Three essays in applied microeconomics: Philly style

Alexander Christian Marsella
West Virginia University, acm0043@mix.wvu.edu

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Part of the Criminology and Criminal Justice Commons, Econometrics Commons, Economic Theory Commons, Health Economics Commons, Public Economics Commons, and the Regional Economics Commons

Recommended Citation
https://researchrepository.wvu.edu/etd/11750

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.
THREE ESSAYS IN APPLIED MICROECONOMICS: PHILLY STYLE.

by

Alexander Marsella

Dissertation submitted to the
John Chambers College of Business and Economics
at West Virginia University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Economics

Brad R. Humphreys, Ph.D., Committee Chairperson
Jane Ruseski, Ph.D.
Daniel Grossman, Ph.D.
Alicia Plemmons, Ph.D.
Department of Economics

Morgantown, West Virginia
2023

Keywords: Crime, Urban Policy, Contingent Valuation, Fecundability, Philadelphia

Copyright © 2023 Alexander Marsella
ABSTRACT

Three essays in applied microeconomics: Philly style.

by Alexander Marsella

My dissertation analyzes several contemporary policy-based and institutional occurrences in an urban setting to help guide further advancements in reducing violence, drug overdose deaths, and other unhealthy behaviors that city governments look to curb. Several recent developments in Philadelphia offer a promising setting for studying policies that have broad implications.

Chapter 1 examines the effect of the West Philadelphia Promise Zone initiative on violent crime rates in a high-crime area of West Philadelphia, where a series of educational, public-safety, and quality-of-life improvement grants were disbursed from 2014 onward. My difference-in-differences analysis with two-way fixed effects and cluster bootstrapped standard errors provides the first causal evidence of a modest (approximately 10%) reduction in violent crime, primarily assaults, attributable to this program. By the end of 2019, violent crime in the Promise Zone descended to around the average level across Philadelphia. A synthetic difference-in-differences estimator corroborates this result. In addition, I find evidence that one of the primary grants of the Promise Zone led to increased standardized test scores.

Chapter 2 examines the effect of an information treatment on openness to a particular social service. Fentanyl overdose is a leading cause of death for Americans ages 18 to 45. Recently, an organization called Safehouse attempted to open a “Supervised-Injection Facility” (SIF) in South Philadelphia. Here, intravenous drug-users would have been able to legally use drugs under medical supervision. After progressing past legal hurdles and planning a relatively short-noticed opening, the organization faced immense backlash and “not in my backyard” (NIMBY) sentiment from the local community, ultimately leading to the cancellation of the site. This paper applies contingent valuation survey techniques to this novel scenario in the city of Philadelphia. I find strong evidence of a NIMBY effect, where approximately one half of respondents who support the opening of an SIF relatively far away from them oppose or are unsure of its placement within a mile of their residence. I also find that a randomly assigned information treatment is effective in increasing respondents’ openness to an SIF in their area. Additionally, I find that the perceived cost to residents of an SIF on their block is high: potentially thousands of dollars per month. Support is substantially higher among respondents from Kensington, the heart of Philadelphia’s drug epidemic.

Chapter 3 provides the first analysis examining whether Sugar-Sweetened Beverage (SSB) taxes inadvertently led to increased birth rates within urban populations. Due to the staggered nature of urban soda tax implementation across the United States (and lack of parallel trends), I implement the Staggered Synthetic Difference-in-Differences estimator on county-level births per birthing aged woman across the United States. Despite there being a link in the medical literature between soda consumption and reduced fertility, and literature finding successful demand reductions from SSB taxes, results suggest that urban soda taxes do not reduce soda consumption enough to have a noticeable effect on birth rates.
Acknowledgments

I am very grateful to my advisor, Dr. Brad R. Humphreys, for his guidance and optimism through the job market season and my dissertation work, as well as my committee members, Dr. Jane Ruseski, Dr. Daniel Grossman, and Dr. Alicia Plemmons. Their insight, guidance, and constructive comments have been invaluable to me in my development as a researcher and the completion of my dissertation. Special thanks as well to Dr. Alex Cardazzi, who provided invaluable guidance in coding and research methods to me as my senior in the PhD program. Thank you to all the wonderful students I have taught at West Virginia University, who have reminded me why, at age 18, I felt so strongly that I needed to become an educator of economics. The joy I feel teaching economics to students is unlike anything else in the world and I feel blessed every day to be following this career path in this discipline I love.

Thank you to my muse, Haley, for showering me with so much love the past four years (and for putting up with me during the stressful job market season). Thank you to our cat, Vash, who has provided the both of us so much joy. Thank you, Mom and Dad, for rooting for me and encouraging me since childhood to pursue my passion. I feel very blessed to have parents who have always encouraged me to pursue the kind of intellectual curiosity that was ultimately necessary to complete this dissertation. Thank you as well to the many professors at my undergraduate institution, West Chester University, who supported and encouraged me when I told them about my plans to follow in their footsteps.

I would also like to thank my close friends Akshay, August, Connor, Dan, Gio, Keith, Matt, Mike, Sean, Tom, and Trevor, who I have kept in constant contact with throughout this entire PhD process and especially during the pandemic. The banter and camaraderie on our Discord server has kept me both sane and highly entertained throughout these tumultuous five years.

