Hits from the Bong: The Impact of Recreational Marijuana Dispensaries on Property Values*†

Danna Thomas
University of South Carolina

Lin Tian
INSEAD & CEPR

December 10, 2020

Abstract

We exploit a natural experiment in Washington state that randomly allocates recreational marijuana retail licenses to estimate the capitalization effects of dispensaries into property sale prices. Developing a new cross-validation procedure to define the treatment radius, we estimate difference-in-differences, triple difference, and instrumental variables models. We find statistically significant negative effects of recreational marijuana dispensaries on housing values that are relatively localized: home prices within a 0.36 mile area around a new dispensary fall by 3-4% on average, relative to control areas. We also explore increased crime near dispensaries as a possible mechanism driving depressed home prices. While we find no evidence of a general increase in crime in Seattle, WA, there is a significant increase in nuisance-related crimes in census tracts with marijuana dispensaries relative to other census tracts in Seattle.

keywords: real estate markets, local externalities, drug legalization

*Declarations of Interest: None
†Thomas - email: danna.thomas@moore.sc.edu. Tian - email: lin.tian@insead.edu.

1 Introduction

Despite an increasing trend of recreational marijuana liberalization across the United States and in other parts of the world, legalizing cannabis continues to be a contentious issue.[1] While there is general consensus on the benefits of legalization, such as cannabis tax revenues and decreased incarceration for drug-related crimes, strong reservations remain on the *local-level* impacts of marijuana businesses to neighborhoods, evidenced by the widespread city-level restrictions of dispensaries as well as neighborhood resistance to dispensary entry within states that have legalized marijuana.[2] Understanding the local-level consequences of marijuana dispensaries on neighborhoods is, therefore, crucial for assessing the aggregate effects of legalization and designing effective public policies to address the localized impact of the legalization.

This paper studies local responses to marijuana dispensaries, exploring how dispensary entry is capitalized into local housing values. If residents perceive a nearby marijuana dispensary as a disamenity (or as an amenity), they can "vote with their feet;" hence, the opening of a cannabis retailer should lead to a decrease (or increase) in property values, reflecting the residents' willingness to pay to live away (or near) the retailer.[3]

However, the endogeneity of dispensary location creates a challenge in identifying the causal effects of cannabis retailers on neighborhood property values: There may be variables unobserved by the econometrician but observed by the cannabis retailers that are correlated with neighborhood outcomes. To overcome this identification problem, we exploit a natural experiment in Washington that randomly allocates recreational marijuana retail licenses to applicants. Following the 2012 legalization of recreational marijuana in Washington, cannabis license applicants were required to provide potential dispensary sites on their applications, and many retail licenses were allocated via lottery. This enables us to assemble a novel data set that connects license lottery winners, losers, and cannabis

[1] Caulkins et al. (2016) offers a comprehensive summary of the issues and much of the existing academic literature.

[2] Cities such as Pasco, WA and Compton, CA have banned recreational marijuana dispensaries. For an example of neighborhood-level resistance in San Francisco, California: "Residents fighting to keep marijuana dispensary out of sunset district neighborhood," *kron4.com*.

[3] This approach to measuring the capitalization of (dis)amenities has been taken in many papers across economics subfields. For example, in the education literature, Figlio and Lucas (2004) and Black (1999) study the impact of school quality on housing markets. In the environmental economics literature, examples include Chay and Greenstone (2005) on the impact of the Clean Air Act, Currie et al. (2015) on toxic planting openings and closings, Davis (2005) on cancer clusters, Davis (2011) on power plants, and Greenstone and Gallagher (2008) on hazardous waste.

retailers to nearby property sales in Washington state. Sites that "lost" the license lottery provide a natural comparison group to control for unobservables related to store site choice, thus alleviating the endogeneity concern. Moreover, we control for other neighborhood-level unobservables in the style of Linden and Rockoff (2008), treating properties within the same neighborhood but farther away from the dispensary as a control group. This approach allows us to estimate difference-in-differences and triple-difference models. Further, as the license lottery is a plausible source of exogenous variation in dispensary location, our setting provides a natural instrumental variables framework where we use the addresses in the applications of license winners as an instrument for the actual marijuana dispensaries' locations.

To complement our empirical strategy, we propose a new cross-validation procedure to define the treatment group. One potential concern in studying how amenities impact nearby neighborhoods is determining what constitutes "nearby." In many studies of how (dis)amenities affect property values, researchers often focus their analysis on properties within concentric rings around the (dis)amenity ("ring method").[4] The inner ring defines the treatment group while the outer ring defines the control group. No standardized method of selecting the radius of the treatment ring exists to our knowledge, leaving this important choice to the researcher.

As a result, the radii of the rings are generally chosen in a somewhat subjective manner even though an arbitrary choice of radius may influence the results. For example, if the treatment effect is decreasing in distance, increasing the radius from the (dis)amenity may decrease the magnitude of the estimate, washing out any promising results. On the other hand, varying the radius changes the number of observations in the control and treatment groups, which may affect the precision of the estimates. To address this issue, Diamond and McQuade (2019) develop a non-parametric difference-in-differences estimator. Complementary to their approach, we propose an easy-to-implement, data-driven procedure—a leave-one-out cross validation—to select the optimal radius with which to conduct the analysis. Our cross validation procedure balances the trade-off between precision and changes in magnitude of the treatment effect estimate.

The cross validation procedure yields an optimal radius of 0.36 miles, i.e., property sales that took place within 0.36 miles of a marijuana dispensary are classified into the treated group. We show that

[4] See studies such as Linden and Rockoff (2008), Currie et al. (2015), Muehlenbachs et al. (2015), Autor et al. (2014), Pope and Pope (2015), and Campbell et al. (2011).

sale prices within this distance of a marijuana dispensary decline: the estimated negative price impact is as low as 3.15% in our triple difference model and as high as 3.92% in our instrumental variables model. This decrease particularly effects younger, more diverse neighborhoods. For the average home sale price in our data, this translates to about a $10,100-$13,500 reduction in prices following the entry of a recreational marijuana dispensary. While this magnitude may seem substantial, our results are consistent with other studies on the impact of disamenities on property values found in the public economics literature.[5]

Nevertheless, our result is highly distinct from prior work on the effects of marijuana liberalization and property values. While Adda et al. (2014) observes that a marijuana de-penalization policy in Lambeth, London decreased borough-wide property values, other studies find that marijuana liberalization *increases* home sale prices. For example, Cheng et al. (2018) compares cities in Colorado that allow recreational marijuana businesses to those municipalities that do not, concluding that cities that legalize marijuana businesses have higher property values which the authors attribute to a possible "green boom." At a more localized level, Conklin et al. (2018) studies conversions from medical marijuana dispensaries into recreational retailers in Denver, CO, estimating that homes within 0.1 miles of a medical-retail conversion increase in value relative to those slightly farther away by 8%. Similarly, Burkhardt and Flyr (2018) examines new medical and recreational dispensary entry in Denver. Using home sales within 0.25 miles of a dispensary opening as the treatment group and properties within 0.25 miles of where a new dispensary would open in the subsequent 6-12 months as a control, they find a 7.7% increase in home sale prices.

In contrast to these previous papers, our study uses extensive data on property sales throughout the entire state of Washington and comprehensive administrative data on retailers. Therefore, we are able to compare local neighborhoods both before and after any recreational retailer entry occurs. Moreover, our research design has the advantage of exogenous license distribution statewide—through the license lottery outcomes—which provide plausible counterfactual locations where no retailer enters. These differences may explain the divergence with previous estimates.

A possible driver of the estimated negative price impact is that communities may perceive that marijuana dispensaries cause crime. For instance, because marijuana is still federally illegal, cannabis

[5] We discuss the comparison of our findings with other estimates in the literature in Section 5.

businesses typically do not have access to banks and, consequently, are cash-only, making them possible targets for robbers.[6] Nonetheless, evidence on the relationship between on local crimes in areas near marijuana dispensaries is mixed.[7] For example, Freisthler et al. (2016) uncovers a positive correlation between dispensary density and violent crime in Long Beach, California while Kepple and Freisthler (2012) does not detect an association between medical dispensary density and violent crime Sacramento. A number of other papers have found correlations between dispensary density and child neglect, marijuana abuse, and youth usage (Freisthler et al., 2015; Mair et al., 2015; Shi, 2016). Using a quasi-experimental approach, Chang and Jacobson (2017) identifies temporary increases in crimes, particularly property crime, during temporary dispensary closures in Los Angeles likely due to fewer "eyes-on-the-street." Brinkman and Mok-Lamme (2019) uses an instrumental variables approach to establish that dispensaries in Denver, CO decrease crime in the census tracts where they are located.

We add to these quasi-experimental approaches by utilizing data on police reports in Seattle, WA. Leveraging the natural experiment setting from the license distribution lottery, we use the lottery results as an instrument for dispensary location in Seattle census tracts. We estimate that overall crime reports decrease by 13.4 per 10,000 residents though our estimate is not statistically significant at 10%. Despite this, when we analyze categories of crime, we find evidence that dispensary entry increases the number of nuisance crime related reports (e.g., disorderly conduct, loitering) by 4.2 per 10,000 residents but decreases the number of drug-related reports by 2.8 per 10,000 residents. Moreover, we also find that nuisance crime reports and violent crime reports increase in adjoining census tracts by 1.8 and 2.5 per 10,000 residents, respectively. Increased nuisance related crime, therefore, may be one contributing factor to depressed home prices in areas near dispensaries.

The paper proceeds as follows. Section 2 offers details about the setting of our empirical exercise in Washington state. Our data and methodology are described in Sections 3 and 4, respectively. Section 5 details the results, and the model and results for crime are discussed in Section 6. Section 7 concludes.

---

[6] Abcarian, Robin. "Your Business is Legal, but You Can't Use Banks. Welcome to the Cannabis All-Cash Nightmare." *Los Angeles Times*, January 29, 2017. `http://www.latimes.com/local/abcarian/la-me-abcarian-cannabis-cash-20170129-story.html`

[7] Studies of marijuana liberalization laws on state-wide crime levels include Lu et al. (2019), Huber III et al. (2008), Anderson et al. (2013), and Anderson et al. (2015). Adda et al. (2014) studies the effects de-penalization of marijuana of bourough-wide crime in Lambeth, London.

2 Background and Institutional Details

2.1 Initiative-502

On November 6, 2012, by a statewide vote of 55.7 percent to 44.3 percent, Washington state voters approved Initiative-502 (I-502), legalizing the possession and consumption of cannabis for adults over twenty-one years of age as well as the production and sale of marijuana in businesses regulated by the state government.[8,9] In order for firms to participate in the legalized recreational marijuana market, I-502 stipulated that a business must hold either a producer (marijuana farmers), processor (creators of joints, edibles, vapor products, etc.), or retailer license. Further, the law allowed state regulators to restrict the number of licenses it issued.

While the state's cannabis market regulator—the Washington Liquor Cannabis Board (WLCB)—opted not to limit the number of licenses issued to upstream firms such as farmers and processors, it capped the number of retail licenses state-wide at 334. It then divided up these licenses among counties using a formula that calculated "the number...[by] minimiz[ing] the population-weighted average" distance from the user to the marijuana retailer.[10,11]

The licenses were then split across the county's incorporated cities according to the proportion of the county's population within the city. The remaining licenses were assigned to the county's rural areas. For example, King County was allocated sixty-one retail licenses to be spread across seventeen incorporated cities and rural King County. Bellevue, which contains about 6% of King County's population, was assigned four; and Seattle, which has around a third of the county's population, was assigned twenty-one. Table 1 provides a detailed breakdown of the number of licenses for each jurisdiction.

[8] Initiative Measure No. 502, Session 2011 (WA 2011)

[9] "Initiative Measure No. 502 Concerns marijuana - County Results," Washington Secretary of State, November 27, 2012, https://results.vote.wa.gov/results/20121106/Initiative-Measure-No-502-Concerns-marijuana_ByCounty.html.

[10] To calculate this average distance, the formula assumes users are spread uniformly across the state and that "stores are placed... to maximize convenience." Hence, the "proxy" for distance is the area of a county divided by the number of stores in the county.

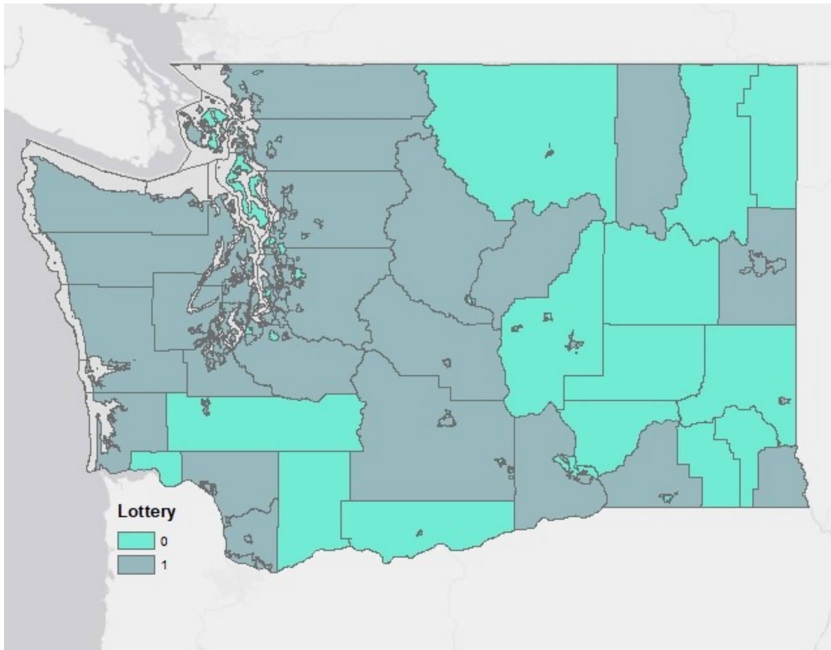
[11] Caulkins, Jonathan P. and Linden Dahlkemper, "Retail Store Allocation," BOTEC Analysis Corporation, Jun. 28, 2013, available from Washington Liquor Cannabis Board.

