
Num of Substates	250	250	250	250	250	250
Observations	2000	2000	2000	750	750	750
	Panel B: Workplace Positivity Rate for Marijuana					
	All Years: 2007-2020			3 Periods: 2009, 2012, 2020		
	(1)	(2)	(3)	(4)	(5)	(6)
Above median	0.227***	0.223***		0.259***	0.257***	
exposure to states with legal use	(0.061)	(0.062)		(0.064)	(0.068)	
Standardized min.		-0.043	-0.063		-0.017	-0.107
distance to a legalized state		(0.057)	(0.057)		(0.089)	(0.095)
Year-state FE	Yes	Yes	Yes	Yes	Yes	Yes
3 digit zip code FE	Yes	Yes	Yes	Yes	Yes	Yes
Num of 3 digit zip code	759	759	759	759	759	759
Observations	10626	10626	10626	2277	2277	2277

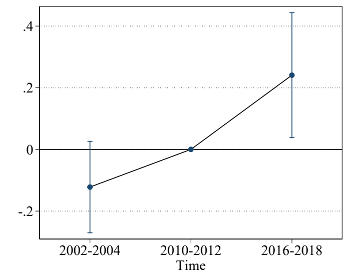
Notes: Columns 1 - 3 use data from all periods while Columns 4-6 only use data from only 3 periods: two periods prior to any marijuana legalization and one period post marijuana legalization.



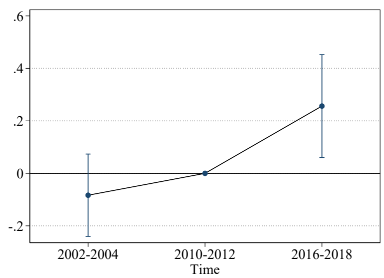
(a) Alaska



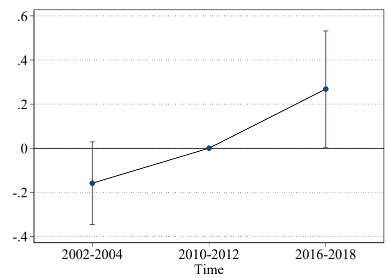
(b) California



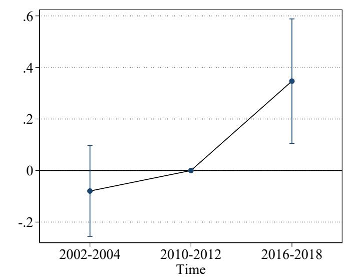
(c) Colorado



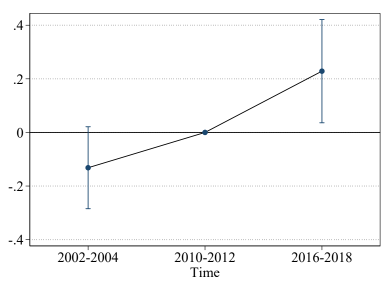
(d) Washington, D.C.



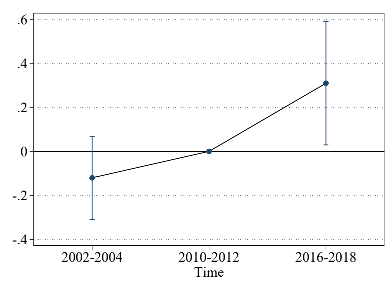
(e) Maine



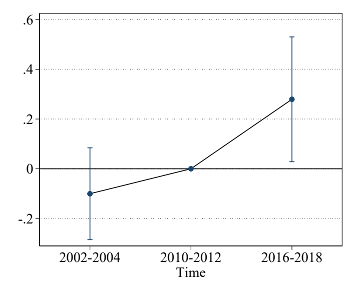
(f) Massachusetts



(g) Nevada



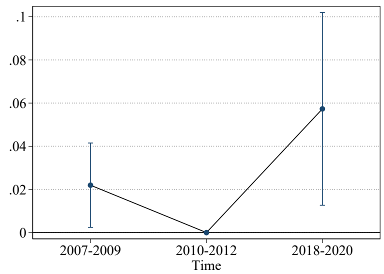
(h) Oregon



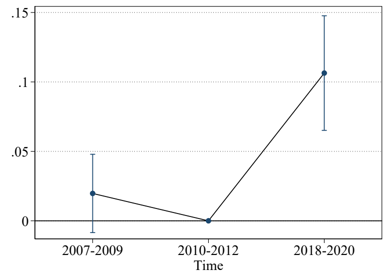
(i) Washington

Figure A.5: Event study separately by each state: the impact of the probability of having out-of-state friends exposed to a new state legalization of marijuana use on the marijuana use in the past year