To all the colleagues, friends, professors, conference attendees, journal referees, and more who have made this whole thing possible: thank you.
## Contents

Abstract ................................................................. ii

Acknowledgments ....................................................... iii

List of Figures .......................................................... vii

List of Tables .......................................................... ix

1 Do place-based crime reduction policies work?: Evidence from the West Philadelphia Promise Zone 1

  1.1 Introduction ....................................................... 1

  1.2 Byrne Criminal Justice Innovation and The Promise Zone ........ 3

    1.2.1 Byrne Criminal Justice Innovation ........................... 3

    1.2.2 The West Philadelphia Promise Zone ......................... 4

  1.3 Context ........................................................... 7

  1.4 Data ........................................................................ 10

    1.4.1 Choice of control variables ................................. 11

  1.5 Methodology ......................................................... 13

  1.6 Results ..................................................................... 14

    1.6.1 Difference-in-Differences ........................................ 14

    1.6.2 Event Study ..................................................... 15

    1.6.3 Synthetic Difference-in-Differences .......................... 15

  1.7 Standardized Test Scores at Treated Schools ......................... 18

  1.8 Discussion .......................................................... 19

iv
# Table of Contents

1.8.1 Possible Mechanism ........................................... 19
1.8.2 Takeaway ....................................................... 20

2 Supervised Injection and NIMBY in Philadelphia: A Contingent Valuation Experiment  
2.1 Introduction ....................................................... 37
2.2 The Economics of Drug Abuse ........................................... 40
2.3 Literary Context ....................................................... 43  
  2.3.1 NIMBY ............................................................ 43
  2.3.2 Contingent Valuation and Public Goods Provision ....................... 44
2.4 Study Design ......................................................... 45
  2.4.1 The Goals .......................................................... 46
  2.4.2 The Subject Pool .................................................. 46
  2.4.3 The Design ......................................................... 49
2.5 Estimation and Results .................................................. 55
  2.5.1 Control variables collected from the survey .............................. 55
  2.5.2 Treatment Effect on Support ...................................... 55
  2.5.3 Heterogeneous Treatment Effects across Demographics ................. 57
  2.5.4 Contingent Valuation .............................................. 58
  2.5.5 Deriving Supply Curves of Respondents ............................... 60
2.6 Conclusion and Takeaway ............................................... 62

3 Soda Taxes and Fecundability: Evidence from the Philadelphia Sugar-Sweetened Beverage Tax  
3.1 Introduction ......................................................... 69
3.2 Literary Context ....................................................... 72
  3.2.1 SSBs and infertility ............................................... 72
  3.2.2 The effect of soda taxes on demand and population health ............. 73
3.3 Economic Theory ....................................................... 74
  3.3.1 Income and Substitution Effect .................................... 74
  3.3.2 The SSB Tax to Increased Fecundability Mechanism ..................... 76
List of Figures

1.1 Philadelphia’s Census Tracts, with the Promise Zone border in red. 22
1.2 The Promise Zone 23
1.3 The Promise Zone in OpenStreetMap 24
1.4 Violent Crime Trends 26
1.5 Event Study 28
1.6 Event Study with Staggered Rollout 29
1.7 Disaggregated Event Study 30
1.8 Synthetic Difference-in-differences with Covariates 31
1.9 Highest Level of Education Attained by Tract Residents 32
1.10 Effect of GEAR UP grant on math PSSA. 34
1.11 Effect of GEAR UP grant on reading/ELA PSSA. 35
1.12 Effect of GEAR UP grant on science PSSA. 36

2.1 Distribution of Durations 48
2.2 Distribution of Durations after Screening on Time 50
2.3 Opening Screen of the Survey. 51
2.4 Order of the Openness Questions 53
2.5 Frequency Distribution of Control Variables 64
2.6 Supply curves by treatment status 65
2.7 Supply curves by addicted family member status 66
2.8 Supply curves by familiarity with SIFs 66
2.9 Supply curves grouped by age: 55 or older vs. younger than 55. 67
List of Tables

1.1 Crime incidents in the Promise Zone ......................................................... 23
1.2 Summary Statistics ...................................................................................... 25
1.3 Difference-in-Differences for Violent Crime ............................................... 27
1.4 Two Period Difference-in-Differences for Education (2013 vs. 2019) ............. 33

2.1 Frequency table: responses with attrition. .................................................. 63
2.2 Frequency table: responses without attrition. .............................................. 63
2.3 Effect of Information Treatment on Support ................................................. 65
2.4 Responses to each money offer exposure. ..................................................... 68
2.5 Repeated Measures Logit for WTA ............................................................. 68

3.1 Summary Statistics ...................................................................................... 80
A.1 Difference-in-Differences for Violent Crime ............................................... 100
A.2 GLMM Models ............................................................................................ 102
A.3 Microsynth SCM Output Table ................................................................. 111
B.1 Allowing WTA to change once respondents are offered money ................. 117
B.2 Including Starting Point Fixed Effects ....................................................... 119
B.3 Baseline Linear Probability Model ............................................................. 121
B.4 Fixed Effects Linear Probability Model ...................................................... 121
Chapter 1

Do place-based crime reduction policies work?: Evidence from the West Philadelphia Promise Zone

1.1 Introduction

Crime causes a plethora of negative externalities in a city in addition to harming the victim; it leads to economic damage, population loss, and psychological damage for residents. Many programs and policies have been proposed to reduce crime in cities, including “place-based” policies. In this paper, I study a recent place-based policy in Philadelphia: the “West Philadelphia Promise Zone”.