Table 1: Schedule of License Quotas

City/Locality	# Licenses	# Apps	# Entrants	City/Locality	# Licenses	# Apps	# Entrants	City/Locality	# Licenses	# Apps	# Entrants
Aberdeen*	1	2	1	Island County	3	3	3	Port Townsend*	1	4	1
Adams County	2	0	0	Issaquah*	1	10	1	Pullman*	3	14	3
Anacortes*	1	3	1	Jefferson County*	3	5	2	Puyallup	2	2	0
Arlington*	1	3	1	Kelso	1	0	0	Quincy	1	0	0
Asotin County*	2	3	2	Kennewick*	4	5	0	Redmond*	2	3	0
Auburn*	2	5	1	Kent*	3	9	0	Renton*	3	9	2
Bainbridge Island*	1	3	1	King County*	11	44	10	Richland	3	0	0
Battle Ground*	1	2	1	Kirkland*	2	12	2	Sammamish	1	0	0
Bellevue*	4	19	3	Kitsap County*	7	40	4	San Juan Island*	1	13	0
Bellingham*	6	27	7	Kittitas County*	2	4	0	SeaTac*	1	2	0
Benton County*	2	3	2	Klickitat County	3	3	2	Seattle*	21	191	21
Bonney Lake*	1	2	0	Lacey*	2	9	2	Sedro-Woolley	1	0	0
Bothell	1	1	1	Lake Stevens*	1	2	1	Selah	1	0	0
Bremerton*	2	16	3	Lakewood*	2	4	2	Sequim*	1	5	0
Burien	1	0	0	Lewis County	4	4	0	Shelton	1	1	1
Burlington	1	1	0	Lincoln County	2	0	0	Shoreline*	2	8	2
Camas*	1	8	0	Longview*	3	10	3	Skagit County*	4	17	1
Centralia*	2	4	0	Lopez Island	1	0	0	Skamania County	2	2	1
Chehalis*	1	2	1	Lynden	1	0	0	Snohomish County*	16	87	12
Chelan County*	3	7	3	Lynnwood	2	2	0	Spokane*	8	58	6
Clallam County*	3	12	3	Maple Valley*	1	3	0	Spokane County*	7	13	6
Clark County*	6	22	1	Marysville*	3	9	0	Spokane Valley*	3	27	3
Columbia County	1	0	0	Mason County*	4	11	3	Stevens County	4	3	4
Cowlitz County*	3	8	2	Mercer Island	1	1	0	Sunnyside	1	1	1
Des Moines*	1	2	1	Mill Creek	1	1	0	Tacoma*	8	44	9
Douglas County*	2	6	1	Monroe	1	0	0	Thurston County*	6	19	6
East Wenatchee*	1	3	1	Moses Lake*	2	5	2	Tukwila*	1	4	0
Edmonds	2	2	0	Mount Vernon*	3	5	2	Tumwater*	1	7	1
Ellensburg*	2	8	2	Mountlake Terrace*	1	16	0	University Place	1	0	0
Ephrata	1	1	1	Mukilteo	1	1	0	Vancouver*	6	48	6
Everett*	5	27	4	Oak Harbor	1	1	1	Wahkiakum County	1	0	0
Federal Way*	3	15	0	Ocean Shores*	1	3	1	Walla Walla	2	2	1
Ferndale	1	1	1	Okanogan County	4	3	2	Walla Walla County*	2	3	1
Ferry County*	1	3	1	Olympia*	2	9	3	Washougal*	1	6	0
Franklin County	1	0	0	Omak	1	1	1	Wenatchee	3	2	2
Garfield County	1	0	0	Orcas Island	1	1	1	West Richland	1	1	0
Goldendale	1	1	1	Pacific County*	2	19	2	Whatcom County*	7	15	6
Grandview	1	0	0	Pasco	4	3	1	Whitman County	1	0	0
Grant County	3	2	2	Pend Oreille County	2	1	0	Yakima	5	7	1
Grays Harbor County*	3	7	3	Pierce County*	17	45	7	Yakima County*	6	10	3
Hoquiam*	1	2	1	Port Angeles*	2	8	2	**State Total**	**334**	**1173**	**212**

Notes: Jurisdictions with * are those in which licenses are allocated via lottery.

Figure 1: Distribution of Jurisdictions in Washington



Notes: The blue-green shaded areas are jurisdictions where the recreational marijuana license quota was not binding, i.e., there are more licenses than the number of applications. The blue-grey areas are jurisdictions where the recreational marijuana license quota was binding and therefore a lottery to allocate the license was carried out within the jurisdiction.

2.2 The Washington Marijuana Retail License Lottery

In November 2013, the WLCB opened a thirty day window during which potential marijuana retailers could apply for a retailer license. Applicants were subject to background checks to determine if they were eligible licensees. Furthermore, as stores were banned from locating within 1000 feet of a "school, playground, ... child care center, public park, public transit center, or library", the license applications required a potential store address so that regulators could determine compliance with the location restrictions.[12]

Prospective firms could submit multiple applications for multiple licenses. However, the state imposed restrictions on the number a firm could obtain: A business could not have more than three licenses and no more than a third of all stores in a jurisdiction. Moreover, while the application fee was a nominal $250, many businesses did not submit several applications. In fact, 99 percent of the 802 individual applicants turned in less than three applications with 68 percent submitting just one,

[12]Initiative Measure No. 502, Session 2011 (WA 2011).

and 51 percent of firms that handed in multiple applications were not petitioning for licenses in the same jurisdiction.

In total 1,173 applications were submitted. Table 1 lists the number of applicants and available licenses for each jurisdiction as well as the number of dispensaries in the market by February 2016. Seventy-five jurisdictions had more applicants than licenses, and the WLCB decided to distribute licenses for these areas via a lottery.[13] Figure 1 highlights which jurisdictions allocated licenses via lottery.

The license lotteries were held April 21-25, 2014. Each applicant in a lottery was randomly assigned a number by the accounting firm Kraght-Snell. The numbers—without any identifying information—were then sent to Washington State University's Social and Economic Sciences Research Center, which ranked the numbers from 1 to n, with n being the number of applicants within a jurisdiction. Kraght-Snell then decoded the rankings. If a ranking number was higher than the number of licenses allocated to a jurisdiction, the firm was a lottery "winner." The results of the lottery, made public on May 2, 2014, were well-publicized in the state and local press.[14]

2.3 Entry in the Recreational Marijuana Market

Contingent on receiving a license, licensees could begin selling marijuana as early as July 2014. Figure 2 shows the evolution of store entry over time. Seventy percent of lottery winners entered the market, and half of those that did not enter (i.e., 15 percent of the lottery winners) were kept from opening due to local bans on marijuana businesses.[15] We cannot account for the other half of lottery winners that did not enter the market though one possibility is that potential firm owners failed subsequent background checks. However, if a lottery winner could not enter due to a failed background check, etc., Washington awarded the license to the next applicant in the lottery ranking.[16]

Importantly, even though application addresses were not legally binding, thirty-six percent of firms in our data located in the exact area specified on their application. Those entrants that did not open at their application address generally moved to locations found on other applications. Specifically,

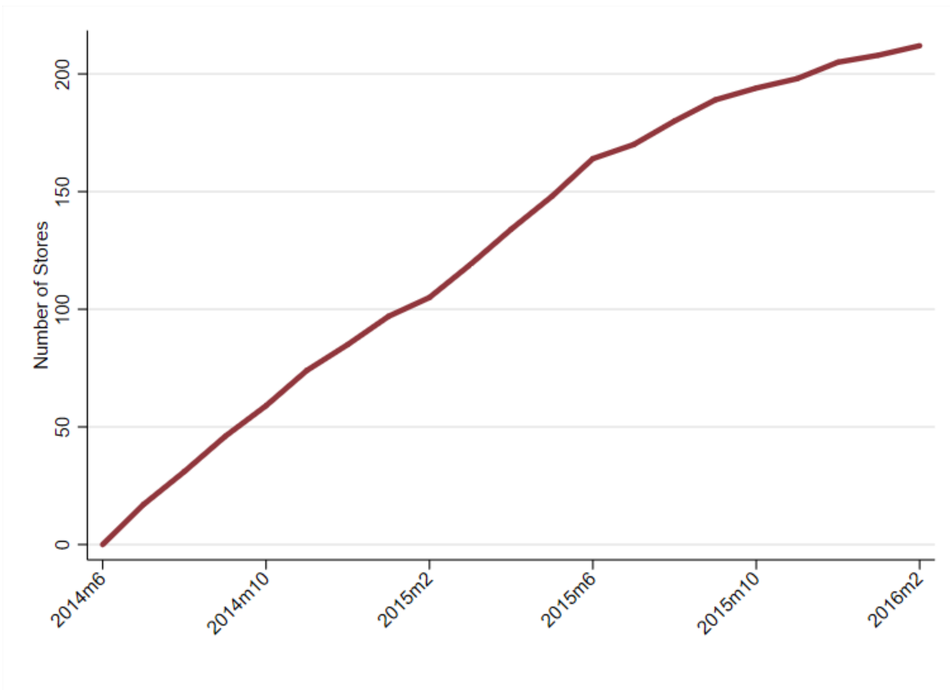
[13]The state did not reallocate the licenses of those jurisdictions whose quota did not bind to other parts of the state which de facto capped the number of licenses at less than 334.

[14]See "Who Won the Pot Shop Jackpot and Where the Stores Might Be," *The Seattle Times*, May 2, 2014.

[15]The WLCB did not reallocate licenses in areas with bans to other jurisdictions.

[16]Another possible explanation for failure to enter is the potential firm holding the license for speculative purposes.

Figure 2: Entry of Recreational Marijuana Dispensaries Over Time



Notes: The figure shows the cumulative number of recreational marijuana dispensaries in Washington state between June 2014 and February 2016.

seventy-five percent of entrants are within one-third of a mile from some lottery application address. This suggests that addresses submitted for the lottery constitute meaningful information and can serve as a good predictor for the actual dispensary location—a fact that is crucial for our empirical strategies.

3 Data

We assemble a novel data set from a variety of sources. For our primary analysis, we first use the results of the marijuana retail license lottery, provided by the WLCB through public records request. The data includes the applicant's tradename, application number, lottery rank, and address for the potential store's location. The WLCB also provided data on jurisdictions' license allocations which enables us to identify lottery "winners"—those applicants whose lottery rank is less than or equal to the number of licenses allocated to the jurisdiction.[17]

Next, we combine the lottery results with information on operating cannabis retailers, allowing us to view which lottery winners eventually entered the market. The retailer data comes from the WLCB's Traceability System, a system that allows the state to track marijuana products through the cannabis supply chain until the products are sold by dispensaries. The data includes retailers' addresses and sales for each day. As not all dispensaries opened on July 2014—the first month of legal sales—we construct the firms' entry dates by defining the entry date as the day of first sale. Figure 3a displays the geographic distribution of retailers and license applicants across Washington.

Our housing sales data is supplied by CoreLogic (previously named DataQuick), a private vendor. The data is generated from public records created by tax assessors as well as from proprietary records created by multiple listing services. While the data set includes properties from all across Washington, as Figure 3b shows, the coverage is the best along the Interstate 5 corridor, the most populous area of the state. The data includes comprehensive housing characteristics for each property—e.g., square footage, year built, whether the property is a single or multi-family home or condo, and address–and information on the property transaction—e.g., sale date and sale price (which we convert to January 2014 dollars).

[17]We also include as winners those applicants that initially just missed the lottery cutoff but were awarded licenses due to failed background checks by higher ranked applicants.

We geocode all addresses using the commercial geocoder *Here*, keeping only those property addresses that are identified with high accuracy at the address number level. We then calculate the geodesic distance from properties to retailers and lottery participant addresses, dropping properties greater than one mile away from a lottery address or retail dispensary. Figure 4 provides a pictorial example of how we construct the data sample. As we are interested in finding comparable neighborhoods that are attractive to retail applicants but where some firms enter and others do not, we limit the sample to the 75 jurisdictions that allocated retailer licenses via lottery and restrict the data to those properties within one mile of a lottery participant or dispensary address. We also narrow the sample used in our analysis to home sales occurring from January 2012 to February 2016, as Washington state expands the number of recreational marijuana licenses in 2016. Consistent with previous work such as Linden and Rockoff (2008), we also drop sales outside the range of $30,714 to $1,482,382 (i.e., keeping only property transactions with sales in the 1st percentile to 99th percentile in the price distribution).

For additional information about neighborhoods, we connect property locations to their census tracts using the TIGER/Line shapefiles from the U.S. Census Bureau. The shapefiles are linked to the 2014 American Community Survey (ACS) which allows us to incorporate tract-level demographic data into the sample. These data include tract population, median income, and binned counts of age, education level, and race. Using this information, we compute the percentage of individuals in the census tract that are high school and college graduates, the percentage of non-Hispanic white individuals in the census tract, as well as the percentage of individuals in the census tract between the ages of 18 and 35.[18]

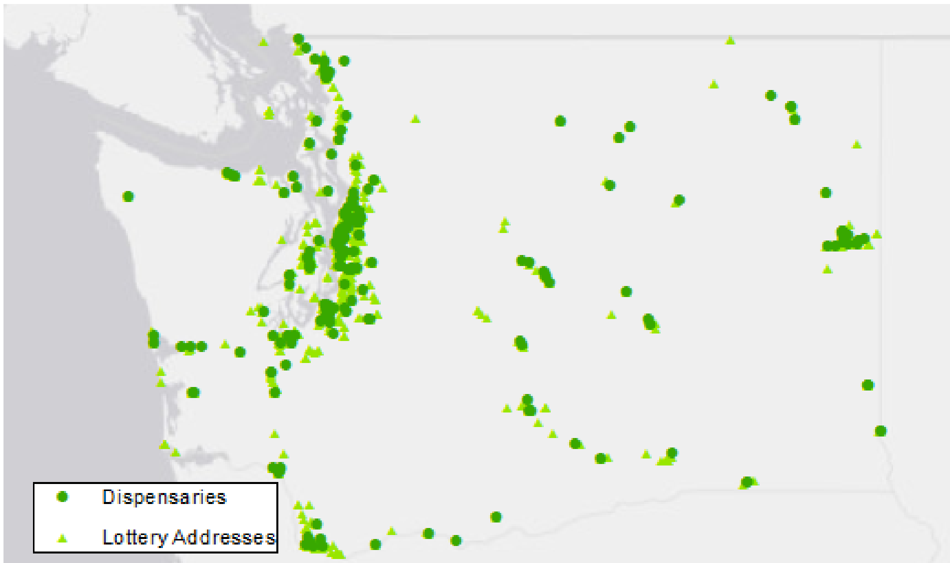
We also link properties to the I-502 referendum results at the properties' precincts using 2012 election results and precinct shapefiles obtained from the Washington Secretary of State. Finally, we calculate the distance between dispensaries and the closest public school using data on the location of schools from the Washington Department of Education.

Overall, the data consists 83,006 records from 25 counties in Washington, 172 stores, and 869 lottery applications.[19] Table 2 provides the overall summary statistics of the data. Properties within

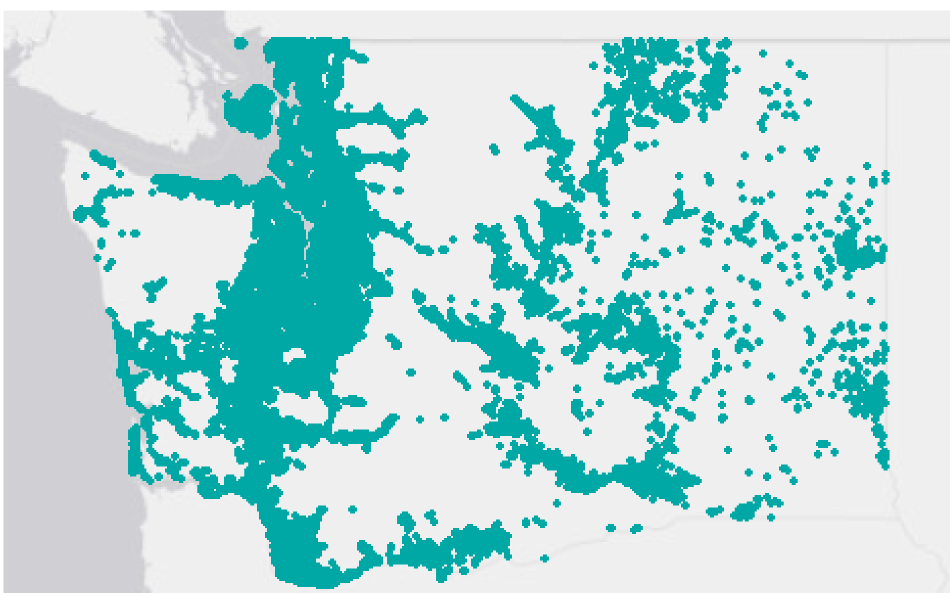
[18] Azofeifa et al. (2016) finds that people under thirty-five are primary users of marijuana.

[19] Top panel of Figure A.1 in the Appendix shows the number of observations by distance to dispensary and lottery addresses.

Figure 3: Geographic Distribution of Retailers and Properties



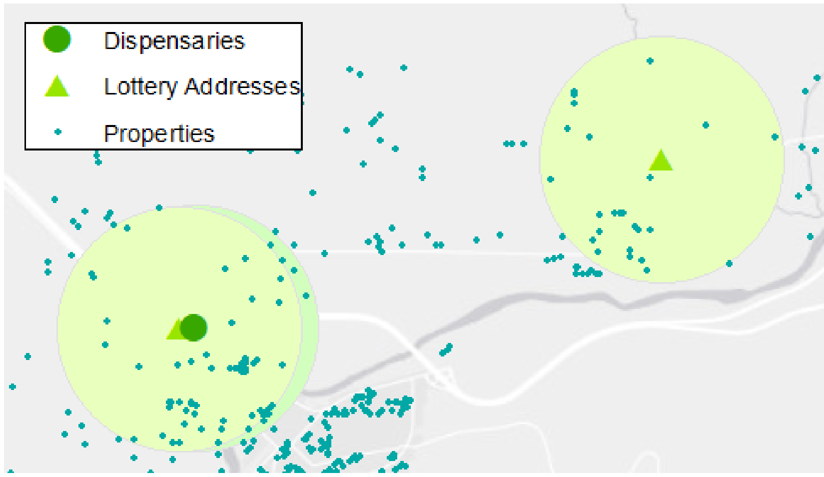
(a) Dispensaries



(b) Properties

Notes: In Figure 3a each dot represents a single dispensary that was open from July 2014–February 2016 while in Figure 3b each dot represents a single property in the Corelogic data set that was sold from July 2014–February 2016.