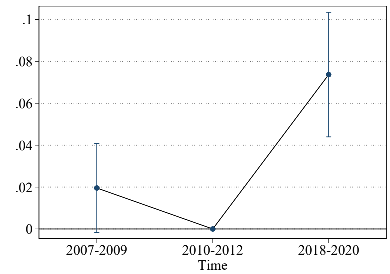
Notes: Includes only 3 time periods: 2002–2004, 2010–2012, 2018-2020. All event studies control for physical proximity, unemployment rate and the logarithm of income. The control variable for physical proximity encompasses the minimum distance to any state that has legalized marijuana. The minimum distance for years before legalization was set to a constant of two times the maximum post-legalization, but the exact specification does not matter as state-year fixed effects absorb it.



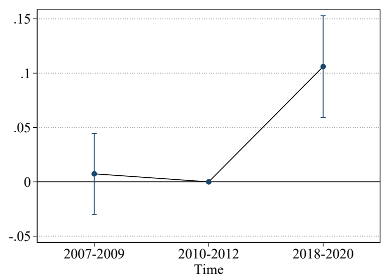
(a) Alaska



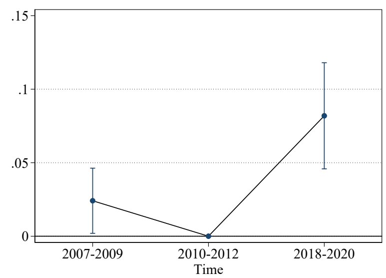
(b) California



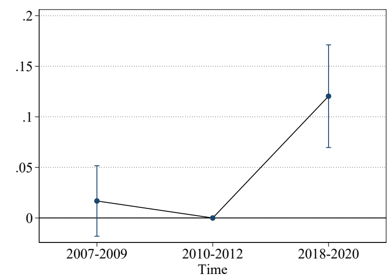
(c) Colorado



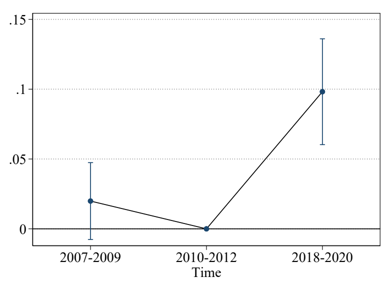
(d) Washington, D.C.



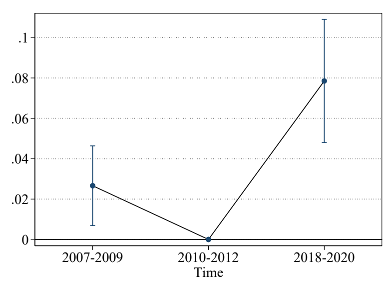
(e) Maine



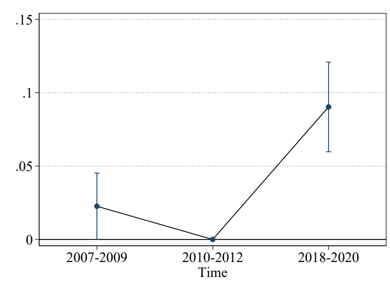
(f) Massachusetts



(g) Nevada



(h) Oregon



(i) Washington

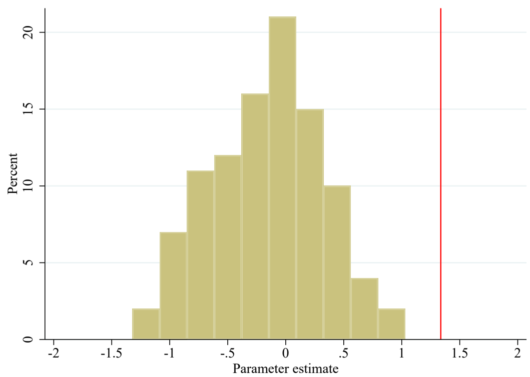
Figure A.6: Event study separately by each state: the impact of the probability of having out-of-state friends exposed to a new state legalization of marijuana use on positivity rate for marijuana in workplace drug tests

Notes: Includes only 3 time periods: 2007–2009, 2010–2012, 2018-2020. All event studies control for unemployment rate and the logarithm of income.