This area, home to approximately 2% of the city’s population incurs over $100 million annually in tangible and intangible costs\(^1\) of violent crime. In 2014, it was designated as the West Philadelphia Promise Zone by the Obama administration. This designation, which was part of a broader Obama-administration urban-revitalization program, fast-tracked grants to community organizations within specific geographic areas and created a network of local organizations to work with local law enforcement and community leaders to address local problems such as violent crime. Through this program, over $70 million in grant funding was disbursed from 2014 to 2019 (Stoker and Rich, 2020).

\(^1\)Based on crime data from OpenDataPhilly.org and estimated crime costs from the RAND Corporation.
This paper fills a niche in the literature regarding place-based policy analysis in that, to my knowledge, no published research has studied the Philadelphia Promise Zone directly and little has studied Promise Zones in general. Previous analyses done by community groups in Philadelphia provide descriptive and correlative statistics but do not use modern causal-inference techniques.

One would expect that releasing large sums of grant money into the communities within the Promise Zone, especially money for diversion programs for at-risk youth and criminal offenders returning to the community, would induce a gradual reduction in violent crime rates as the treatment takes hold. In addition to this, various quality-of-life improvement grants, with millions of dollars of funding being released every year, should result in gradual but varying reductions in violent crime. Further, large grants targeted at area schools should reduce violent crime in and around those schools.

That being said, the literature is mixed on whether federally sponsored place-based policies as a general concept are effective in improving areas with high rates of poverty and violence. Some research finds that certain place-based policies were ineffective at helping area residents (Neumark and Young, 2019; Chen et al., 2020; Freedman et al., 2021; Sage et al., 2021) or were too expensive given the mild improvements created (Glaeser and Gottlieb, 2008). On the other hand, there is research that finds more positive effects from place-based policies (Busso et al., 2013; Austin et al., 2018).

This paper first provides a brief description of the developments within the Philadelphia neighborhood of Mantua, the expansion of these developments to the area surrounding Mantua (the Promise Zone), and background on research regarding previous place-based programs (the Promise Zone being one of the most recent). Then, I discuss various assumptions I make about the data-generating process of the Promise Zone and I implement causal-inference methods fit for each assumption. Under the various assumptions about the way in which the treatment unfolds in the area, I apply methods such as difference-in-differences, event study, and synthetic difference-in-differences. In addition to the overall analysis of the Promise Zone as a unified treatment over its area, I examine the direct effect of crime around schools in response to the two largest grants in the program. All my analyses point to the general conclusion that reduced violent crime is attributable to the Promise Zone as a whole and the aforementioned grants to schools. The results are robust to the inclusion of recently developed estimators such as those in Sun and Abraham (2021) and
1.2 Byrne Criminal Justice Innovation and The Promise Zone

1.2.1 Byrne Criminal Justice Innovation

The Byrne Criminal Justice Innovation (BCJI) facilitated the process of getting a grant for the neighborhood of Mantua from 2012 to 2016 focused on data-driven policing strategies in coordination with community residents and leaders to study criminogenic areas. For six months, Mount Vernon Manor Community Development Corporation (MVM) planned and coordinated with the police department, the US Attorney’s Office, and a research partner from Drexel University. Any planned community intervention had to be motivated by data and other evidence facilitated through the research partner. During this time, the team gathered community feedback, administrated focus groups, interviewed residents, and studied crime occurrences geographically using the OpenDataPhilly API, which tracks crime incidents.

This program aimed to develop hot spot policing strategies and data-driven crime-reduction policies and to engage with local community leaders. It also used a competitive grant system in support of partnerships between the local government and nonprofit organizations in the area. At one particular crime hot spot—the corner of 34th and Haverford Avenue—Mount Vernon Manor reported a 65% reduction in 911 calls and an elimination of all arrests. It is not clear that this claim was based on a causal-inference analysis, but it furnishes us the hypothesis that the BCJI affected crime.

The BCJI facilitated several programs, implemented in the fall of 2013, in Mantua that may have reduced violent crime, such as removal of blight, community collaboration with police, more youth programs, and a school-based youth court.

Stokes (2020) provides an in-depth descriptive analysis of this program along with the Promise Zone. He argues that the early stages of the BCJI were contentious for several reasons. First,
the Philadelphia Police Department (PPD) misunderstood parts of the proposal; for example, a misunderstanding regarding whether the grant would support police equipment (it did not) delayed a memorandum of understanding that would establish a budget for police overtime pay in Mantua. Second, since Mount Vernon Manor was the beneficiary of the grant and allocated grant funds in the area, this made things more difficult for city planners. This resulted in a two-year-long reduction in trust between the police department and Mount Vernon Manor, which was reversed when a new police captain took office two years into the program. Coincidentally or not, this was around when the Promise Zone initiative began in West Philadelphia (which contains Mantua). Stokes (2020) goes on to assert that it was not until the Promise Zone designation for West Philadelphia that the effort in Mantua became more effective. The actual BCJI treatments did not begin until the fall of 2013.

1.2.2 The West Philadelphia Promise Zone

As of January 2014, much of West Philadelphia, including Mantua and surrounding neighborhoods, was contained in the Promise Zone. The Promise Zone is depicted in Figure 1.1 as the area outlined in Red. In Figure 1.2, which zooms in on this area, Mantua is comprised of tracts 108 and 109. One of the requirements for the designation of the West Philadelphia Promise Zone was the presence of some form of preexisting place-based program. In this case, an active grant from the BCJI qualified the area. In addition, Promise Zones must be contiguous, contain between 10,000 and 200,000 residents, and have a poverty rate exceeding 32.5%. The area chosen experienced a poverty rate around 50%, high crime rates, and low rates of educational attainment, and it contained numerous abandoned homes. Each city was responsible for outlining the goals of its promise zone. Philadelphia’s zone was implemented to improve education, create jobs, stimulate the local economy, and reduce violent crime. Zones do not receive extra funding outright, but the governmental and nongovernmental organizations within them receive fast-tracked approval for federal grants (Stoker and Rich, 2020). Ultimately, Promise Zones receive more federal grant funding than comparable areas that are not Promise Zone designees. 