Figure 4: Dispensaries, Lottery Addresses, and Surrounding Neighborhoods



Notes: Buffers around dispensaries and lottery addresses have a radius of one mile.

one mile of a dispensary tend to be less expensive, older, and slightly smaller, while the neighborhoods tend to be slightly younger, slightly more diverse, and less wealthy. This pattern also repeats itself when comparing areas around "winning" addresses versus "losing" addresses. This could be suggestive of selection of stores into neighborhoods. We discuss our strategies to identify causal effects in spite of selection in Section 4.

4 Empirical Methodology

As is well understood, a major challenge in identifying the capitalization of local (dis)amenities such as cannabis retailers into housing prices is that variation in the amenity is rarely exogenous and is likely correlated with factors which affect location choices of the amenity but are unobserved by the econometrician. For example, in our setting, dispensaries will likely open in geographic locations that already have high marijuana demand. If latent cannabis demand is correlated with housing prices, then estimates of the effects of dispensary entry will be biased.

Specifically, there are two selection issues that need to be addressed: (1) selection of neighborhoods; and (2) selection of exact locations within a given neighborhood. In what follows, we consider three different empirical strategies and discuss how each of them deals with these selection issues. In particular, we describe how our unique setting allows us to identify the treatment effect under a less

Table 2: Summary Statistics

Panel A: Property and Neighborhood Characteristics

	All Properties	Entrants	No Entrants	Winners	Losers
Closing Price	324,095	313,347	341,208	308,008	348,733
	(219,822)	(215,255)	(225,855)	(211,227)	(230,204)
Beds	2.886	2.89	2.881	2.898	2.869
	(1.018)	(1.013)	(1.026)	(.9784)	(1.076)
Baths	1.803	1.785	1.831	1.794	1.817
	(.7682)	(.771)	(.7628)	(.7671)	(.7697)
Home Age	42.84	43.78	41.34	43.08	42.48
	(34)	(34.45)	(33.2)	(33.84)	(34.23)
Square Footage	1,666	1,657	1,680	1,647	1,696
	(847.1)	(894.2)	(765.9)	(698.8)	(1,033)
% 18-24 y.o.	.2859	.2964	.2693	.2804	.2944
	(.106)	(.1097)	(.0974)	(.0979)	(.1167)
% White	.6797	.6717	.6924	.679	.6807
	(.1558)	(.1615)	(.1453)	(.1567)	(.1543)
% H.S. Grad.	.6426	.6362	.6529	.6366	.6517
	(.1011)	(.1012)	(.0999)	(.0979)	(.1051)
Median Income	63,491	61,739	66,280	62,506	64,999
	(21,612)	(21,726)	(21,132)	(21,524)	(21,660)
% Yes I-502	.6286	.628	.6296	.6216	.6394
	(.107)	(.1101)	(.1018)	(.1023)	(.113)
Miles to Dispensary	.6133	.6214	.6002	.6034	.6284
	(.2499)	(.2433)	(.2596)	(.2444)	(.2573)
Miles to Nearest	.4649	.4515	.4861	.4695	.4579
School	(.2886)	(.2735)	(.3099)	(.293)	(.2816)
Miles to Medical	6.916	6.186	8.078	6.129	8.12
Dispensary	(15.58)	(15.03)	(16.36)	(14.24)	(17.38)
% Condo	.2174	.2128	.2248	.2097	.2292
% Multi-Family	.0303	.0348	.0232	.0318	.0281
% Single-Family		.7524	.7519	.7585	.7427
Observations	83,006	50,985	32,021	50,218	32,788

Panel B: Applicants and Entrants

	Entrants	Winners	Losers
Number of Firms	172	197	672
Percentage of Applications	.153	.175	.597

restrictive set of assumptions compared to conventional literature.

4.1 Difference-in-Differences (DD) Method

A popular strategy to correct for selection, commonly referred to as the "ring approach," is to compare properties close—within an inner ring of r mile radius–to a (dis)amenity to properties that are slightly farther away—within an outer ring of $R > r$ mile radius. Identification relies on the assumption that nearby properties are the most likely to be adversely or positively affected by the (dis)amenities while properties slightly farther away should share many characteristics and price trends of the nearby properties but be less impacted by localized effects, making those properties a desirable control group.

To elaborate further, we adopt an expository style similar to that of Muehlenbachs et al. (2015). We define the change in a property's price after dispensary entry as ΔP_k^j with $j = \mathbb{1}(d \leq R)$ and $k = \mathbb{1}(d \leq r)$, and the variable d as the distance to the closest cannabis dispensary.[20] For ΔP_k^j,

$$\begin{aligned} \Delta P_0^1 &= \Delta Neighborhood + \Delta Macro, \\ \Delta P_1^1 &= \Delta Dispensary + \Delta Neighborhood + \Delta Macro. \end{aligned}$$

Price changes can be driven by macroeconomic price trends ($\Delta Macro$) as well as neighborhood-level price trends ($\Delta Neighborhood$). However, changes due to store entry ($\Delta Dispensary$) impact only the closest properties. Hence, the difference-in-differences (DD) calculation identifies the effect of dispensary entry:

$$\Delta Dispensary = \Delta P_1^1 - \Delta P_0^1. \tag{1}$$

Adopting this approach to our context motivates the following DD regression on all properties within R miles of a dispensary or lottery address:

$$\ln(p_{ijt}) = \beta_0 + \beta_1 D_{it}^r + \beta_2 D_i^r + \beta_3 D_{it}^R + \beta_4 D_i^R + \beta_5 Post_{it} + \beta_6 X_{ijt} + \varepsilon_{ijt}, \tag{2}$$

[20] ΔP_0^1 (ΔP_1^1) thus denotes the change in price for a property in the outer (respectively, inner) ring.

where p_{ijt} is the sale price of observation i in area j at time t.[21] The variable $Post_{it}$ is an indicator function that equals one after the announcement of the license winners, consistent with a model of forward-looking consumers. The variables D_i^r and D_{it}^r are indicator functions with $D_i^r = \mathbb{1}(d_i \leq r)$, and $D_{it}^r = \mathbb{1}(d_i \leq r) \cdot Post_{it}$, where d_i is the distance to the closest dispensary. Similarly, $D_i^R = \mathbb{1}(d_i \leq R)$ and $D_{it}^R = \mathbb{1}(d_i \leq R) \cdot Post_{it}$. The vector X_{ijt} is a vector of property characteristics (number of bedrooms and bathrooms, home age, log square footage, property type—e.g.condo, single family) and census tract characteristics (median tract income, percentage of high school graduates, percentage of individuals between 18 and 35, percentage of the tract population that is non-Hispanic white), the percentage of yes votes in the property's precinct for I-502 in the 2012 referendum, quarter-year fixed effects, and area (city or zipcode-city) fixed effects. The coefficient of interest is β_1, the treatment effect of dispensary entry.

The identification assumption for the DD approach is that housing prices close to the marijuana dispensaries would have trended similarly to house prices farther away. While the specification controls for many neighborhood-level unobservables that effect marijuana demand, there could still exist systematic unobserved price trends between properties near dispensaries and properties farther away that are not adequately captured. In other words, the DD approach addresses potential selection of neighborhoods, but not within-neighborhood selection. If there is selection of the exact geographic location of the marijuana dispensaries within a neighborhood, then the DD estimates will be biased.

4.2 Addressing Within-Neighborhood Location Selection

The conventional DD approach is not adequate to address both neighborhood selection and location selection within neighborhoods. However, our setting in Washington allows for us to overcome both identification challenges. Specifically, we can account for the differences between properties near and farther away for dispensaries because we observe properties in our data that were located in a location attractive to marijuana firms but miss out on being close to a dispensary due to the license quota. We discuss two different empirical strategies both leveraging the unique setting in Washington

[21] We keep all properties next to a dispensary or lottery address for the DD analysis, including those in neighborhoods with at least one lottery application but no dispensary entries, which are not in the treatment or control group in this model. We prefer this approach for two reasons: (i), it keeps the number of observations the same for the DD and the DDD approaches, thereby facilitating a direct comparison of results; and (ii), the additional samples help to identify coefficients on the property and neighborhood controls, i.e., β_6. Results keeping only observations by dispensaries are very similar to the model that includes all properties and are available upon request.

marijuana market, and examine their respective strengths and weaknesses.

Difference-in-Difference-in-Differences (DDD)

The first approach extends the DD method by introducing an additional control group to account for within-neighborhood differences—properties within neighborhoods that contain a lottery address. These potential dispensary sites reflect the desirability of a particular location to a dispensary operator, which may be correlated with prices trends for the surrounding properties.

To further illustrate our approach, we now denote the change in a property's price after dispensary entry as $\Delta P_{k,n}^{j,m}$ with j and k defined as in Section 4.1, and $m = \mathbb{1}(\tilde{d} \leq R)$ and $n = \mathbb{1}(\tilde{d} \leq r)$. The variable $\tilde{d}$ is the distance to the closest dispensary or lottery address.[22,23] For $\Delta P_{k,n}^{j,m}$,

$$\begin{aligned}
\Delta P_{0,0}^{1,1} &= \Delta Neighborhood + \Delta Macro, \\
\Delta P_{1,1}^{1,1} &= \Delta Dispensary + \delta + \Delta Neighborhood + \Delta Macro, \\
\Delta P_{0,1}^{0,1} &= \delta + \Delta Neighborhood + \Delta Macro \\
\Delta P_{0,0}^{0,1} &= \Delta Neighborhood + \Delta Macro,
\end{aligned}$$

where δ denotes within-neighborhood location selection, i.e., price trends common to areas that are/would be close to marijuana dispensaries. (In the DD approach, δ is assumed to be 0.) A triple difference identifies the treatment effect, i.e., $\Delta Dispensary$.

$$\Delta Dispensary = (\Delta P_{1,1}^{1,1} - \Delta P_{0,0}^{1,1}) - (\Delta P_{0,1}^{0,1} - \Delta P_{0,0}^{0,1}). \tag{3}$$

This motivates the following triple-difference estimating equation:

$$\ln(p_{ijt}) = \alpha_0 + \alpha_1 D_{it}^r + \alpha_2 D_i^r + \alpha_3 D_{it}^R + \alpha_4 D_i^R + \alpha_5 \tilde{D}_{it}^r + \alpha_6 \tilde{D}_i^r + \alpha_7 Post_{it} + \alpha_8 X_{ijt} + u_{ijt}, \tag{4}$$

where α_1 is the triple-difference estimate. As in Section 4.1, X_{ijt} is a vector of property and census

[22] Using all lottery addresses or only lottery "losers" does not result in a statistically significant change in our estimates.

[23] In order to keep sharp distinctions between treatment and control groups, we exclude properties that would fit into both control groups in our analysis. These are those close to lottery addresses and within one mile of a dispensary, but are more than 0.36 miles from a dispensary, i.e., $\mathbb{1}(d_i \leq R) \cdot \mathbb{1}(\tilde{d}_i \leq r) = 1$ but $\mathbb{1}(d_i \leq r) = 0$.

tract characteristics, the precinct-level 2012 referendum results, as well as quarter-year and city or zipcode-city fixed effects, and the indicator function $Post_{it}$ is equal to one after the announcement of the license winners. The variables D_i^r and D_{it}^r are indicator functions with $D_i^r = \mathbb{1}(d_i \leq r)$ and $D_{it}^r = \mathbb{1}(d_i \leq r) \cdot Post_{it}$, where d_i is the distance to the closest dispensary. D_i^R and D_{it}^R are defined analogously. The variables $\tilde{D}_i^r$ and $\tilde{D}_{it}^r$ are a dummy variables with $\tilde{D}_i^r = \mathbb{1}(\tilde{d}_i \leq r)$ and $\tilde{D}_{it}^r = \mathbb{1}(\tilde{d}_i \leq r) \cdot Post_{it}$. Specifically, $\tilde{d}_i$ is i's distance from the closest dispensary or lottery address, so $\tilde{D}_i^r$ equals one if the property is within r miles of a dispensary or lottery address. $\tilde{D}_i^R$ and $\tilde{D}_{it}^R$ are similarly defined.

Instrumental Variables (IV)

We also consider a related but distinct empirical strategy in the spirit of Chaisemartin and D'Haultfoeuille (2018) that leverages the license lottery outcomes. The license lottery generates plausibly exogenous variation, which we use to construct two instrumental variables that assign neighborhoods into treatment and control areas, as well as properties within a neighborhood into treatment and control sites.

We define the dummy variable denoting those properties randomly assigned to treatment by the lottery (i.e. properties within r miles of a lottery winner), $G_i^r = \mathbb{1}(d_i^W \leq r)$, along with an indicator of those properties assigned to a treatment neighborhood, $G_i^R = \mathbb{1}(d_i^W \leq R)$, where the variable d_i^W is the distance to the closest "winning" lottery address. The corresponding post-lottery variables are $G_{it}^R = \mathbb{1}(d_i^W \leq R) \cdot Post_{it}$ and $G_{it}^r = \mathbb{1}(d_i^W \leq r) \cdot Post_{it}$, respectively.

Realized treatment after dispensary entry is defined as before with $D_{it}^r = \mathbb{1}(d_i \leq r) \cdot Post_{it}$ where $Post_{it}$ is set equal to one after the lottery winners are announced. D_{it}^R is defined analogously. Because dispensaries can locate away from their listed lottery address, D_{it}^r is not random, and $D_{it}^r \neq G_{it}^r$. However, being close to a winning address, i.e., $G_{it}^r = 1$, is a strong predictor for being treated, i.e., $D_{it}^r = 1$, so we use G_{it}^r and G_{it}^R as instruments for D_{it}^r and D_{it}^R.

Nevertheless, although lottery outcomes are random, the number of available licenses varies across cities, and many homes within the same city may be relatively close to multiple lottery addresses. In other words, some properties have a higher chance of being assigned to the treatment ($G_i^r = 1$) than others, and these properties may be systematically different from properties with a low probability of treatment. Without controlling adequately for the probability of treatment, the instrumental variables exclusion restriction will be violated.

Therefore, we construct W_i, a vector of variables that control for the probability of treatment. To do this, we start by calculating the probability that i has a lottery winner within r miles.[24] We bin these probabilities into 20 quantiles and create dummy variables for each bin. We create similar dummy variables for all properties with lottery applications within R miles and fully interact these variables with each other and with dummy variables for property type. We repeat this process for the number of license applications within r and R miles. If W_i sufficiently controls for treatment probability, G_i^R and G_i^r are random conditional on the covariates, which include, in addition to W_i, X_{ijt}—a vector of property characteristics (number of beds and baths, log square footage, property age, and property type), precinct-level I-502 referendum results, census tract characteristics (median tract income, percentage of high school graduates, percentage of individuals between 18 and 35, percentage of the tract population that is non-Hispanic white), area (city and zipcode-city) fixed effects, and quarter-year fixed effects—, and an indicator equal to one if a property is located near a license application address interacted with $Post_{it}$.