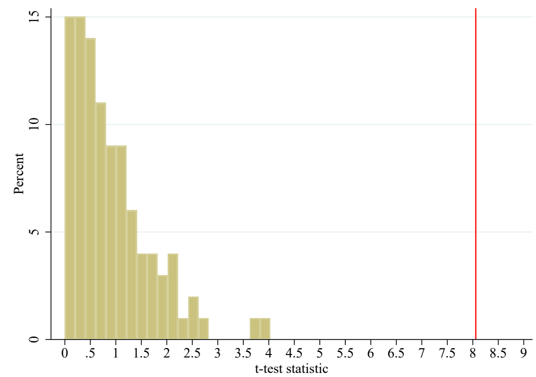
Table A.4: Impact of exposure to other states that have legalized marijuana on marijuana outcomes with changing age groups

	Panel A: Marijuana Use in Past Year for older than 11					
	All Years: 2004-2018			3 Periods: 2004, 2012, 2018		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized recreational cannabis	3.508***			3.633***		
	(1.021)			(1.181)		
Log. exposure to states with legal use		0.921***	0.953***		1.230***	1.264***
		(0.247)	(0.282)		(0.415)	(0.442)
Standardized min. distance to a legalized state			0.130			0.154
			(0.301)			(0.329)
Year-state FE	No	Yes	Yes	No	Yes	Yes
Substate FE	Yes	Yes	Yes	Yes	Yes	Yes
Num of Substates	250	250	250	250	250	250
Observations	2000	2000	2000	750	750	750
	Panel B: Marijuana Use in Past Year for older than 25					
	All Years: 2004-2018			3 Periods: 2004, 2012, 2018		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized recreational cannabis	3.435**			3.651**		
	(1.378)			(1.508)		
Log. exposure to states with legal use		0.985***	1.010***		1.573***	1.555***
		(0.231)	(0.249)		(0.370)	(0.377)
Standardized min. distance to a legalized state			-0.067			-0.038
			(0.275)			(0.295)
Year-state FE	No	Yes	Yes	No	Yes	Yes
Substate FE	Yes	Yes	Yes	Yes	Yes	Yes
Num of Substates	245	245	250	248	248	250
Observations	1960	1960	1994	744	744	748

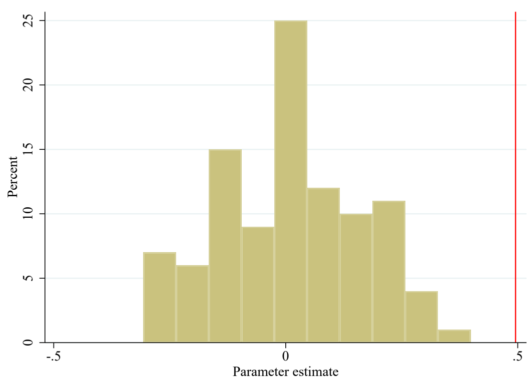
Notes: In Panel A the outcome variable is marijuana use in the past year for older than 11, and in Panel B is for older than 25. Columns 1 and 4 are the direct impact of own state's marijuana legalization. Columns 2, 3, 5, and 6 are the impact of exposure to other states that have legalized marijuana. Columns 1 - 3 use data from all periods while Columns 4-6 only use data from only 3 periods: two periods prior to any marijuana legalization and one period post marijuana legalization.



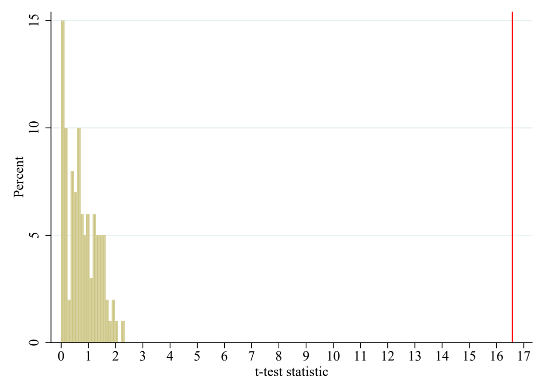
(a) Coefficients from Marijuana Use



(b) T-tests from Marijuana Use



(c) Coefficients from Workplace Drug Test



(d) T-Tests from Workplace Drug Test

Figure A.7: Histogram of Permutation Test Estimates

Notes: Each panel indicates the histogram of coefficients or t-tests from 500 permutations where the specification is the same as Equation 1 but we randomly assign network exposure within a given state and year. The red bar indicates the results from Column 3 from Table 2. Panels A.7a and A.7b utilize NSDUH while Panel A.7c and A.7d use Quest Workplace Test Dataset.