6https://www.hud.gov/sites/documents/PZ_R3_APP_GUIDE_URBAN.PDF
7The City of Philadelphia states that the designation was created to “ensure that the ZIP code a person is born in does not determine their future.” https://www.phila.gov/programs/west-philadelphia-promise-zone/
8Upon receiving a direct request for information, AmeriCorps VISTAs working for Philadelphia who coordinate grants through in the Promise Zone kindly provided a list of all grants from 2014 to 2020. The full version of the
Promise Zones differ from Empowerment Zones, Enterprise Zones, and Opportunity Zones. Instead of focusing on spurring outside business investment, the Promise Zone program coordinates federal grant money for a wide variety of urban-renewal programs, many of which focus on improving schools and opportunities for youths. Philadelphia’s Promise Zone designation expanded the BCJI’s strategies, programs, and initiatives in Mantua to the rest of West Philadelphia while also establishing new ones. This expansion along with the preferential treatment federal grants gave to community organizations in West Philadelphia turned the BCJI efforts into a much more substantial force in Philadelphia. To facilitate the goals of the Promise Zone, the Office of Community Empowerment and Opportunity was established in 2014. Under the administration of that office, each initiative of the zone is spearheaded by a specialized organization, of which there are approximately thirty. These include community-development corporations such as Mount Vernon Manor, the Local Initiative Support Corporation, and People’s Emergency Center; public institutions such as the Philadelphia Housing Authority, the police department, and department of commerce; and universities such as Drexel University and the University of Pennsylvania.

The grants supported a wide range of programs: programs matching employees to employers, job-training programs for residents, antirecidivism programs targeted at adjudicated youth, better preschool programs, expanded educational support, and more. They intended to reduce violent crime by implementing community-oriented policing strategies, removing blight, and maintaining vacant lots. They also had some less specifically outlined goals, such as reducing poverty and encouraging healthy eating. The primary goal of this paper is to study the designation’s effect on violent crime.

Unlike police-patrol interventions that begin at time T and can be clearly measured, the plethora of programs that were funded by grants within the Promise Zone and that sought to reduce violent crime began at various times. While the Promise Zone was designated and officially began in January, a tabular list is available in Appendix B, along with the infographics they provided.

9 For reference, Mantua houses approximately 6,000 residents, while the area spanned by the Promise Zone houses over 30,000 residents.

10 For example, Drexel University coordinates the “improved education” goal, while Mount Vernon Manor works with the Philadelphia Police Department to improve public safety and the Housing Authority to improve housing access.

11 https://www.hudexchange.info/sites/onecpd/assets/File/Promise-Zones-Designee-West-Philadelphia.pdf

12 Additionally, analyses for non-violent crime, education, and a comparison of benefit to cost are performed ad hoc in various capacities within the appendix.
uary 2014, it took time for the grants to be allocated and programs to be created. For example, the Face Forward 2 program, a public-safety treatment that provides diversion programs for hundreds of 14- to 24-year-old delinquents living in the Promise Zone, did not begin operating until January 2015.\textsuperscript{13}

Among the largest grants are the GEAR UP and Promise Neighborhood grants, which provided direct and thorough support to schools and their (mostly teenaged) students. This provides the setup for a staggered-rollout event study. In 2014, GEAR UP awarded $14.5 million to the following schools: Martha Washington, Middle Years Alternative, Rudolph Blankenburg, Morton McMichael, Alain Locke, School of the Future, Overbrook High, and West Philadelphia High School. In 2017, the $30 million Promise Neighborhood grant\textsuperscript{14} began to be disbursed to the following schools: Morton McMichael, Samuel Powel, West Philadelphia High, Alain Locke, Martha Washington, Belmont Charter, and Science Leadership Academy Middle School.\textsuperscript{15} According to Becker and Mulligan (1997), not only does education make people less prone to crime because of the increased opportunity cost it creates, but it also makes individuals less impatient and more risk averse.

Over $3 million in grants were dedicated specifically to crime-reduction social interventions. For the Training to Work and Face Forward grants, hundreds of offenders and at-risk youth were treated directly with interventions. These grants were among the first to be disbursed. While this is a small number of people in absolute terms, it is large relative to the offending population and the population of the zone. It is difficult to track individuals directly; one can imagine that some of the crimes (likely assaults, as they are the most common crime) recorded in the pretreatment period were committed by individuals eventually involved in these two grant programs. Treatments such as these should reduce crime rates.

Much of the crime literature argues that making an outside option more desirable than criminal activity makes people less likely to be involved in crime. This is also supported by economic theory since the Promise Zone programs make youth and adult offenders who are released from detention more desirable job candidates and give them extracurricular activities that make them less likely

\textsuperscript{13}It is also important to note that some grants should not affect violent crime in the period studied: for example, I do not expect the “Action for Early Learning Grant”, which provided $4.7 Million in 2014 to support Pre-K through 3rd grade teaching programs, to have any effect on crime by 2019 (though it certainly opens the door for crime research in the mid 2020s, when its age cohort will be reaching adulthood).