The IV model is then

$$\ln(p_{ijt}) = \tau_0 + \tau_1 D_{it}^r + \tau_2 D_{it}^R + \tau_3 G_i^r + \tau_4 G_i^R + \tau_5 Post_{it} + \tau_6 W_i + \tau_7 X_{ijt} + \nu_{ijt}, \tag{5}$$

$$D_{it}^r = \lambda_0 + \lambda_1 G_{it}^r + \lambda_2 G_{it}^R + \lambda_3 G_i^r + \lambda_4 G_i^R + \lambda_4 Post_{it} + \lambda_6 W_i + \lambda_7 X_{ijt} + \mu_{ijt}, \tag{6}$$

$$D_{it}^R = \pi_0 + \pi_1 G_{it}^r + \pi_2 G_{it}^R + \pi_3 G_i^r + \pi_4 G_i^R + \pi_5 Post_{it} + \pi_6 W_i + \pi_7 X_{ijt} + \xi_{ijt}. \tag{7}$$

We also study the intent-to-treat (ITT) model

$$\ln(p_{ijt}) = \sigma_0 + \sigma_1 G_{it}^r + \sigma_2 G_{it}^R + \sigma_3 G_i^r + \sigma_4 G_i^R + \sigma_5 Post_{it} + \sigma_6 W_i + \sigma_7 X_{ijt} + \chi_{ijt}. \tag{8}$$

The coefficients of particular interest are τ_1 in (5), the average treatment effect on the treated, and σ_1 in (8), the intent-to-treat effect.

A practical issue to address is that properties in losing neighborhoods and/or near losing addresses, which are in the instrument control group (i.e., $G_i = 0$), may become treated in the post-period when

[24] Details on calculating this probability are found in Appendix A.1.

the operators locate the dispensary at an address different from the lottery address (i.e., $D_{it} = 1$). In other words, the rate of treatment increases in the control group across the pre- and post-periods. To deal with this, we have two options. The first one is to retain the whole sample, including observations in losing neighborhoods and/or near losing addresses that ended up in treated groups. Alternatively, we can construct a time-consistent control group by dropping those observations so that the rate of treatment in the control group does not change over time. In the first option, τ_1 does not identify the average treatment on the treated unless the researcher makes the strong assumption that local average treatment effects are homogeneous between treatment and control groups (Chaisemartin and D'Haultfoeuille, 2018). To avoid making such strong assumptions, we follow the recommendation in Chaisemartin and D'Haultfoeuille (2018) and adopt the second option. Specifically, we define control groups where the distribution of treatment does not change over time, i.e., for $G_i = 0$, $G_{it} = D_{it} = 0$. Properties in losing neighborhoods and/or near losing addresses that ended up in treated groups, i.e., for $G_i = 0$, $D_i = 1$, are therefore dropped from the analysis.

Comparison of DDD and IV

Both the DDD and IV methods provide a way to control for within-neighborhood selection that the DD approach assumes away; hence, we view the two methods as complementary. First, the two methods are equivalent if there is full compliance, i.e., all lottery winners enter at the same exact locations as the addresses declared during the lottery application. However, as seen in Figure 4, though many licensees were close to their original stated addresses, stores were allowed to move locations. Indeed, as explained in Section 2.3, 30% of lottery winners chose not to enter at all. And out of the remaining 70%, 64% of them chose a different location for the actual dispensary. In light of this, each method has unique benefits and potential drawbacks, as elaborated below.

The identification assumption of the DDD estimate is simply that there was no shock during our study period that differentially affected prices of only the treatment location in the treatment neighborhood. The addition of control neighborhoods differences out systematic differences between properties close to and farther away from the dispensaries (or potential dispensary locations from lottery applications). However, if factors underlying firms' location switching decisions between the lottery address and the actual address are correlated with property price trends, then the DDD estimates may be inconsistent. To explore this issue further, we compare realized neighborhoods versus intended neighborhoods for those firms whose actual address and license application address differed by more than two miles. Table 3 displays the average sale price for homes within half a mile of the firm's realized address versus the application address, characteristics of the addresses' census tracts (the percentage of the tract population 18-24 years old, the percentage of the population that is non-hispanic white, the percentage of the population with a high school diploma, and the median income of the tract), the percentage of yes votes for I-502 in the addresses' precincts, and the distance of the addresses to the nearest school. While realized addresses tend to be in less diverse, more educated, and wealthier census tracts, these differences are not materially large. Indeed, in the paired sample t-test, neighborhood differences between firms' application address and actual address are not statistically significant, as shown in Column (3) of Table 3.[25] Hence, we believe the endogenous location switching problem is unlikely to pose a serious identification threat.

The IV approach, on the other hand, relies on the license lottery outcomes—in which some properties are assigned to treatment and others to control—and allows for partial compliance. As long as enough lottery winners locate their dispensaries close to the stated addresses in the lottery, the lottery results serve as a good predictor of treatment. To identify the treatment effect, there needs to be a strong association between winning the lottery and the actual entry of a marijuana dispensary as well as satisfaction of the exclusion restriction. This means the instrument must be as good as randomly assigned, conditional on the probability of treatment, i.e., W_i, and have no effect on property prices other than through affecting the entry of marijuana dispensary or the estimates will be biased. Given the random nature of the lottery, if W_i adequately controls for treatment probability, then exclusion restriction is unlikely to be violated. We explore this more through a series of covariate tests in Section 5.

4.3 Selection of Treatment Radius

In all three estimation approaches, an important decision is selecting the radius of the inner ring, r, that defines the "nearby" treated group. The choice of r affects both the bias and the variance of the treatment effect estimator. On the former, a very small r means that many treated observations

[25] We perform an additional robustness check in which we drop cases where the entrants are more than 2 miles from the lottery application address. As reported in Table A.1, the results are similar to the baseline estimates.

Table 3: Realized Neighborhood Versus Lottery Application Neighborhood

	Application Address	Realized Address	p-value
Average Sale Price	303,683	322,563	.3255
	(143,692)	(153,850)	
% 18-24 y.o.	.2578	.2814	.1388
	(.1021)	(.1223)	
% White	.7478	.7019	.1245
	(.1332)	(.1671)	
% H.S. Grad.	.6318	.6134	.1797
	(.081)	(.1046)	
Median Income	61,238	57,993	.4151
	(23,788)	(21,022)	
% Y for I-502	.5856	.5918	.6165
	(.0952)	(.1035)	
Miles to Nearest School	.9314	.764	.1488
	(.7083)	(.8515)	

Notes: The column "Realized Address" shows characteristics for the neighborhood in which the firm located while "Application Address" displays characteristics for the neighborhood the firm indicated on their license application. The column "p-value" displays the results of a paired sample t-test testing the differences in characteristics between the firms' realized neighborhoods and application neighborhoods. The reported p-value is based on standard errors clustered at the city level. Characteristics include the average sale price of properties within 0.5 miles of the address, the percentage of the census tract population 18-24 years old, the percentage of the census tract population that is non-hispanic white, the percentage of the tract population with a high school diploma, and the median income of the tract, the percentage of yes votes for I-502 in the addresses' precincts, and the distance of the addresses to the nearest school. The sample includes only those firms that moved greater than two miles from their application address.

would be assigned to the neighborhood control group by the researcher. Similarly, a very large r would include many untreated observations in the treatment group. In both cases, it would bias the average treatment effect estimate toward zero. On the latter, a very small r means that there will be few observations in the treatment group, implying a higher variance of the treatment effect estimator. However, increasing r too much would leave too few observations in the control group, resulting again in a larger variance in the estimate of the treatment effect.[26]

Although the choice of r could have significant impacts on the results, the conventional approach in dealing with this choice is somewhat arbitrary.[27] Hence, in our study, we propose a data-driven and easy-to-implement selection process for choice of r. As explained above, extreme values of r would lead to both bias and imprecision (high variance) in the estimates. Intuitively, starting from small values of r, slightly increasing it would increase the number of observations in the treatment group, decreasing estimator variance and improving the precision of the estimate. It would also more accurately delineate between the treatment and control group. The considerations in distance selection somewhat echos the classic "bias-variance trade-off" of bandwidth selection in non-parametric econometric analysis. Our key insight is to choose an r that minimizes the sample mean squared prediction error, i.e.,

$$\min_r \frac{1}{n} \sum_i (y_i - \hat{y}_i^r)^2 \tag{9}$$

where $\hat{y}_i^r$ is the leave-one-out predicted value of y_i at radius r. Additional details are contained in Appendix A.3.

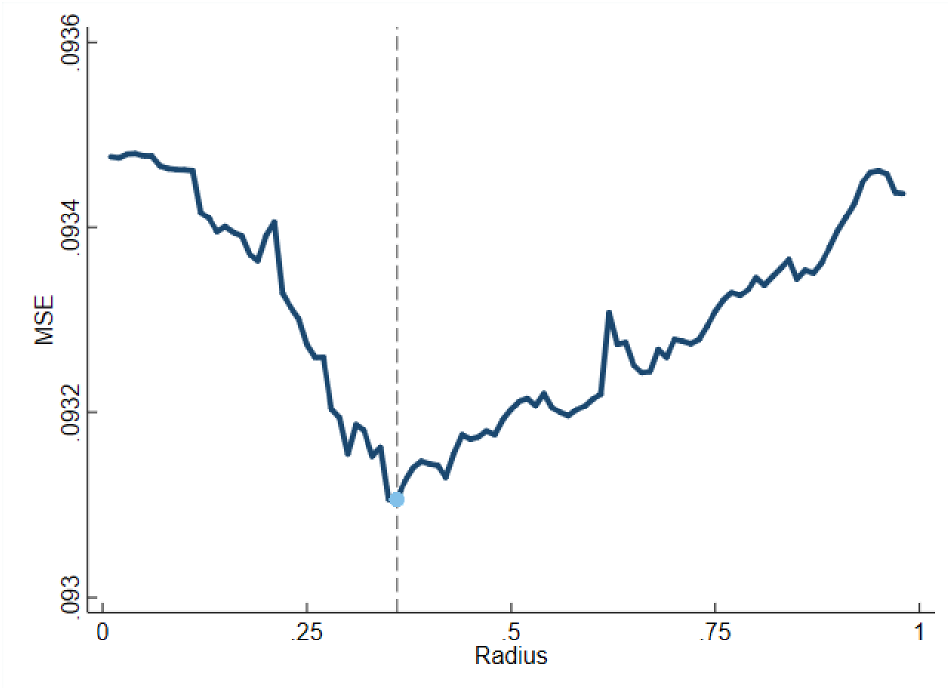
5 Empirical Results

Guided by the empirical strategies, we study the impact of marijuana dispensaries on property prices. We begin by showing the results of the cross-validation procedure that selects the treatment radius. We then evaluate the identifying assumptions of the various estimation methods. We finally

[26] A simple example to illustrate the trade-offs further is contained in Appendix A.2.

[27] The most careful approaches such as Currie et al. (2015) on air pollution use a selection process that is informed by the science. Specifically, they use data on ambient levels of hazardous air pollution to define the treatment radius. However, when the scientific literature offers less direction, others such as Linden and Rockoff (2008) and Muehlenbachs et al. (2015) guide the choice of radius by plotting the non-parametric price gradient and finding a "sensible" distance at which to define the nearby properties.

Figure 5: Cross Validation Results



Notes: This figure plots the mean squared errors, as specified in Equation (9), under different treatment radii (in miles).

present the estimation results including the robustness checks.

5.1 Treatment Radius

In section 4.3, we develop a cross-validation procedure to select a treatment radius in a data-driven manner by minimizing the sample mean squared prediction errors. Results of the cross-validation procedure on Equation (2) are reported in Figure 5. The estimated mean squared error takes the classic U-shape and is minimized at a distance $r^* = 0.36$ miles.[28] We, therefore, adopt a treatment radius of 0.36 miles throughout our empirical analysis. In Section 5.3, we assess how the estimated treatment effect changes as we vary the values of r.

[28] In Appendix A.4, we follow Linden and Rockoff (2008) and Muehlenbachs et al. (2015) and carry out a "sanity check" for the cross-validation result to assess the reasonableness of the treatment radius by estimating local linear regressions of log home sale price on distance to nearest dispensary (or lottery address).

5.2 Identification Assumptions

5.2.1 Parallel Trends

With the cross-validation results in hand, we can study the identification assumptions of our empirical approaches further. In a standard difference-in-differences design, the primary assumption is that absent of the treatment, the treated and control groups would evolve along parallel trends. We examine price trends directly by estimating a local polynomial regression of log home sale prices on days before and after the license lottery results are announced.

Figure 6 displays the results. Figure 6a shows the evolution in prices of properties within one mile to a dispensary, with "near" and "far" denoting prices in the control and treatment groups in the DD model, respectively. We can see that prices in the two groups evolve somewhat similarly pre-lottery. In the triple difference setting, we have an additional control group using lottery addresses. As shown in Figure 6b, price trends for the properties around lottery addresses evolve almost identically. Together, Figures 6a and 6b show that there is no significant difference in property prices in the pre-treatment period between the treatment and control groups. After the lottery announcement, however, prices diverge. This change is most noticeable around 6 months after the lottery winners announcement.[29] As seen in Figure 2, only a few stores enter after July 2014, which may contribute to the delayed response.

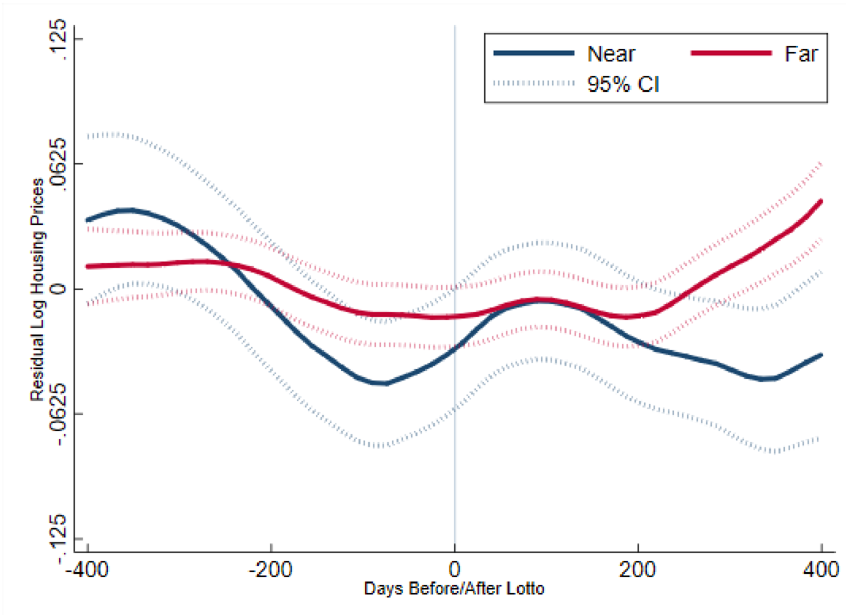
5.2.2 Covariate Balance

As explained in Section 4.2, to address potential selection caused by lottery winners changing locations, we propose an IV approach, in which we use lottery outcomes to predict the actual treatment. For the IV specification, causal inference rests on the assumption that G_i^r is random conditional on vector W_i.

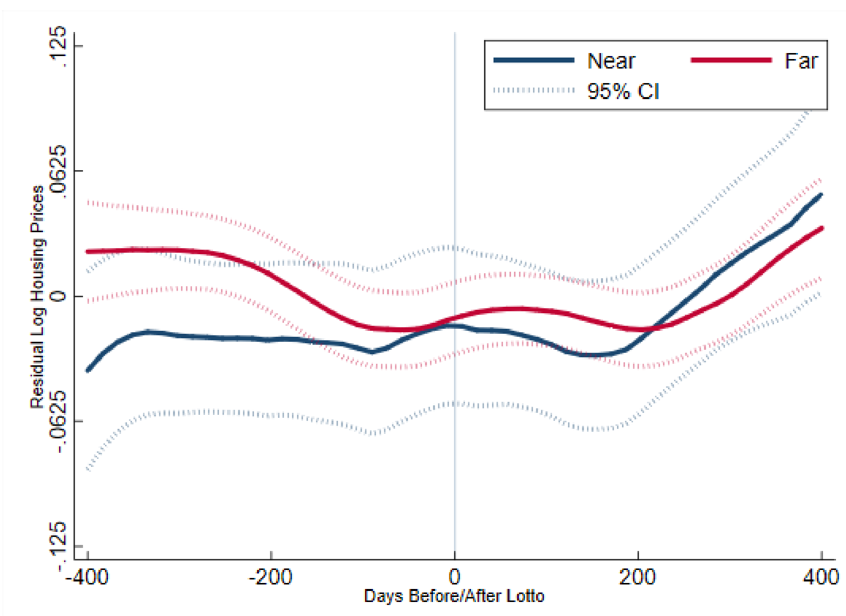
To investigate this further, we study differences in the pre-randomization characteristics of properties and neighborhoods. Table 4 reports the results of the covariate analysis. Looking at simple differences between the groups $G_i^r = 0$ and $G_i^r = 1$ reveals that properties near lottery winners are smaller, older, and in younger neighborhoods in the pre-lottery period. However, raw differences

[29] This pattern is mirrored in our event study graph Figure A.4.