B Robustness Analysis: NSDUH Larger Sample

In order to increase our sample we use an alternative matching of substates over time without aggregating substates (for robustness). One alternative sample was created by selecting the largest county in the substate in the first time period (NSDUH wave 2002-2004) and match that substate to a substate in the next period that includes this county, and so on. For robustness, we repeat the exercise by started from the last period (NSDUH wave 2016-2018) instead. The benefit of these alternative samples is less aggregation.

When using the main sample for the analysis, all variables are weighted by population at the new (aggregate) substate level. When using the alternative samples, all variables are weighted by the population of the baseline wave, that is, either the first wave or last.

Table B.1: Direct and indirect impact of exposure to states that have legalized marijuana on marijuana outcomes

	Panel A: Marijuana Use in Past Year, Baseline: 2002-2004					
	All Years: 2004-2018			3 Periods: 2004, 2012, 2018		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized recreational cannabis	3.931***			4.430***		
	(0.749)			(1.081)		
Log. exposure to states with legal use		1.070***			1.437***	
		(0.260)			(0.409)	
Above median exposure to states with legal use			0.548***			0.513***
			(0.129)			(0.149)
Year-state FE	No	Yes	Yes	No	Yes	Yes
Substate FE	Yes	Yes	Yes	Yes	Yes	Yes
Num of Substates	325	325	325	325	325	325
Observations	2599	2599	2599	974	974	974
	Panel B: Marijuana Use in Past Year, Baseline: 2016-2018					
	All Years: 2004-2018			3 Periods: 2004, 2012, 2018		
	(1)	(2)	(3)	(4)	(5)	(6)
Legalized recreational cannabis	4.374***			4.572***		
	(0.897)			(1.384)		
Log. exposure to states with legal use		1.083***			1.304***	
		(0.284)			(0.445)	
Above median exposure to states with legal use			0.415***			0.395**
			(0.153)			(0.178)
Year-state FE	No	Yes	Yes	No	Yes	Yes
Substate FE	Yes	Yes	Yes	Yes	Yes	Yes
Num of Substates	244	244	244	244	244	244
Observations	1952	1952	1952	732	732	732

Notes: Panel A use an alternative sample that was created by selecting the largest county in the substate in the first time period (NSDUH wave 2002-2004) and match that substate to a substate in the next period that includes this county, and so on. Panel B use an alternative sample were we repeat the exercise by started from the last period (NSDUH wave 2016-2018) instead. Columns 1 and 4 are the direct impact of own state's marijuana legalization. Columns 2 and 5 are the impact of exposure to other states that have legalized marijuana. Column 3 and 6 uses the above median exposure measure. Columns 1-3 use data from all periods while Columns 4-6 only use data from only 3 periods: two periods prior to any marijuana legalization and one period post marijuana legalization.

C Mechanism

To study the role of geographic characteristics, we use a three-step estimation. In the first step, to predict the social network, we regress on the logarithm of the social connectedness index between locality i and j, the logarithm of distance between the localities, and fixed effects for both localities:

$$\log(SCI_{ij}) = \alpha_1 \log(Distance_{ij}) + LocalityFE_i + LocalityFE_j + \varepsilon_{ij} \tag{6}$$

The estimates confirm what was known from the literature that distance has a high predictive power of the social network (Table C.1). Using the estimates we obtain the predicted social connectedness index. Using the predicted index we calculate the predicted exposure to legalization according to equation (1).