\textsuperscript{14}https://drexel.edu/civicengagement/centers-initiatives/promise-neighborhood/

\textsuperscript{15}Among these, West Philadelphia High, Overbrook High and School of the Future are situated 100-200 meters outside the border of the Promise Zone.
to return to crime. In a meta-analysis of youth diversion programs (one of the treatments included in the Promise Zone grants), Wilson and Hoge (2013) find that these programs are effective in reducing recidivism. More broadly, I expect that the zone itself, by establishing an initiative to induce more community cooperation with police, also reduced violent crime by helping to catch more criminals outside the groups treated directly with grant-based programs; there could be potential spillover effects among individuals within the zone. While this mechanism is difficult to isolate, the direction of the effect should be a reduction in crime captured in an analysis of the overall effect of the program.

1.3 Context

Becker (1968) theorizes that crime is incentivized by expected reward to the criminal and disincentivized by a higher risk of the criminal being caught or a greater sanction if caught. Freeman (1999) expands on this theory by arguing that, while sanctions for a crime (such as imprisonment) may disincentivize crime, new criminals simply replace imprisoned criminals in the market for crime. He offers economic opportunity as an alternative to sanctions. If individuals have a better outside option, primarily in the form of gainful employment, they will have less need to commit crimes.

That crime occurs more often in cities than elsewhere is often taken for granted, and it is also a well-studied empirical fact. Glaeser and Sacerdote (1999) dig into the factors behind this phenomenon, focusing primarily on the Becker model but adding external factors such as female-headed households. In line with Becker, they find that approximately one-quarter of crime variation can be explained by higher possible reward due to more concentrated wealth in cities. Moving beyond Becker’s model, they claim that one-third to one-half of the additional crime in urban areas can be explained by the more concentrated presence of female-headed households.

Regarding place-based policies within urban settings, there is some debate in the literature around their effectiveness. Well-known federally guided place-based policies include Empowerment Zones, Enterprise Communities, and Renewal Communities—all of which were designated from 1993 to 2000—and, more recently, Opportunity Zones, designated in 2017. For the first three, their overall effectiveness is generally considered minimal and their cost-effectiveness is often called into

---

There is a near 1-to-1 relationship between female-headed and “single-parent, no father” household. They essentially proxy using data that describes the sex of the head of household, not the number of parents.
question. For example, Glaeser and Gottlieb (2008) study the federal government’s use of localized policy to assist specific regions and neighborhoods through the Empowerment Zone program. This program created eight urban zones across the country that provided tax and regulatory waivers to firms along with block grants for infrastructure spending; though it helped obtain place-based grants, this program was quite different from the Promise Zone program. The authors find that Empowerment Zone neighborhoods experienced a small reduction in poverty and unemployment. These areas also experienced a mild increase in housing prices and rents. But the authors calculate that the program, which cost $3 billion, only increased economic output by $1 billion. Thus, Glaeser and Gottlieb (2008) argue that Empowerment Zones are cost-inefficient.

In contrast, Busso et al. (2013) find, in their welfare analysis of the first round of Empowerment Zone grants, that they created approximately $750 million in value while only costing $400 million over the period studied. They also document reductions in poverty and unemployment rates, similarly to Glaeser and Gottlieb (2008), and ultimately argue that Empowerment Zones were modestly cost-efficient. Reynolds and Rohlin (2015) examine heterogeneous effects across the household income distribution, finding that Empowerment Zones do little to help low-income residents while potentially benefiting high-income residents.

Enterprise Zones, which focus specifically on spurring business, are praised as cost-efficient in Glaeser and Gottlieb (2008), but criticized in more recent literature. Neumark and Young (2019), in a review of the literature on Enterprise Zones, conclude that they do very little to improve employment or income for individuals living in poor neighborhoods. The largest place-based policy initiative since the Empowerment Zones of the 1990s is the Opportunity Zones of 2017, some of which intersect with Promise Zones. Sage et al. (2021) find no evidence of land value appreciation in Opportunity Zones. Chen et al. (2020), studying their effect on housing prices, conclude that the outside investment spurred by the zones does not raise demand for housing in those areas (implying that there is no perceived amenity improvement in the neighborhood). Freedman et al. (2021) find little to no effect on poverty, employment, or earnings. With this in mind, it is unlikely that the establishment of Opportunity Zones in 2017 would reduce the precision of the estimates in a study of a Promise Zone.\footnote{Originally, I stated that, for this reason, Opportunity Zones could not act as confounders. In reality, they can not act as confounders since they do not have a causal link to Promise Zones.}
place-based policies (which focus on incentivizing outside capital investment into areas) is that, even if they create some economic surplus, they do little to benefit the low-income residents of these areas (Reynolds and Rohlin, 2015; Neumark and Young, 2019).

Austin et al. (2018)’s study represents a major departure from some of the earlier literature regarding place-based policies. The authors find, in regard to place-based policies aiming at increasing economic opportunities, programs tailored to specific locations are more effective than large-scale transfers that do not account for local circumstances. This is particularly relevant regarding Promise Zones, since they involve not a general transfer of funds (for example, more funding for police or more money for schools across the board) but funding for specific programs facilitated and crafted by and in collaboration from locals, taking into account the culture and circumstances of the neighborhoods in the zones.