Figure 6: Time Trends



(a) $D_i^R = 1$



(b) $D_i^R = 0$

Notes: Figures 6a and 6b display the results of a local polynomial regression of residual log sale prices (from a linear regression of log sale prices on $Post_t$, $\mathbb{1}(d_i \leq 1 \text{ mi})$, and $\mathbb{1}(d_i \leq 1 \text{ mi}) \cdot Post_t$) on days since the lottery. Figure 6a shows trends for properties within one mile of a marijuana dispensary, while Figure 6b shows trends for properties one mile of a lottery address. Each graph divides the results by properties ≤ 0.36 miles away ("Near") and > 0.36 miles away ("Far"). The bandwidth is 75 days.

do not control for the fact that the probability of treatment differs across cities and neighborhoods. Therefore, we estimate the following regression:

$$y_{ijt} \quad = \quad \rho_0 + \rho_1 G_i^r + \rho_2 W_i + \epsilon_{ijt}, \tag{10}$$

where y_{ijt} are the property and neighborhood characteristics found in X_{ijt}. Zipcode-city fixed effects are also included in W_i. We report the p-value of the regression in the third column of Table 4. Reassuringly, we are unable to reject the null of balance for all but one characteristics: miles to the dispensary (which is different by construction).

We also conduct a similar analysis for G_i^R. The results are shown in the sixth column of Table 4. Two variables, the percentage of the census tract population between the ages of 18 and 35, and number of beds have significant coefficients on G_i^R. However, the raw differences between columns for these variables are not large. For example, a difference of 0.02 separates the average number of beds between the two columns. We cannot reject the null of balance for all other variables.

5.3 Estimation Results

We start with estimation results from the conventional DD approach. The estimation results for the difference-in-differences model are reported in Columns (1) and (2) of Table 5. Even after controlling for zipcode-city fixed effects (FE), the estimated effects of dispensary entry on nearby property values are close to zero and insignificant.[30] However, as shown in Figure A.2, the DD model may not adequately control for unobservables in nearby properties, which justifies the need for the DDD and IV models that control for within-neighborhood selection.

The estimates of our DDD and IV models are found in Columns (3)–(6) of Table 5.[31] Under our DDD specification with zipcode-city fixed effects in Column (4), we find a statistically significant decrease of 3.15% in property prices within 0.36 mile to marijuana dispensaries. Additionally, the estimates of the IV and ITT models, at -.0392 and -.0268 respectively, are qualitatively consistent and quantitatively similar to the DDD estimates. First stage results, given by Sanderson–Windmeijer

[30] We include zipcode-City FE, rather than zipcode FE, because some zipcodes cross city boundaries.

[31] Rather than cross-validating each model separately, we only report estimates using $r = 0.36$, the results of the CV procedure on Equation 2 as this is the model typically used in the literature. This keeps treatment groups consistent across models.

Table 4: Covariate Balance

	G_i^r			G_i^R		
	(1)	(2)	(3)	(4)	(5)	(6)
	0	1	p-value	0	1	p-value
Beds	2.895	2.61	.36141	2.856	2.876	.01274
	(1.037)	(1.008)		(1.104)	(.9901)	
Baths	1.804	1.691	.73685	1.808	1.784	.20207
	(.7752)	(.7563)		(.7868)	(.7654)	
Home Age	42.52	41.64	.349	42.16	42.62	.35698
	(34.52)	(34.24)		(34.95)	(34.18)	
Square Footage	1,683	1,517	.61274	1,699	1,647	.35145
	(761.9)	(604.8)		(811.1)	(706)	
% 18-24 y.o.	.2843	.3085	.71413	.2944	.2814	.02034
	(.1049)	(.1124)		(.1167)	(.0977)	
% White	.6803	.6722	.77541	.6825	.6776	.12417
	(.1583)	(.1358)		(.1537)	(.158)	
% H.S. Grad.	.6422	.6628	.47662	.6516	.6392	.44083
	(.1013)	(.1015)		(.1054)	(.0984)	
Median Income	63,885	65,299	.60295	65,282	63,179	.27056
	(21,953)	(19,820)		(21,995)	(21,570)	
Miles to Dispensary	.6524	.2423	.0168	.6324	.6015	.8675
or Lotto Address	(.2285)	(.0822)		(.2555)	(.2448)	
N	38,594	4,012		16,995	25,611	

Notes: The sample of analysis includes property transactions taking place during the pre-treatment period. Columns (1) and (2) show the average property or neighborhood characteristics for properties where $G_i^r = 0$ (either > 0.36 miles away from a lottery winner or in the lottery loser neighborhood) and where $G_i^r = 1$ (≤ 0.36 miles away from a lottery winner), respectively. Columns (4) and (5) show the average property or neighborhood characteristics for properties where $G_i^R = 0$ (in neighborhoods containing lottery losers) and where $G_i^R = 1$ (in neighborhoods containing lottery winners), respectively. Column (3) shows the p-values for the coefficient estimates of ρ_1 in (10), whereas Column (6) shows the p-values for the analogous analysis for G_i^R. Standard errors in all regressions are clustered at the jurisdiction level.

test statistics, are significant and robust across all instrumental variables regressions. The difference between these estimates and the coefficient estimates in the DD model underscores the importance of having a valid control group to address all endogeneity concerns, especially the potential selection bias in site selection *within* a neighborhood.

As shown in the table, the DDD and IV results with Zipcode-City FEs (Columns (4) and (6)) are similar to each other. The IV estimate (-0.0391) is somewhat larger in magnitude than the DDD estimate (-0.0315), although the difference is not statistically significant. Together, the results imply an estimated negative price impact of around 3%-4%. For the average home sale price in our data, of $332,486, this implies a willingness-to-pay to avoid the disamentity of $10,100-$13,500, a non-trivial amount. It is however worth-noting that, while sizeable, the magnitude is comparable to other estimates in the economics literature. For example, Linden and Rockoff (2008) identifies a 4.1% drop in property values after the arrival of a sex offender in neighborhoods, an implied decrease of $5,500 given median home prices in their area of study: Mecklenberg County, North Carolina. In Davis (2011), neighborhoods within two miles of a power plant experience a 3-7% decreases in housing values and rents while toxic plants lead to declines in property values by 11% for homes within 0.5 miles of the plants in Currie et al. (2015).

An important caveat to this analysis is that we only observe market prices for homes that sell and do not have data on the changing composition of neighborhoods after dispensary entry. Kuminoff and Pope (2014) state that using time series variation as in difference-in-difference estimation could fail to identify the hedonic price function if neighborhood composition changes over time. Along similar lines, if the types of individuals in the housing market changes after cannabis firm entry, the observed prices may not reflect average willingness to pay. Rather, property sellers may be those that have a higher than average willingness-to-pay to avoid dispensaries while buyers have a lower than average willingness-to-pay. Data such as long-run demographic data or data on buyer and sellers would be needed study any neighborhood compositional changes.

However, as our analysis is short-term, it is unlikely that neighborhoods experienced huge changes in the time period studied. Nonetheless, identifying longer-run effects and studying changes in neighborhood composition remains an important area of future study, particularly if as attitudes evolve over time and citizens become more accustomed or hostile to nearby recreational marijuana dispensaries.

Table 5: Regression Model Results

	DD		DDD		IV		ITT	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D_{it}^r	.0051	-.0006	-.0354***	-.0315***	-.0225	-.0392**		
	(.0093)	(.0089)	(.0124)	(.0114)	(.0192)	(.0188)		
D_{it}^R	-.0919***	-.045***	-.0465	-.0067	.0128	.0137		
	(.0243)	(.016)	(.0288)	(.0194)	(.0143)	(.0143)		
G_{it}^r							-.0144	-.0268**
							(.0133)	(.0132)
G_{it}^R							.009	.0096
							(.01)	(.0101)
F-Stat (D_{it}^r)					512.2583	517.3231		
F-Stat (D_{it}^R)					740.2693	742.3055		
City FEs	Y	N	Y	N	Y	N	Y	N
Zip-City FEs	N	Y	N	Y	N	Y	N	Y
N	82461	82289	82461	82289	67282	67135	67282	67135
R^2	.794	.821	.795	.821			.816	.835

Notes: The estimating equation for Columns (1)-(2) is Equation (2) while Equation (4) is the estimating equation for Columns (3)-(4). The estimating equations for Columns (5)-(6) and Columns (7)-(8) are Equation (5) and (8), respectively. Observations are for each property transaction within the study period. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type), census tract characteristics (share of white, median income, median age, share of high school graduates), and precinct-level I-502 referendum results. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

Table 6: Results Varying r

	$r = 0.25$	$r = 0.35$	$r^* = 0.36$	$r = 0.45$	$r = 0.55$	$r = 0.65$	$r = 0.75$
D_{it}^{r}	.0056	-.0317***	-.0315***	-.026***	-.0212***	-.012	-.0122
	(.0228)	(.0098)	(.0114)	(.0084)	(.008)	(.0082)	(.0092)
D_{it}^{R}	.0067	.0132	-.0067	.0123	.0123	.0117	.013*
	(.0091)	(.0093)	(.0194)	(.0088)	(.0081)	(.0082)	(.0076)
N	82289	82289	82289	82289	82289	82289	82289
$\sum_i D_{it}^{r}$	2,254	4,543	4,831	6,477	9,025	12,181	15,772

Notes: The estimating equation is (4). Each column includes zipcode-city fixed effects. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type), census tract characteristics (share of white, median income, median age, share of high school graduates), and precinct-level I-502 referendum results. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

Additional Results and Robustness Checks

To examine how varying the distance used to define the treatment group will impact the estimated effects, we report results with differing treatment radii for the DDD model.[32] Table 6 displays estimates for models that use 0.25, 0.35, 0.45, 0.55, 0.65, and 0.75 miles to define the treatment group. In general, the estimates decrease in magnitude and precision increases as r increases, showing the need for data-driven procedures such as cross-validation to define optimal radius. Only $r = 0.25$ breaks this trend—likely due to high price variability in the relatively few properties with $d_i < 0.25$.[33]

To explore the role of distance from the disamenity with respect to treatment effects further, we estimate the following regression that interacts distance from the dispensary or lottery address with treatment status:

$$\ln(p_{ijt}) = \theta_0 + \theta_1 D_{it}^{R} + \theta_2 D_i^{R} \cdot \tilde{d}_i + \theta_3 \tilde{d}_i \cdot Post_{it} + \theta_4 D_i^{R} + \theta_5 \tilde{d}_i + \theta_6 Post_{it} + \theta_7 X_{ijt} + \upsilon_{ijt}, \qquad (11)$$

As before, the variables D_i^R and D_{it}^R are indicator functions with $D_i^R = \mathbb{1}(d_i \leq R)$ and $D_{it}^R =$

[32] Table A.2 in the Appendix reports the results for the IV model.

[33] Figure A.1 in the Appendix provides further details on this. The top panel in Figure A.1 shows the number of property transactions by their respective distance to the closest dispensary (or lottery address). During our study period, there are few transactions that took place for small values of r. The bottom panel of Figure A.1 compares the average natural log housing prices in the control and treatment neighborhoods against distance. For small r, the confidence intervals are too wide to obtain significant treatment effects.

Table 7: Treatment Effects with Distance

	(1)	(2)
D_{it}^R	-.0459**	-.0333**
	(.0188)	(.0149)
$D_{it}^R \cdot \tilde{d}_i$	.0788***	.0602***
	(.0217)	(.0198)
City FEs	Yes	No
Zipcode-City FEs	No	Yes
N	81740	81570
R^2	.744	.783

Notes: The estimating equation is (11). In addition to area fixed effects, controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type), census tract characteristics (share of white, median income, median age, share of high school graduates), and precinct-level I-502 referendum results. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

$\mathbb{1}(d_i \leq R) \cdot Post_{it}$, where d_i is the distance to the closest dispensary. The variable $\tilde{d}_i$ is i's distance from the closest dispensary or lottery address, and X_{ijt} is a vector of property characteristics (number of bedrooms and bathrooms, home age, square footage, property type) and census tract characteristics (median tract income, percentage of high school graduates, percentage of individuals between 18 and 35, percentage of the tract population that is non-Hispanic white), the precinct-level 2012 referendum results, as well as quarter-year and city or zipcode-city fixed effects, and the indicator function $Post_{it}$ is equal to one after the announcement of the license winners.

The results are included in Table 7. The estimates imply a decrease in home values near the dispensary that decreases in magnitude as properties get farther away from the disamenity. All negative effects dissipate at around 0.55 miles.

We also study how varying the choice of R affects the results. In our analysis, we follow Diamond and McQuade (2019) and define a neighborhood by drawing a one-mile circle around a lottery address or marijuana dispensary. We do not want a R too small because it would include too few observations in the control group, and we do not want a R too large because the properties very far away may not be subject to the same price trends and, hence, violate the parallel trend assumption. On balance, we argue that a one-mile radius is a reasonable proxy to cover properties within the same neighborhood.

Table 8: Results Varying R

	$R = 1.00$ (baseline)	$R = 1.25$	$R = 1.50$	$R = 1.75$	$R = 2.0$
D^r_{it}	.-315***	-.0294***	-.0337***	-.0292**	-.0673***
	(.0114)	(.0098)	(.0097)	(.0114)	(.0125)
D^R_{it}	-.0067	.0108	.0157*	.0106	.0489***
	(.0194)	(.0088)	(.0088)	(.0083)	(.01)
N	82289	100861	114949	126354	133864
R^2	.821	.818	.814	.809	.808

Notes: The estimating equation is (4). Each column includes zipcode-city fixed effects. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type), census tract characteristics (share of white, median income, median age, share of high school graduates), and precinct-level I-502 referendum results. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

Nonetheless, in Table 8, we examine how the choice of R affects the estimated effects under the DDD specification.[34] In general, the results remain qualitatively consistent as we increase R.

In addition, we explore how our choice of *post* period impacts our estimates. We define the post period for our initial analysis as after the announcement of lottery winners, consistent with a model of forward-looking consumers. However, even though winning locations were well publicized, it could still be the case that dispensary location is not salient until after dispensary entry. (Response in the post-period appears delayed in Figure 6.) Therefore, we estimate our models using actual store entry as the post-period. However, a difficulty in implementing this approach lies in defining the post period for neighborhoods that do not have a store: While the data has counterfactual locations, it does not have counterfactual entry dates. Therefore, we use the date of the closest dispensary entrant to a lottery loser as the counterfactual entry date.

The results, as reported in Table 9, are qualitatively and quantitatively consistent with our baseline estimates. Specifically, while the magnitude of the coefficients all increase, the estimates are not statistically significantly different from the original point estimates.

Finally, to provide additional evidence for our identification assumptions, we conduct placebo tests

[34]The results for the IV specification can be found in Table A.3 in the Appendix. The results are consistent, but they are imprecisely estimated as R increases, revealing a potential violation of the parallel trend assumption.