We then use a control function approach and the predicted exposure as an instrument for exposure, to estimate the following cross-sectional regression in differences:

$$\begin{aligned} MarijuanaUse_{i,T} - MarijuanaUse_{i,2012} = \beta \log[Exposure_{i,T} - Exposure_{i,2012}] \\ + StateFE_s + \varepsilon_i \end{aligned} \tag{7}$$

where the outcome variable is the change (from 2012 to the latest period T in our sample) in marijuana use in locality i in state s. The outcome variable is regressed on the logarithm of the change in exposure and state fixed effects. Note that the exposure in 2012 (before legalization) equals zero and hence, $\log[Exposure_{i,T} - Exposure_{i,2012}] = \log Exposure_{i,T}$.

To obtain the control function estimates, first, we regress the logarithm of predicted exposure and state fixed effects on the logarithm of exposure (Table C.2):

$$\log Exposure_{i,T} = \log PredictedExposure_{i,T} + StateFE_s + \xi_i \tag{8}$$

From this regression, we obtain the residuals and include these residuals as an additional regressor in regression (7). Estimates of regression (7) without control function are presented in columns 1 and 3 of Table C.3 and with control function (residuals) in columns 2 and 4.

Table C.1: First step; outcome variable: logarithm of the social connectedness index

	Substate level	3-digit zip code level
	(1)	(2)
α_1: Log. distance	-1.200***	-1.124***
	(0.047)	(0.065)
F-statistic of H0: $\alpha_1 = 0$	642.5	296.7
Number of states	43	50
Number of observations	72500	673729

Notes: An observation is a locality-to-locality pair. Fixed effects for both localities are included to control for the locality-specific characteristics. Standard errors are clustered at the state level.

Table C.2: Second step; outcome variable: the logarithm of exposure to legalization

	Substate level	3-digit zip code level
	(1)	(2)
Log. predicted exposure	0.912***	0.958***
	(0.076)	(0.040)
	[0.096]	[0.074]
State FE	Yes	Yes
Number of states	43	50
Number of observations	250	763

Notes: Observation is a locality. The outcome variable is the logarithm of exposure in the latest year in the sample. The explanatory variable is the logarithm of predicted exposure in the latest year in the sample, it is predicted based on the estimates in table C.1 and calculated according to equation (1). All regressions include state fixed effects. Standard errors in parentheses are bootstrapped over steps 1–2 of the estimation with 200 bootstrap samples. Standard errors in square brackets are the naive standard errors clustered at the state level, without taking into account that the variable *Log. predicted exposure* is obtained using the estimates in step one. Stars to indicate significance are based on the bootstrapped standard errors.

Table C.3: Third step; outcome variable: change in marijuana use

	Marijuana use last year		Workplace positivity rate	
	(1)	(2)	(3)	(4)
Log. exposure	1.144**	1.300**	0.424***	0.244**
		(0.592)		(0.111)
	[0.455]	[0.591]	[0.099]	[0.104]
Residuals		-0.313		0.356**
		(0.846)		(0.180)
		[0.816]		[0.166]
State FE	Yes	Yes	Yes	Yes
Number of states	43	43	50	50
Number of observations	250	250	763	763

Notes: The table presents the control function estimates of regression (7) using the logarithm of the predicted exposure as an instrument. Observation is a locality. The outcome variable is the change (from 2012 to the latest period T in our sample) in marijuana use. The explanatory variables are the logarithm of exposure to legalization and residuals from the regression in Table C.2. All regressions include state fixed effects. Standard errors in parentheses are bootstrapped over three steps of the estimation with 200 bootstrap samples. Standard errors in square brackets are clustered at the state level, without taking into account that the variable *Residuals* is obtained using the estimates in steps 1–2. Stars to indicate significance are based on the square bracket standard errors in columns 1 and 3, and on the bootstrapped standard errors in columns 2 and 4.

D Online Appendix: Dataset construction

D.1 Facebook Social Connectedness Index

Facebook Social Connectedness Index (SCI) is available at: `https://dataforgood.facebook.com/` The measure was first constructed and analyzed by Bailey et al. (2018*b*).[13] To construct the index, Facebook users were assigned a location based on the users' information and activity on Facebook, including the stated city on their profile, and device and connection information. The 2021 public release version of the index is constructed in two steps. In step one, the index between counties i and j is calculated as:

$$SocialConnectednessIndex_{ij} = \frac{FacebookConnections_{ij}}{FacebookUsers_i \times FacebookUsers_j} \tag{9}$$

where $FacebookConnections_{ij}$ is the total number of friendship links between the two counties and $FacebookUsers_i$ is the total number of Facebook users in location i. In step two, the index is scaled to have a maximum value of one billion and a minimum of one. For privacy reasons, a small amount of random noise is added, and locations with too few number of users are excluded. The index measures the relative probability of a Facebook friendship link between a given Facebook user in location i and a given Facebook user in location j. If the index is twice as large, a given Facebook user in location i is about twice as likely to be friends with a given Facebook user in location j. The 2021 public release version of the index is calculated as of October 2021.