Since the Philadelphia Promise Zone seeks to coordinate public, private, and nonprofit organizations with the ultimate goal of improving a disadvantaged urban area, it best fits the definition of *neighborhood-renewal program* provided by Alonso et al. (2019). Alonso et al. (2019) examine the effect of England’s Neighborhood Renewal Fund (NRF) on violent crime. Examining 345 localities from 2000 to 2007, they study the effect of fund resources that were distributed to 81 of those areas. Their main empirical strategy involves a two-way fixed-effects difference-in-differences approach examining how yearly crime rates are affected after funding is made available to these areas. Similar to the Promise Zone, the NRF involved partnerships among local governments and community organizations in applying renewal interventions using UK government funds. They find that the binary effect of receiving funding reduced burglary by 13%, robbery by 24%, and violence by 13%. Renewal of vacant lots and removal of blight reduce violent crime as well, according to Branas et al. (2018), and Paredes and Skidmore (2017) find that removing dilapidated housing raises nearby property values. That said, it is not clear from the master list of grants that any grants specifically focused on removing blight or dilapidated buildings, even though this goal was laid out explicitly by the facilitators of the Promise Zone.

Kitchens and Wallace (2021) examine the Los Angeles Promise Zone and its effect on local

---

18 “Neighborhood renewal programs are place-based interventions for the regeneration of distressed urban areas” Alonso et al. (2019).

housing prices. The authors find that the Los Angeles Promise Zone caused property value to increase by 6-11 percent\textsuperscript{20}, or about $50,000 on average. They explore potential mechanisms but do not isolate a specific mechanism through which property values increase. They find no significant change in building-permitting, crime, or educational outcomes.

To my knowledge, the only study that addresses the Philadelphia Promise Zone directly is Stokes (2020), which is part of a special report studying the Byrne Criminal Justice Innovation as a whole. This paper provides in-depth background to the BCJI, Mantua, the Promise Zone, and the interactions among the three. It also provides a descriptive analysis of changes in rates of crime in various hot spots across Mantua. In general, I find that through the BCJI, mostly after the Promise Zone took effect, Mantua moved from a high-crime neighborhood toward a more average Philadelphia neighborhood. When we consider that Promise Zone programs are implemented differently from city to city, it is clear that this research supplements Stokes (2020) by using causal-inference techniques and serves as a complement to Kitchens and Wallace (2021) in a (hopefully) growing literature around the effects of the federal Promise Zone program.

1.4 Data

Figure 1.1 shows the entire area studied, including the boundaries of the Promise Zone, attained as a shapefile for all Philadelphia census tracts. Census tracts with fewer than 100 residents at any point during the period studied are excluded. Figure 1.2 provides a close-up view of the treated area. Figure 1.3 provides a richer view with the OpenStreetMap API in Leaflet. The area is bounded by the Schuylkill River to the east, Girard Avenue to the north, 48th Street to the west, and Sansom Street to the south. Using the establishment of the Promise Zone initiative in 2014, I study the effect of these changes on violent crimes in the West Philadelphia Promise Zone.

Crime data for Philadelphia from January 2010 to December 2019 were obtained from the OpenDataPhilly\textsuperscript{21} tool provided by the City of Philadelphia. This API provides access to a data set (updated in real time) that tracks all crime incidents reported by the police. The data set includes the type of crime committed, exact location by coordinates, and exact time down to the

\textsuperscript{20}Using matching, they find a more modest effect of 3-5 percent. Interestingly enough, matching increases the absolute value of my points estimates.

\textsuperscript{21}https://data.phila.gov/visualizations/crime-incidents
minute. This allows for a robust understanding of exactly when and where any given crime occurred. Shapefiles of the City of Philadelphia’s census tracts were also acquired, allowing individual crime occurrences to be mapped and grouped by census tract.

For the purpose of this paper, the over one million observations in this longitudinal crime data were grouped as 3,750 tract-year observations. While there are 384 census tracts in Philadelphia, 9 census tracts that had fewer than 100 residents for at least one year were removed. Violent-crime occurrences within the Promise Zone during the entire period studied are displayed and categorized in Table 1.1. Complete summary statistics for the data are displayed in Table 3.1.

Figure 1.4 shows the average number of crimes per year per 1,000 residents within tracts in and outside the Promise Zone. This figure suggests a downward trend in violent crime in the pre-period; the trend flattens out in most of Philadelphia but continues to fall sharply in the Promise Zone until 2019, when it rises across the city. The zone descends to a similar level of violent crime to the rest of the city by the end of the period studied.

### 1.4.1 Choice of control variables

For yearly demographic controls from the census for units with fewer than 65,000 residents, block-level data are unavailable but both block-group and tract-level five-year American Community Survey (ACS) estimates are available. Tract-level estimates have substantially less sampling error than block-group-level estimates, so I choose to use tract-level controls and aggregation. To understand the incidence of crime, population estimates must be as accurate as possible. Annual population estimates and demographic information from the US Census Bureau were obtained through the ACS at the census-tract level for 2010 to 2019. This paper uses the ACS five-year estimates from the Census Bureau\(^\text{22}\) of racial composition (percent Black, white, and Hispanic); proportion of tract population 25 or older who have not completed high school, completed high school or a GED, completed some college, or completed a bachelor’s degree; proportion of children under 18 living in a single-mother household; proportion of the tract population that are boys or men 15 to 29; proportion of tract population 16 to 64 who have not worked in the past 12 months;

---

\(^{22}\)According to [https://www.census.gov/programs-surveys/acs/guidance/estimates.html](https://www.census.gov/programs-surveys/acs/guidance/estimates.html), the 5-year estimates are more reliable than the 1-year or 3-year estimates. These estimates are recommended by the Census Bureau for performing research at the tract level. It is also important to note that, while these go back to 2009, the Census Bureau cautions against comparing population parameters from the 2009 ACS 5-year to later years due to methodological changes. Therefore I exclude 2009 and begin at 2010.
and real per capita income.