Table 9: Using Store Entry as Post-period

	DD		DDD		IV		ITT	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D^r_{it}	.0056	.0024	-.0538***	-.0533**	-.0502	-.0753**		
	(.0118)	(.0103)	(.0205)	(.0229)	(.0372)	(.0374)		
D^R_{it}	.0056	.0024	.027**	.0236**	.0362	.0369		
	(.0118)	(.0103)	(.0108)	(.0093)	(.0237)	(.0232)		
G^r_{it}							-.013	-.0189*
							(.0105)	(.0113)
G^R_{it}							.0133	.0128
							(.009)	(.0086)
F-Stat (D^r_{it})					292.1038	289.8311		
F-Stat (D^R_{it})					473.5281	480.1176		
City FEs	Y	N	Y	N	Y	N	Y	N
Zipcode-City FEs	N	Y	N	Y	N	Y	N	Y
R^2	.796	.817	.795	.821			.816	.835
N	50683	50560	82461	82289	67282	67135	67282	67135

Notes: The estimating equation for Columns (1)-(2) is (2) while the estimating equation for Columns (3)-(4) is (4). The estimating equation for Columns (5)-(6) is (5). Controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type), census tract characteristics (share of white, median income, median age, share of high school graduates), and precinct-level I-502 referendum results. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

Table 10: Placebo Test Results

	May 1, 2013		June 1, 2013		July 1, 2013	
	DDD	*IV*	*DDD*	*IV*	*DDD*	*IV*
	(1)	(2)	(3)	(4)	(5)	(6)
D^r_{it}	.0084	.0031	.0061	.0308	-.0054	.0087
	(.0188)	(.022)	(.0155)	(.0238)	(.0129)	(.0216)
D^R_{it}	-.0295	.0097	-.0275	.0082	-.0207	.0038
	(.0256)	(.0144)	(.0225)	(.0125)	(.0201)	(.0123)
F-Stat (D^r_{it})		908.1146		842.7838		827.2036
F-Stat (D^R_{it})		778.136		803.4284		744.3264
R^2	.817		.817		.817	
N	21502	17402	21502	17402	21502	17402

Notes: The estimating equation for Columns (1), (3), and (5) is (4) while the estimating equation for Columns (2), (4), and (6) is (5) with $r^* = 0.36$. Columns under the heading "May 1, 2013" are those specification defining the placebo lottery date as on May 1, 2013. Other columns are defined similarly. Controls are quarter-year fixed effects, zipcode-city fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type), census tract characteristics (share of white, median income, median age, share of high school graduates), and precinct-level I-502 referendum results. Robust standard errors clustered at the jurisdiction-level are in parentheses. An observation is a property sold in the year 2013. Significance levels: * 10%, ** 5 %, *** 1%.

by estimating equations (4) and (5) on all property sales from 2013. The lottery was not conducted until the Spring of 2014, so effects found in 2013 could be indicative of pre-trends. We choose a date in 2013 to serve as placebo lottery announcement dates and estimate equations (4) and (5), defining $Post_{it} = 1$ for each property transaction after this date. We do this three times for placebo lottery dates May 1, June 1, and July 1, The results are presented in Table 10. Reassuringly, most of the estimates are close to zero or positive, and none of the estimates are statistically significant.

Differences by Neighborhood Characteristics

To analyze how dispensary entry may have heterogeneous effects across neighborhoods, we stratify the sample by demographic variables and estimate Equation (4). The results are reported in Table 11. We first study a sample split across income level, dividing by whether the property is located in a census tract where the median income is above the 2014 median income of Washington state (around \$61,000). While the coefficient estimates suggest that properties in higher income tracts are the ones affected by dispensaries, an F-test cannot reject the null hypothesis of equality between the

DDD coefficients. An F-test between the DDD estimates when the sample is divided by whether the property is in a tract where more than 50% of residents have a bachelor's degree also cannot reject the null hypothesis of equality.

Hansen et al. (2020) describes different types of sales activity in recreational dispensaries near the Washington state border. Therefore, we also divide the sample by whether a property is within thirty-five miles of the state border. While the interior of the state seems to experience a larger decline in prices, an F-test cannot reject equality between the DDD coefficients.

Statistically significant differences between the DDD coefficients emerge when the sample is divided by whether or not the property's census tract is greater than 70 percent non-hispanic white. (The population of Washington is about 76 percent non-hispanic white.) Those properties in more diverse census tracts seem to drive the estimated decrease in property values while mostly white census tracts experience no change in housing prices. Moreover, census tracts where the tract's median age is below the state-wide median age of 37 also seem to drive the decline in sale prices after dispensary entry. Properties in older census tracts do not see a decline in prices.

Significant differences also occur between properties in precincts that had greater than 50% support for I-502. Interestingly, those properties in areas that voted against I-502 but still were close to a dispensary experienced an *increase* in prices while those that voted overwhelmingly for I-502 experienced a decline in prices. Though beyond the scope of this paper, digging into the underlying mechanisms driving these heterogeneous results with respect to neighborhood characteristics is a promising and policy relevant research area.

6 Crime

Thus far we have remained agnostic to what could be driving any changes in home sale price. To that end, we analyze crime as a possible mechanism for depressed prices and use census tract-level police response data from the city of Seattle (which we discuss at length in Appendix A.6) to estimate

Table 11: Estimated Effects by Neighborhood Characteristics

	> Median Income		>50 % Bachelor's		> 70% White		< Median Age		> 50% Pro I-502		< 35 mi. from Border	
	0	1	0	1	0	1	0	1	0	1	0	1
D_{it}^r	-.016	-.0469***	-.031*	-.0334**	-.0568***	.0031	.0084	-.0461***	.1068*	-.0398***	-.0344***	-.0157
	(.0195)	(.0146)	(.0181)	(.017)	(.0168)	(.0173)	(.0124)	(.0145)	(.0622)	(.0103)	(.0125)	(.0282)
D_{it}^R	.0137	.004	.0143	.0081	.0105	.0042	.0161*	.0055	-.0082	.0134	.0132	-.0025
	(.0133)	(.0088)	(.0112)	(.015)	(.013)	(.0094)	(.0087)	(.012)	(.0239)	(.0091)	(.0093)	(.0155)
N	40923	41340	43309	38958	41217	41037	30389	51853	6478	75792	71975	10314
$F_{1,77}$	1.224		.0066		5.0429		7.9016		5.5948		.3834	

Notes: The estimating equation is (4). Columns under the heading "> 50 % Bachelor's" split the sample by whether or not the property is in a census tract where > 50 percent of individuals have a bachelor's degree. '> Median Income" refers to whether properties in census tracts where the median income exceeds the statewide median income. "> 70% White" references whether properties are in census tracts that have a population share of non-hispanic white that exceeds 70 percent. "< Median Age" divides the sample by whether the property is in a census tract where the median age exceeds the statewide median age. "> 50% Pro I-502" divides the sample by whether the property is in a voting precinct where the majority of voters voted "Yes" on Initiative-502. "< 35 mi. from Border" divides the sample by whether the property is within 35 miles of the Washington state border. Observations are for each property transaction within the study period. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type) and zip code fixed effects. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

the following instrumental variables model:

$$crime_{jt} = \gamma_0 + \gamma_1 \cdot Dispensary_{jt} + \gamma_t + \gamma_j + \omega_{jt}, \tag{12}$$

$$Dispensary_{jt} = \rho_0 + \rho_1 \cdot Winner_j \cdot Post_t + \rho_t + \rho_j + \upsilon_{jt}, \tag{13}$$

where $crime_{jt}$ is the number of police responses per 10,000 residents of census tract j at month-year t. The variables γ_t and γ_j are month-year and census tract fixed effects, respectively. The time fixed effects control for any overall cyclical trends in crime while the census tract fixed effects control for any within-tract crime trends. The variable $Dispensary_{jt}$ is an indicator function equal to 1 if a dispensary locates in census tract j at time t. As before, it is likely that dispensary entry in a census tract is correlated with unobserved variables that impact crime rates. Therefore, we instrument $Dispensary_{jt}$ using lottery outcomes, where $Winner_j$ is 1 if there is at least one lottery winner in census tract j, and $post_t$ is equal to 1 after the license lottery announcement. The coefficient of interest is γ_1.[35]

The results are reported in the top panel of Table 12. We find that the number of police response reports decreases by around 13 per 10,000 census tract residents—a magnitude consistent with Brinkman and Mok-Lamme (2019), but this decrease is not statistically significant at 10% level. After sub-dividing crimes by category, drug-related crimes experience a small but statistically significant decrease while no significant change is found for property crimes and violent crimes. In contrast, nuisance-related crime reports such as loitering, disturbing the peace, or traffic crimes increase by 4.2 per 10,000 residents. The increase in nuisance-related crime reports may be driven by a direct effect of marijuana, i.e., consumption of marijuana increases criminality, or an indirect effect in which the marijuana dispensaries draw younger people or other individuals more likely to commit such crimes. Without additional data on the demographic characteristics of those involved in these crime reports, we cannot separate these channels. However, it is worth highlighting that it is possible marijuana lowers aggregate crime, while bringing in those to stores more likely to commit it. Additionally, the cash-only business of marijuana could also encourage street robberies.

We also estimate Equation (12) using only census tracts that neighbor those tracts with dispen-

[35] As the crime data is only available at the census tract level, we cannot replicate the exercise in Section 4, which requires data on the exact locations of crime occurrences.

saries, to study possible spillover effects. As shown in the bottom panel of Table 12, these adjacent tracts see small increases in nuisance crimes as well as a small increase in violent crimes. It is also worth highlighting that the differences between effects in Table 12 are driven by the precision in the estimates rather than the magnitude of the effect.

As these changes to crime in adjacent tracts do not directly correspond with decreases in crimes of directly treated census tracts, our findings are not consistent with the some-what prevalent belief that crimes are "displaced" by marijuana dispensaries. Rather, they are more consistent with a theory where marijuana dispensaries increase the number of people in an area, creating more unwanted disturbances. These perceived nuisances may make neighborhoods more unpleasant places to live, which in turn may induce a drop in property values. However, if the greater number of people is a consequence of increased economic activity generated by dispensaries, then these costs and benefits must be carefully weighed when implementing any local policies intended to mitigate any negative externalities caused by marijuana firms.

Table 12: Crime Rates

	All	Nusiance	Drug	Property	Violent
Panel A: Treated Tracts					
Treated	-13.39	4.216**	-2.78***	-10.88	2.308
	(9.543)	(2.104)	(.9859)	(7.378)	(2.045)
N	2728	2728	2728	2728	2728
F	387.4583	387.4583	387.4583	387.4583	387.4583
Panel B: Adjacent Tracts					
Adjacent	2.692	1.813*	.888	-1.221	2.46**
	(4.697)	(.9297)	(.7157)	(3.519)	(1.018)
N	6200	6200	6200	6200	6200
F	454.58	454.58	454.58	454.58	454.58

Notes: The estimating equation is (12). Each column includes month-year as well as census tract fixed effects. Robust standard errors are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

7 Conclusion

While legalization is often voted on by citizens and legislatures at the state- or national-level, municipalities and neighborhoods are left to deal with any localized negative effects of marijuana businesses. Therefore, significant research on the causal effects of cannabis businesses on neighborhoods is needed to inform local public policy.

To this end, we study the impact of cannabis dispensaries on surrounding property values. A recreational marijuana retail license lottery held in Washington state provides plausibly exogenous variation to neighborhoods that were affected by marijuana retailer entry. Further, because participants in the license lottery were required to submit potential addresses due to location restrictions, we have novel data on both actual entrants' addresses and the addresses for license lottery winners and losers. This allows us to estimate difference-in-differences, triple difference, and instrumental variables empirical models. In addition, we develop a data-driven cross validation procedure to determine the optimal radius from the dispensary to study any effects. Our estimates indicate a marijuana dispensary decreases property values by 3%-4% for homes within 0.36 miles of the retailer, a decline of about $10,000-$15,000 based on the average home values in Washington. These results imply a high willingness to pay to avoid the local negative externalities.

A hypothesized mechanism driving the decline in property values is crime around dispensaries. To investigate this, we study changes in police reports after recreational dispensary entry in the Seattle, WA. While we find limited evidence of a general decrease in overall crime and statistically significant evidence of a decrease in drug-related reports, we estimate that nuisance-related crime reports increase by about 4.2 per 10,000 census tract residents. Moreover, we find evidence that violent crime slightly increases in census tracts adjoining those where dispensaries locate.

Our findings suggest that crime risk could be a contributing factor to the negative price impacts of dispensaries, but that is likely only a partial explanation. Research on the localized impact of cannabis businesses, particularly studies that focus on long-run effects and their interactions with local neighborhood characteristics, remains a vital area of research.

References

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

Anderson, D. M., B. Hansen, and D. I. Rees (2013). Medical Marijuana Laws, Traffic Fatalities, and Alcohol Consumption. *Journal of Law and Economics 56*(2).

Anderson, D. M., B. Hansen, and D. I. Rees (2015). Medical Marijuana Laws and Teenage Marijuana Use. *American Law and Economics Review 17*(2).

Autor, D. H., C. J. Palmer, and P. A. Pathak (2014). Housing market spillovers: Evidence from the end of rent control in cambridge, massachusetts. *Journal of Political Economy 122*(3), 661–717.

Azofeifa, A., M. E. Mattson, G. Scauer, T. McAfee, A. Grant, and R. Lyerla (2016). National Estimates of Marijuana Use and Related Indicators – National Survey on Drug Use and Health, United States, 2002-2014. *MMWR Surveillance Summaries 65*(11).

Black, S. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics 114*(2), 577–599.

Brinkman, J. and D. Mok-Lamme (2019). Not in My Backyard? Not So Fast. The Effect of Marijuana Legalization on Neighborhood Crime. *Regional Science and Urban Economics 78*.

Burkhardt, J. and M. Flyr (2018). The effect of marijuana dispensary openings on housing prices. *Contemporary Economic Policy*.

Campbell, J. Y., S. Giglio, and P. Pathak (2011). Forced sales and house prices. *American Economic Review 101*(5), 2108–31.

Caulkins, J. P., B. Kilmer, and M. A. Kleiman (2016). *Marijuana Legalization: What Everyone Needs to Know?* (Second ed.). Oxford University Press.

Chaisemartin, C. d. and X. D'Haultfoeuille (2018). Fuzzy Differences-in-Differences. *The Review of Economic Studies 85*(2), 999–1028.

Chang, T. Y. and M. Jacobson (2017). Going to pot? the impact of dispensary closures on crime. *Journal of Urban Economics 100*, 120–136.

Chay, K. Y. and M. Greenstone (2005). Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy 113*(2), 376–424.

Cheng, C., W. J. Mayer, and Y. Mayer (2018). The Effect of Legalizing Retail Marijuana on Housing Values: Evidence from Colorado. *Economic Inquiry 56*(3).

Conklin, J., M. Doup, and H. Li (2018). Contact High: The External Effects of Retail Marijuana Establishments on House Prices. *Real Estate Economics*.

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015). Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings. *American Economic Review 105*(2), 678–709.

Davis, L. W. (2005). The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster. *American Economic Review 94*(5), 1693–1704.

Davis, L. W. (2011). The Effect of Power Plants on Local Housing Values and Rents. *Review of Economics and Statistics 93*(4), 1391–1402.

Diamond, R. and T. McQuade (2019). Who wants affordable housing in their backyard? an equilibrium analysis of low-income property development. *Journal of Political Economy*.

Figlio, D. N. and M. E. Lucas (2004). What's in a Grade? School Report Cards and the Housing Market. *American Economic Review 94*(3), 591–604.

Freisthler, B., P. J. Gruenewald, and J. P. Wolf (2015). Examining the Relationship Between Marijuana Use, Medical Marijuana Dispensaries, and Abusive and Neglectful Parenting. *Child Abuse and Neglect 48*, 170–178.

Freisthler, B., W. R. Ponicki, A. Gaidus, and P. J. Gruenewald (2016). A micro-temporal geospatial analysis of medical marijuana dispensaries and crime in long beach, california. *Addiction 111*(6), 1027–1035.

Greenstone, M. and J. Gallagher (2008). Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program. *The Quarterly Journal of Economics 123*(3), 951–1003.

Hansen, B., K. Miller, and C. Weber (2020). Federalism, Partial Prohibition, and Cross-Border Sales: Evidence from Recreational Marijuana. *Journal of Public Economics 123*, 104–159.