D.2 Aggregation of the SCI

The finest geographical levels available of the 2021 SCI are US ZCTA (zip code tabulation) and county. Our outcome variables are at the substate level (aggregation of counties) for the NSDUH data and 3 digit-zip code level (aggregation of zip codes) for the Drug Testing Index from Quest Diagnostic. Thus, we need to construct measures of the SCI at the substate and 3 digit-zip code level to match with the level of our outcome variables. To do this we follow the aggregation proposed by Bailey et al. (2021*b*).

For the aggregation of the county level SCI to the substate level SCI, let us index the counties in each substate i by $r_i \in R(i)$, let $Friendships_{ri,rj}$ count the total number of friendship links between individuals in counties r_i and r_j, let Pop_{ri}, denote the total (Facebook) population in county r_i, and denote $PopShare_{ri}$, denote the share of that population

[13]The terminology has changed over time: the measure that is called the Social Connectedness Index in the public release data of 2021, was called the relative probability of friendship by Bailey et al. (2018*b*).

in county r_i in substate i: $\sum_{ri \in R(i)} PopShare_{ri} = 1$. Then the SCI between substate i and substate j is given by:

$$\begin{aligned} SCI_{i,j} = \frac{Friendships_{i,j}}{Pop_i \times Pop_j} &= \frac{\sum_{ri \in R(i)} \sum_{rj \in R(j)} Friendships_{ri,rj}}{\left(\sum_{ri \in R(i)} Pop_{ri}\right) \times \left(\sum_{rj \in R(i)} Pop_{rj}\right)} \\ &= \sum_{ri \in R(i)} \sum_{rj \in R(j)} \frac{Pop_{r_i}}{\sum_{ri \in R(i)} Pop_{ri}} \frac{Pop_{rj}}{\sum_{rj \in R(j)} Pop_{rj}} \frac{Friendships_{ri,rj}}{Pop_{ri} \times Pop_{rj}} \\ &= \sum_{ri \in R(i)} \sum_{rj \in R(j)} PopShare_{r_i} \times PopShare_{r_j} \times SCI_{ri,rj} \end{aligned} \tag{10}$$

Similarly, we use equation (10) to aggregate the ZCTA level SCI to the 3 digit-zip code level SCI. We assume that zip code and ZCTA are the same.

D.3 SCI between substate to state and 3-digit zipcode to state

$$SCI_{ik} = \sum_{j \in k} \frac{Population_j}{\sum_{j \in k} Population_j} \times SCI_{ij} \tag{11}$$

where $Population_j$ is the population in county j in state k and SCI_{ij} is the Social Connectedness Index between counties i and j.

D.4 Construction of locality level unemployment and income measures

We use the county-level data to construct population-weighted substate and 3-digit zip code measures. In our main sample, 235 substates are defined in terms of counties, while 15 substates are defined in terms of census tracts.[14] For these 15 substates, we use income and unemployment rate for the largest county (measured by population) in each substate.

For the 2010 3-digit ZIP Code Tabulation Areas (ZCTAs), there are 891 unique codes, and 88% of these fall into more than one county.[15] In addition, 73 of the total 3-digit ZCTA's cross state-boundaries. To address these overlapping issues, we match each ZCTA with the county with the most population overlapped.[16]

[14]Specifically, six substates in Massachusetts, five in Connecticut, and four in Rhode Island. Starting 2014-2016, NSDUH define the substates in Rhode Island in terms of census tracts, instead of counties.

[15]We exclude Puerto Rico.

[16]Out of 32,989 unique 2010 ZIP Code Tabulation Areas (ZCTAs), around 73% fall only into a single county, 21% fall into two counties, and the rest fall into more than two counties. There are 103 ZCTAs that cross two states.