Higher levels of educational attainment, particularly high school completion for boys and men,\(^\text{23}\) are found to be causally linked to lower levels of violent crime (Lochner and Moretti, 2004; Lochner, 2020). Machin et al. (2011) find that more education substantially reduces property crime. Improving educational quality and attainment is a key goal of Promise Zones, and it determines where much of the funding goes. One discrepancy between the Promise Zones’ educational targets and the way the ACS measures educational attainment is that grade schoolers and high schoolers are targeted by the Promise Zones’ educational grants, while the ACS measures educational attainment for individuals 25 and older. Therefore, the effect of educational grants disbursed throughout the zone on educational attainment as measured by the ACS will not be observed. However, the Face Forward 2 grant and the Training to Work 1 grant, which are considered public-safety grants because they target youths and adults currently being adjudicated for crimes, fund GED and college-preparation services for adults and soon-to-be adults. For this reason, an improvement in educational attainment may partially mediate the zone’s effect on violent crime.

Growing up in a single-mother household\(^\text{24}\) is associated with higher youth involvement in crime (Glaeser and Sacerdote, 1999). For that reason, I include the percentage of children under 18 living in a female-headed household with no father\(^\text{25}\) present. According to the FBI, young men commit the majority of violent crime (Ulmer and Steffensmeier, 2014). A change in the share of population that are 15- to 29-year-old boys and men might not be picked up by fixed effects if some exogenous change occurs at the tract level. Neither single-mother households nor presence of young men are confounders, since neither one should have affected the choice of whether to establish a Promise Zone. Failing to control for them does not open a backdoor path\(^\text{26}\), but it still reduces the precision of my estimate. If the share of children living in single-mother households or the share of the population that are young men changes in a tract-specific way (uncaptured by year fixed effects), then it could bias the coefficient of the treatment on violent crime. Race is similar in this regard, as it did not have any direct causal link to the establishment of the zone but could affect

\(^{23}\)This is a key finding of Lochner and Moretti (2004).

\(^{24}\)More specifically, Glaeser and Sacerdote (1999) claim that higher concentrations of female-headed households in cities can help explain much of the crime differential between urban and non-urban areas.

\(^{25}\)Literally “no husband present” in the ACS.

\(^{26}\)For more on backdoor paths and causality, see Cunningham (2021), which is also available at https://mixtape.scunning.com/dag.html.
criminality through a mediator variable such as income or education. I include it as well to increase
the precision of the estimate. Lack of employment can be another causal factor behind committing
violent crimes. I calculate the share of the population 16 to 64 that has not worked in the last 12
months as a measure of long-term joblessness.

1.5 Methodology

The first goal of this paper is to study the overarching effect of the Promise Zone as a whole, which
began in 2014. While parallel pre-trends are not the only requirement for a difference-in-differences
analysis, the clear division between before and after for treated and untreated tracts makes this a
reasonable setting for difference-in-differences. I employ a model of the following form:

\[
Crime_{it} = \beta_1 Post_t + \beta_2 Zone_i + \beta_3 Post_t \times Zone_i + A_{it} + \alpha_i + \gamma_t + \mu_i
\]  

(1.1)

Here, \(Crime\) is the number of crime incidents per 1,000 residents in a tract \(i\) and time \(t\). \(Zone\) is a
dummy variable equal to 1 for any census tract \(i\) fully lying within the boundaries of what would
be the Promise Zone at any time in the period studied. \(Post\) is a dummy variable equal to 1 for any
year \(t\) 2014 or later. It should be noted that both \(Post\) and \(Zone\) are subsumed by the fixed effect
terms when applicable. The interaction term is the basis of the difference-in-differences analysis;
it is equal to 1 for a zone tract in 2014 or later. \(\alpha_i\) represents a time-invariant tract fixed effect
to capture unobserved characteristics of tracts; \(\gamma_t\) represents year fixed effects to capture universal
time trends throughout the city. \(A_{it}\) is a vector of controls, namely the five-year ACS estimates
discussed in Section 1.4, which are measured at the tract-year level. These variables include the
percent of the population that are white, percent that are Black, percent that are Hispanic, percent
of children under 18 who live in a household headed by a single mother, percent who are boys or
men 15 to 29, percent with various levels of highest educational attainment (no high school diploma,
high school or GED, some college, associate’s degree, bachelor’s degree, and the reference category
of graduate degree), percent who are 16 to 64 and who have not worked in the past 12 months,
and real income per capita. \(\mu\) is the error term. Standard errors are cluster-robust at the tract and
year level and are wild-cluster-bootstrapped for the interaction term when significant.
It is not clear that the treatment effect is consistent over time given that many grants were disbursed over several years. It could be useful to understand the heterogeneity of the treatment effect with respect to time. Therefore, I also employ a dynamic difference-in-differences (event study) model in which Zone is interacted with each year in the period studied, using 2013 as the reference year (t=-1). This allows me to further test for parallel pre-trends and to determine whether the strength of the effect varies across the post-period. The model is specified as follows:

\[ Crime_{it} = \sum_{t=-4}^{-2} \beta_k \times Treat_i + \sum_{t=0}^{5} \beta_k \times Treat_i + A_{it} + \alpha_i + \gamma_t + \mu_i \]  \hspace{1cm} (1.2)

Here, Treat is equal to 1 only for tracts within the Zone when t\geq0. This model uses the same controls, fixed effects, and error clustering as the previous difference-in-differences model.