Huber III, A., R. Newman, and D. LaFave (2008). Cannabis Control and Crime: Medicinal Use, Depenalization and the War on Drugs. *The B.E. Journal of Economic Analysis and Policy 16*(4).

Kepple, N. J. and B. Freisthler (2012). Exploring the ecological association between crime and medical marijuana dispensaries. *Journal of Studies on Alcohol and Drugs 73*(4), 523–530.

Kuminoff, N. and J. C. Pope (2014). Do "Capitalization Effects" for Public Goods Reveal the Public's Willingness to Pay? *International Economic Review 4*(4), 1227–1250.

Linden, L. and J. Rockoff (2008). Estimates of the Impact of Crime Risk on Property Values from Megan's Laws. *American Economic Review 98*(3), 1103–1127.

Lu, R., D. Willits, M. K. Stohr, D. Makin, J. Snyder, N. Lovrich, M. Meize, D. Stanton, G. Wu, and C. Hemmens (2019). The Cannabis Effect on Crime: Time-Series Analysis of Crime in Colorado and Washington State. *Justice Quarterly*.

Mair, C., B. Freisthler, W. R. Ponicki, and A. Gaidus (2015). The Impacts of Marijuana Dispensary Density and Neighborhood Ecology on Marijuana Abuse and Dependence. *Drug and Alcohol Dependence 154*, 111–116.

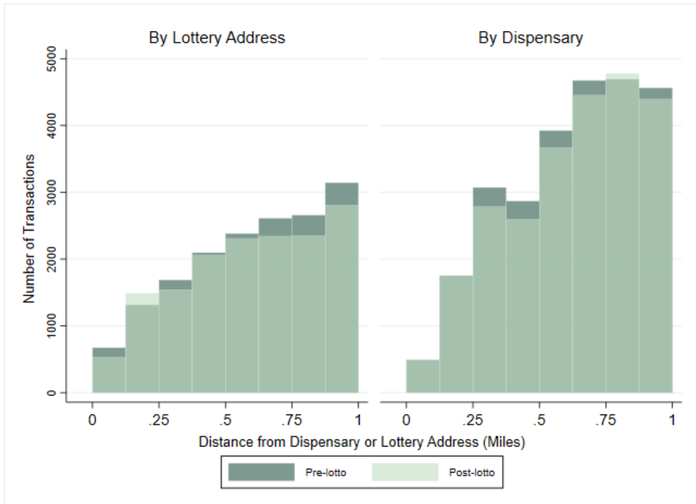
Muehlenbachs, L., E. Spiller, and C. Timmins (2015). The Housing Market Impacts of Shale Gas Development. *American Economic Review 105*(12), 3633–59.

Pope, D. G. and J. C. Pope (2015). When Walmart Comes to Town: Always Low Housing Prices? Always? *Journal of Urban Economics 87*(5), 1–13.

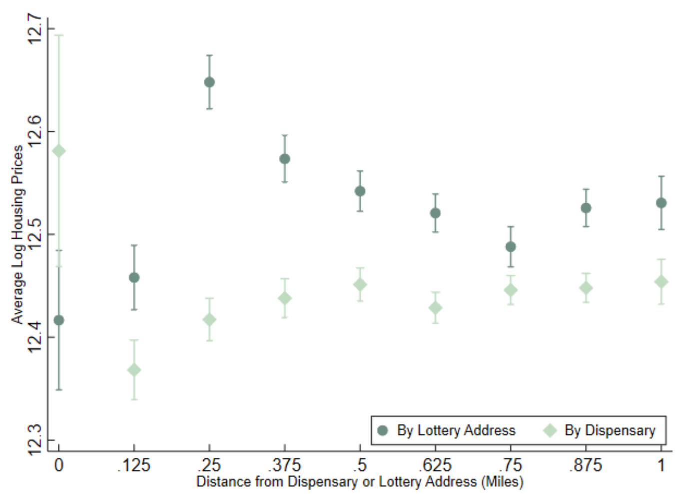
Shi, Y. (2016). The Availability of Medical Marijuana Dispensary and Adolescent Marijuana Use. *Preventative Medicine 91*, 1–7.

Appendix

Figure A.1: Summary of Data by Distance to Dispensary or Lottery Address



(a) Number of Observations by Distance



(b) Price of Properties by Distance

Notes: Figure A.1a displays a histogram of the number of observations by distance to a dispensary or lottery address. The data is further divided by whether the transaction occurred before or after the lottery announcement. Figure A.1b displays the average log home price by distance for properties around lottery addresses and recreational dispensaries. The bars show the standard error of the mean. For both figures, "By Dispensary" properties refer to those properties in the same neighborhood as a dispensary while "By Lottery Address" refer to properties in neighborhoods that missed out on dispensary entry. The data is binned by 0.125 miles.

A.1 Calculating the Probability of Treatment for IV Model

For a given property to be in a winning neighborhood (i.e., $G_i^r = 1$) means that it is within r miles to at least one winning address. The probability of this is therefore one minus the probability that all lottery applicants within r miles of this property lose. All lottery applicants within r miles of a given property lose if all licenses allocated to a jurisdiction are won by applicants further than r miles away to the property. Formally, the probability of treatment group assignment for property i in jurisdiction j is

$$P_i^r = 1 - P(\text{all applicants within } r \text{ miles lose}) = 1 - \frac{{}_{NumApps_j - NumApps_i}C_{Numlicenses_j}}{(NumApps_j - NumApps_i - Numlicenses_j)!},$$

where $NumApps_j$ is the number of license applications in jurisdiction j, $NumApps_i$ is the number of license applications within r miles of i, and $Numlicenses_j$ is the number of licenses in j.

A.2 A Simple Illustration of Trade-offs in Radius Selection

Consider the following simple empirical model:

$$y_i = \alpha + X_i^r \beta^r + \varepsilon_i$$

where $X_i^r = \mathbb{I}(d_i \leq r)$ and r is the radius of the inner ring. A well-established result is that the ordinary least squares estimate of β is

$$\hat{\beta}^r = \frac{\frac{1}{N}\sum_{i=1}^N X_i \cdot y_i - \left(\frac{1}{N}\sum_{i=1}^N X_i\right) \cdot \left(\frac{1}{N}\sum_{i=1}^N y_i\right)}{\frac{1}{N}\sum_{i=1}^N X_i^2 - \left(\frac{1}{N}\sum_{i=1}^N X_i\right)^2} \tag{A.14}$$

By defining $N_r = \sum_{i=1}^N \mathbb{I}(d_i \leq r)$, equation A.14 can be simplified to

$$\hat{\beta}^r = \frac{1}{N_r} \cdot \sum_{i=1}^{N_r} y_i - \frac{1}{(N - N_r)} \cdot \sum_{i=N+1}^{N} y_i. \tag{A.15}$$

Equation A.15 can be interpreted very intuitively: the estimated average treatment effect is the sample average of those i in where $d_i \leq r$ minus the sample average of those observations such that $d_i > r$. However, as mentioned above, if r is increased, then some observations that were originally included in the average for the control group are now included into the average for the treatment group. Hence, $\hat{\beta}^r$ may be sensitive to the choice of radius.

Further, increasing the radius changes the variance of $\hat{\beta}^r$. The heteroskedastic robust standard error of the parameter estimates $\hat{\theta} = (\hat{\alpha}, \hat{\beta}^r)$ is

$$V_{\hat{\theta}} = \left(\bar{X}'\bar{X}\right)^{-1} \left(\bar{X}'D\bar{X}\right) \left(\bar{X}'\bar{X}\right)^{-1}, \tag{A.16}$$

where

$$\left(\bar{X}'\bar{X}\right)^{-1} = \left(\begin{pmatrix} \mathbf{1}'_{N_r \times 1} \\ \mathbb{I}(d \leq r)' \end{pmatrix} \cdot \begin{pmatrix} \mathbf{1}_{N \times 1} & \mathbb{I}(d \leq r) \end{pmatrix} \right)^{-1} = \frac{1}{NN_r - N_r^2} \cdot \begin{pmatrix} N_r & -N_r \\ -N_r & N \end{pmatrix}; \tag{A.17}$$

and

$$\bar{X}'D\bar{X} = \begin{pmatrix} \sum_{i=1}^{N_R} \sigma_i^2 & \sum_{i=1}^{N_R} \sigma_i^2 \\ \sum_{i=1}^{N_r} \sigma_i^2 & \sum_{i=1}^{N_r} \sigma_i^2 \end{pmatrix}. \tag{A.18}$$

Then plugging Equations (A.17) and (A.18) into (A.16) yields

$$V_{\hat{\theta}} = \begin{pmatrix} \frac{\sum_{i=N_r+1}^{N} \sigma_i^2}{(N-N_r)^2} & -\frac{\sum_{i=N_r+1}^{N} \sigma_i^2}{(N-N_r)^2} \\ -\frac{\sum_{i=N_r+1}^{N} \sigma_i^2}{(N-N_r)^2} & \frac{\sum_{i=1}^{N_r} \sigma_i^2}{N_r^2} + \frac{\sum_{i=N_r+1}^{N} \sigma_i^2}{(N-N_r)^2} \end{pmatrix}. \tag{A.19}$$

Therefore, the variance of $\hat{\beta}$ is

$$\frac{\sum_{i=1}^{N_r} \sigma_i^2}{N_r^2} + \frac{\sum_{i=N_r+1}^{N} \sigma_i^2}{(N-N_r)^2}. \tag{A.20}$$

The special case of $\sigma_i = \sigma$ makes clear how the variance changes as N_r changes. The variance of β under homoskedasticity is

$$\frac{\sigma}{N_r} + \frac{\sigma}{(N-N_r)}. \tag{A.21}$$

If N_r is small, the variance is large. (If $N_r \to 0$, the first half of (A.21) approaches infinity.) Increasing r a small amount decreases the variance. However, as $N_r \to N$, the variance will again approach infinity.

A.3 Details of Cross Validation

We aim to formalize an inner-ring radius selection process by developing an easy to implement, *data-dependent* rule for choice of r. The ring method posits the following linear relationship between y and x:

$$y = x^{*'}\beta^* + e, \tag{A.22}$$

where e is i.i.d. random noise with $\mathbb{E}(e) = 0$. The vector x^* is a vector of covariates some of which are dependent on r^* (for example $\mathbb{I}(d_i \leq r^*)$), and r^* is the threshold distance after which y is no longer affected by the disamenity. However, r^* is unknown. Instead, the econometrician estimates

$$y = x^{r'}\beta^r + \varepsilon, \tag{A.23}$$

where $\hat{\beta}_r$ be the linear estimator of (A.23) at radius r. The mean squared error is

$$\begin{aligned}
\mathbb{E}\left(y - x^{r'}\hat{\beta}^r\right)^2 &= \mathbb{E}\left(x^{*'}\beta^* + e - x^{r'}\hat{\beta}^r\right)^2 \\
&= \mathbb{E}\left((x^{*'}\beta^*)^2 + 2x^{*'}\beta^* \cdot e - 2x^{*'}\beta^* x^{r'}\hat{\beta}^r + e^2 - 2e \cdot x^{r'}\hat{\beta}^r + (x^{r'}\hat{\beta}^r)^2\right) \\
&= \mathbb{E}\left((x^{*'}\beta^*)^2 - 2x^{*'}\beta^* x^{r'}\hat{\beta}^r + e^2 + (x^{r'}\hat{\beta}^r)^2\right) + \left(\mathbb{E}(x^{r'}\hat{\beta}^r)\right)^2 - \left(\mathbb{E}(x^{r'}\hat{\beta}^r)\right)^2 \\
&= (x^{*'}\beta^*)^2 - 2x^{*'}\beta^*\mathbb{E}(x^{r'}\hat{\beta}^r) + \left(\mathbb{E}(x^{r'}\hat{\beta}^r)\right)^2 + \mathbb{E}(x^{r'}\hat{\beta}^r)^2 - \left(\mathbb{E}(x^{r'}\hat{\beta}^r)\right)^2 + \mathbb{E}(e^2) \\
&= \underbrace{\left(x^{*'}\beta^* - \mathbb{E}\left(x^{r'}\hat{\beta}^r\right)\right)^2}_{Bias^2} + \underbrace{\mathbb{E}(x^{r'}\hat{\beta}^r)^2 - \left(\mathbb{E}(x^{r'}\hat{\beta}^r)\right)^2}_{Variance} + \underbrace{\mathbb{E}(e^2)}_{Noise}
\end{aligned} \tag{A.24}$$

The first expression in on the right side of Equation (A.24) is the squared bias of the estimator while the second expression is the variance of the estimator.

The best choice of r is one that minimizes the mean squared error. In practice, we minimize the sample mean squared prediction error, i.e.,

$$\min_r \frac{1}{n}\sum_i \left(y_i - \hat{y}^r_{-i}\right)^2 \tag{A.25}$$

where $\hat{y}^r_{-i}$ is the leave-one-out predicted value of y_i at radius r. While $k-$fold cross validation can be used, leave-one-out cross validation is relatively inexpensive from a computational standpoint due to the fact that

$$\sum_i (y_i - \hat{y}_i^r)^2 = \sum_i \left(\frac{y_i - \hat{y}_i^r}{1 - h_{ii}}\right)^2$$

where h_{ii} is the leverage value of the hat matrix $\mathbf{P} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$.

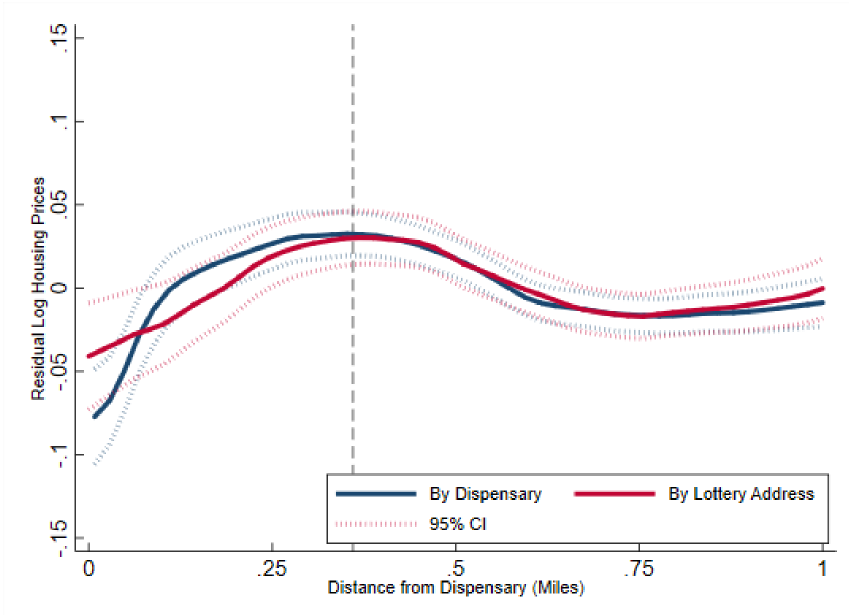
A.4 Cross-validation Results

To assess the reasonableness of this result, we follow Linden and Rockoff (2008) and Muehlenbachs et al. (2015) and estimate local linear regressions of log home sale price on distance to nearest dispensary (or lottery address). The results are displayed in Figure A.2. "By Dispensary" in Figure A.2 refers to those properties within one mile of a marijuana retailer, while "By Lottery Address" references properties that were within one mile of a lottery address but missed being in the same neighborhood as a dispensary. Figure A.2a plots the pre-lottery prices with respect to distance, showing that prices around dispensaries and around lottery addresses have prices that evolve similarly with respect to distance prior to retailer entry. Figure A.2b shows the post-license lottery period. Sale prices in areas near dispensaries decrease after the license lottery. The differences between the curves in Figures A.2a and A.2b, which are plotted in Figure A.3, seem particularly pronounced between 0.15-0.4 miles. The vertical grey dashed line denotes the 0.36 miles, which on visual inspection draws a sensible boundary between properties affected and unaffected by the marijuana dispensary location.

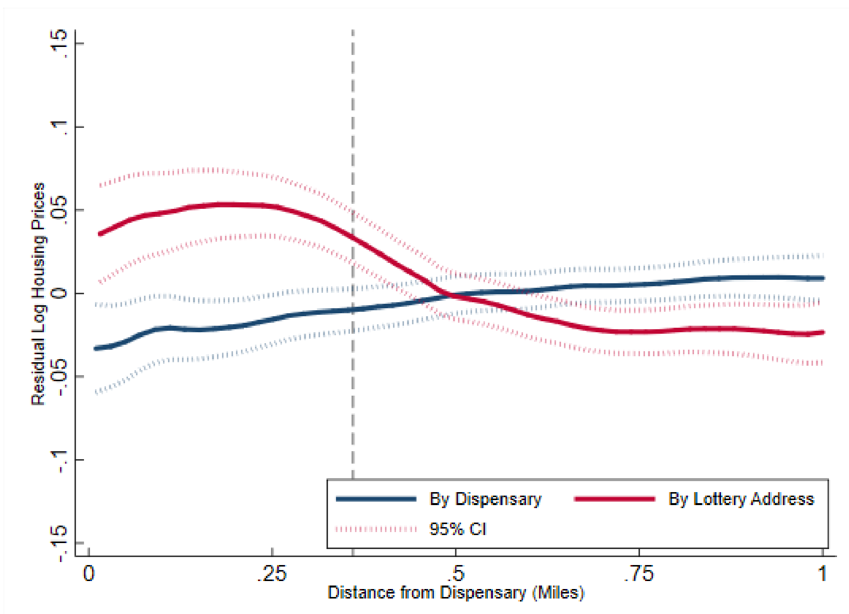
A.5 Additional Robustness Checks

In this section, we present results from additional robustness checks. We first show in Table A.1 that the DDD results are robust when we control for potential selection on entry by dropping observations around dispensaries that move more than 2 miles from the application address. Table A.2 and Table A.3 show how the treatment effect varies when we vary r and R, respectively, under the IV model. These results are similar to the DDD results reported in Table 6 and Table 8. Finally, Figure A.4 displays the event study results.

Figure A.2: Sale Price Gradient of Distance from Closest Dispensary



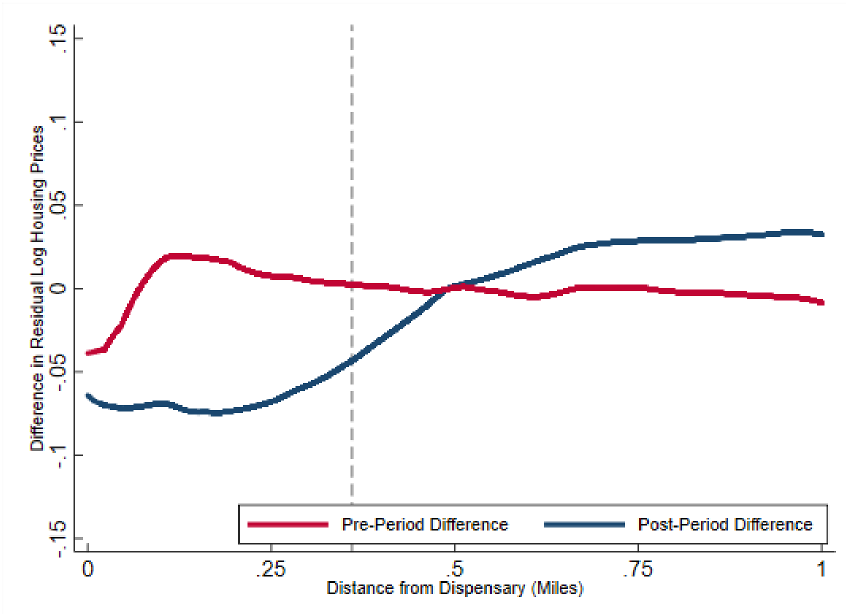
(a) $Post = 0$



(b) $Post = 1$

Notes: Figures A.2a and A.2b display the results of a local polynomial regression of residual log sale prices (from a linear regression of log sale prices on $Post_t$, $\mathbb{1}(d_i \leq 1 \text{ mi})$ and $\mathbb{1}(d_i \leq 1 \text{ mi}) \cdot Post_t$) on the distance to the nearest dispensary or lottery address. "By Dispensary" refers to those properties in the same neighborhood as a dispensary while "By Lottery Address" refers to properties in neighborhoods that missed out on dispensary entry. The bandwidth is 0.125 miles. The gray dashed line denotes the $r^* = 0.36$ found by the cross-validation procedure.

Figure A.3: Price Differences between Treatment and Control Properties by Distance



Notes: The above figure shows the differences between the curves in Figures A.2a and A.2b. The bandwidth is 0.125 miles. The gray dashed line denotes the $r^* = 0.36$ found by the cross-validation procedure.

Table A.1: Triple-Difference Model Controlling for Selection

	(1)	(2)
D_{it}^r	-.0327**	-.0321**
	(.0142)	(.014)
D_{it}^R	-.0565***	-.0399**
	(.0204)	(.0186)
City FEs	Y	N
Zipcode-City FEs	N	Y
R^2	.805	.832
N	59033	58903

Notes: The estimating equation is (4). Observations include only those properties around firms where the distance between the application address and the realized address is less than two miles. Each column includes zipcode-city fixed effects. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type), census tract characteristics (share of white, median income, median age, share of high school graduates), and precinct-level I-502 referendum results. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

Table A.2: Instrumental Variables Results Varying r

	$r = 0.25$	$r = 0.35$	$r^* = 0.36$	$r = 0.45$	$r = 0.55$	$r = 0.65$	$r = 0.75$
D_{it}^r	-.0044	-.0354*	-.0392**	-.0235*	-.0062	.0199	-.0011
	(.0314)	(.019)	(.0188)	(.0133)	(.0116)	(.0154)	(.0129)
D_{it}^R	.0088	.0138	.0137	.0124	.0104	.0024	.0106
	(.0138)	(.0143)	(.0143)	(.0137)	(.0141)	(.0155)	(.0166)
N	67135	67135	67135	67135	67135	67135	67135

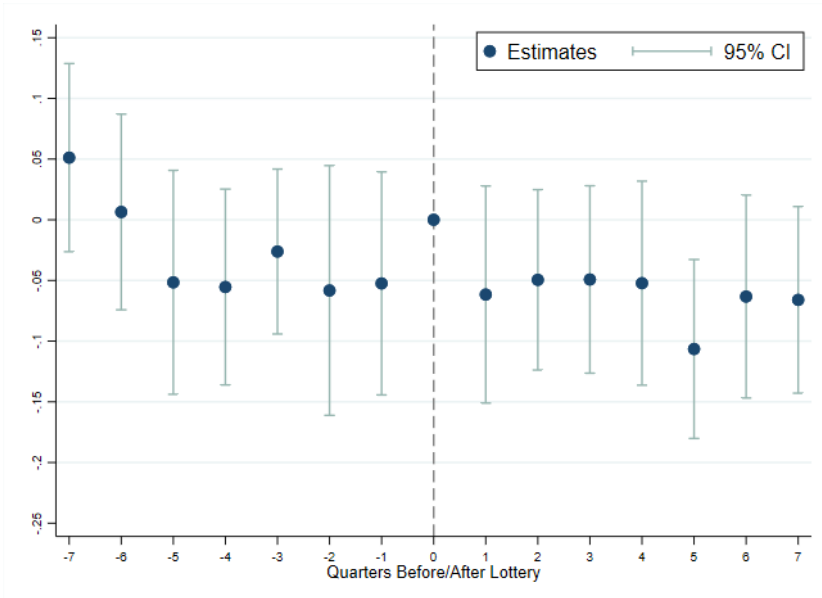
Notes: The estimating equation is (4). Each column includes zipcode-city fixed effects. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type), census tract characteristics (share of white, median income, median age, share of high school graduates), and precinct-level I-502 referendum results. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

Table A.3: Instrumental Variables Results Varying R

	$R = 1.25$	$R = 1.50$	$R = 1.75$	$R = 2.0$
D_{it}^r	-.0382*	-.0312**	-.0263	-.0549*
	(.0211)	(.0148)	(.018)	(.0295)
D_{it}^R	.021	.0131	.0076	.0261
	(.0154)	(.0122)	(.0123)	(.0254)
N	83116	95369	105179	111929

Notes: The estimating equation is (4). Each column includes zipcode-city fixed effects. Other controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type), census tract characteristics (share of white, median income, median age, share of high school graduates), and precinct-level I-502 referendum results. Robust standard errors clustered at the jurisdiction-level are in parentheses. Significance levels: * 10%, ** 5 %, *** 1%.

Figure A.4: Event Study



Notes: Controls are quarter-year fixed effects, property characteristics (number of bedrooms, number of bathrooms, age of property, square footage, property type) and zip code fixed effects Robust standard errors clustered at the jurisdiction-level are in parentheses.

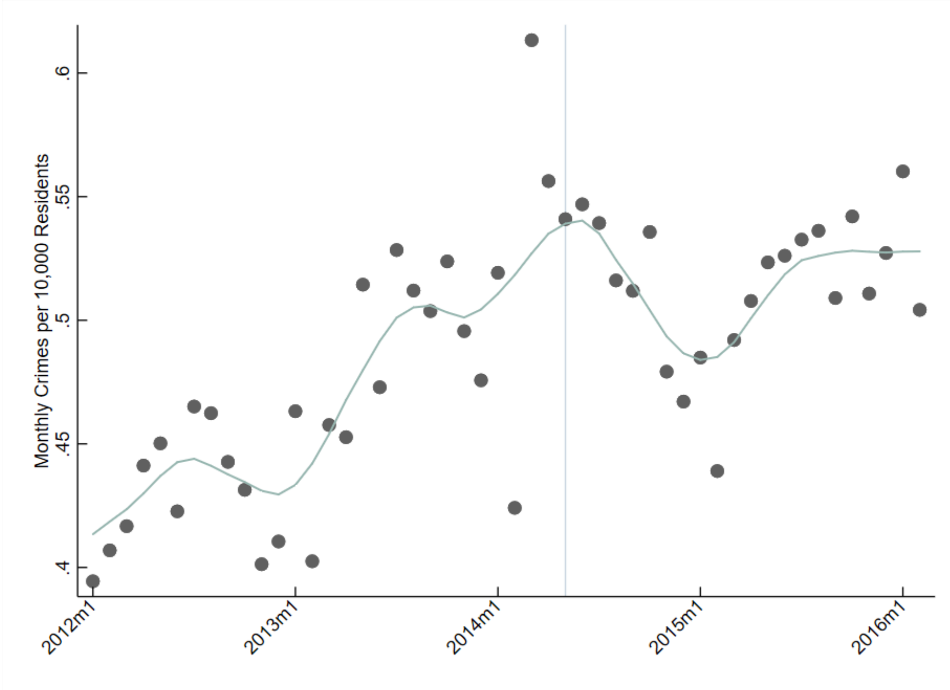
A.6 Seattle Crime Data

Data on reported crimes comes from the City of Seattle Open Data Portal. This comprises all reports to the Seattle Police Department (SPD) and includes the type of offense (e.g., burglary), along with a summarized description of the event (e.g., vehicle robbery), date of the police report, the date the crime occurred, and the location of the offense at the census tract level and block level.[36] We merge the crime data with the previously described data 2014 ACS data from the U.S. Census Bureau in order to calculate crimes per 10,000 residents.

Using the TIGER/Line shapefiles, we also match lottery applicants and operating dispensary addresses to their corresponding census tracts. This enables us to find crime rates in Seattle census tracts that are near marijuana retailers. Further, we limit the analysis to census tracts that contain a lottery address or an operating dispensary or adjoin a census tract contain a lottery address or dispensary. Figure A.5 shows the evolution of crime rates over time while Figure A.6 provides a heat map of police reports in Seattle census tracts along with the locations of dispensaries and lottery

[36]The City of Seattle recently stripped the publicly available data of block and census tract information.

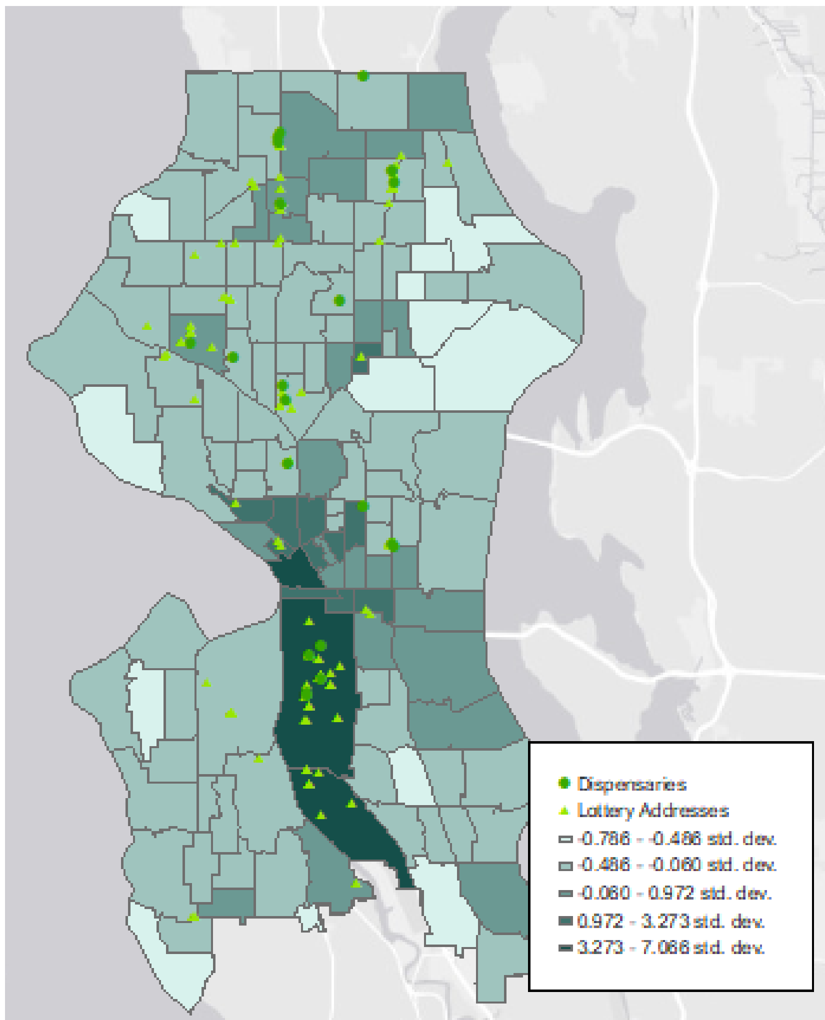
Figure A.5: Crime Rates in Seattle



Notes: The dots show monthly crime rates per 10,000 residents of Seattle. The solid line corresponds to a two-month moving average of crime rates. The vertical gray line denotes the lottery announcement month.

addresses. As seen in the map, marijuana dispensaries tend to locate in areas with a higher number of police reports per 10,000 residents. We also separate out crime rates by types of crime separating out violent, property, drug-related, and nuisance crimes. Table A.4 explains how we categorize these crimes in detail.

Figure A.6: Crime Rates in Seattle Census Tracts



Notes: The heat map displays standardized crime rates (police reports per 10,000 residents) for each census tract in Seattle.

Table A.4: Crime Classifications

Offense Type	Violent	Property	Drug-related	Nuisance
Animal				X
Assault	X			
Burglary		X		
Disorderly Conduct				X
Disturbance				X
Drive-by		X		
DUI			X	
Fireworks				X
Gambling				X
Harassment				X
Homicide	X			
Illegal Dumping				X
Liquor Law Violation			X	
Loitering				X
Narcotics (Possession, Production, etc.)			X	
Pornography				X
Property Damage		X		
Property Stolen		X		
Prostitution				X
Reckless Burning				X
Robbery	X			
Theft		X		
Threats	X			
Traffic				X
Trespass		X		
Public Urination/Defecation				X
Vehicle Theft		X		
Weapon	X			