An alternative assumption should be considered. On paper, Mantua, comprising Tracts 108 and 109, was treated beginning in 2013. The BCJI, which qualified the area for the Promise Zone in the first place, was fully implemented by 2013. I employ an additional event study using 2013 as the treatment start date for Tracts 108 and 109. Sun and Abraham (2021) demonstrate the shortcomings of event-study specifications with staggered treatment rollout. Following their method, I use their estimator along with a placebo-treatment year far outside the data.27

1.6 Results
1.6.1 Difference-in-Differences

Table 1.3 displays the results of the difference-in-differences model, both with and without the full set of controls and two-way fixed effects. When the full set of controls and fixed effects are used, the Promise Zone in the post-period is associated with approximately four fewer violent crimes per thousand individuals in a tract. This means that the average fully contained Promise Zone tract, consisting of approximately 3,500 individuals, sees 15 fewer violent crimes per year attributable to the treatment. With eight tracts lying fully in the zone, this corresponds to 120 fewer violent crimes.

27This was applied using fixest in R. For more, see this tutorial.
crimes per year in total. This is approximately a 10% reduction in violent crime. Note that the result is robust to the addition of controls and two-way fixed effects. I also include bootstrapped standard errors clustered at the tract and year level due to there being few treated clusters (Webb, 2014; Cameron and Miller, 2015).

1.6.2 Event Study

The results of the first event study are displayed in Figure 1.5. This event study compares the difference in violent crime between the Promise Zone and the rest of Philadelphia for each year in comparison to the reference year of 2013. The event study suggests that there is no significant divergence in trends between the zone and the rest of Philadelphia during the pre-period. In the post-period, all years other than 2015 experience a statistically significant reduction in violent crime. The downward trend for the mean effect of the interaction term suggests the treatment effect increased over the years following the Promise Zone’s establishment. This result is not sensitive to the alternative assumption that Mantua’s treatment begins in 2013. Figure 1.6 shows the results of the event study when treatment rollout is assumed to be staggered, with Mantua receiving treatment one period before the rest of the zone. I report both the standard TWFE event study results and the modified Sun and Abraham (2021) staggered rollout estimator. The results suggest that the treatment effects are significant in all post-periods other than the treatment start period. For Mantua, this is 2014-2019; for the rest of the zone, this is 2015-2019. Under either assumption (simultaneous or staggered rollout), the Promise Zone experiences a statistically significant decline in violent in the years following its establishment. That being said, the staggered rollout model’s Sun and Abraham estimator shows a divergence in the differences between treatment and control between the year before treatment and the period two years before treatment. While the point estimates vary depending on the assumptions made, that the general result (causal evidence of reduced violent crime) is not sensitive to this change in assumptions suggests this study is capturing a true reduction in violent crime.

1.6.3 Synthetic Difference-in-Differences

Synthetic control is particularly useful as an alternative to difference-in-differences in this case since graphically it may seem that the treatment had no effect on an already downward-trending violent crime.
crime rate. I argue that violent crime would have stopped trending downward and possibly trended upward had the Promise Zone not been established in 2014. One benefit of synthetic control is that it works well in situations in which the number of treated units is small in comparison to the number of untreated units; as mentioned, it does not require parallel trends to exist naturally in the data since it reweights units to match on pretreatment trends and levels (Abadie and Gardeazabal, 2003; Abadie et al., 2015; Arkhangelsky et al., 2021).

Arkhangelsky et al. (2021) develop an estimator that they demonstrate performs as well as or better than difference-in-differences and classic synthetic control (from Abadie and Gardeazabal (2003)) in settings in which one or the other would normally be appropriate. In a typical synthetic control, statistical software estimates unit weights (generally along with covariate and pretreatment outcome weights) using pretreatment data to simulate the treated unit’s outcome variable (and covariates if applicable) during its pre-period and post-period using this weighted combination of control units. This imposes parallel trends in the pre-period econometrically; ideally, the synthetic control unit has the same trend and level\(^28\) as the actual treated unit. Synthetic difference-in-differences introduces time weights in addition to the unit weights normally used, thus giving more weight to units that are historically similar to the treated unit. In contrast to synthetic control, synthetic difference-in-differences imposes parallel trends but not identical levels. Since the control unit’s outcome begins at a different level from the treated unit’s outcome (similar to a classic difference-in-differences setting) the ATT (average treatment effect on the treated) point estimate is often more conservative than both synthetic control and difference-in-differences.

I implement this estimator using the `synthdid\(^29\)` package in Rstudio. This method provides a point estimate for the treatment effect of -4.395 with 95% CI [-8.86, 0.07] calculated using a cluster bootstrap algorithm (with 10,000 iterations) as recommended by the authors.

I have argued in this paper that covariates matter: some of them are confounders that must be controlled for, while others are partial mediators that should be controlled for to increase precision. For that reason, I believe that the synthetic difference-in-differences estimator with covariates is a more accurate measurement of the causal effect. The method in Arkhangelsky et al. (2021) is focused primarily on cases without time-varying exogenous covariates, but a method including

\(^{28}\)This is something Kahn-Lang and Lang (2020) note can strengthen classic DiD approaches if a suitable control group exists.

covariates is briefly described in a footnote of theirs. Kranz (2021) demonstrates that the method for covariate implementation suggested by Arkhangelsky et al. (2021) might not provide consistent estimates when covariates are correlated with both time period and group. While my difference-in-differences estimates barely change from the addition of the full set of controls, I do believe that the controls I use are correlated with time and group. For example, I know that certain demographic characteristics are systematically different30 in the Promise Zone compared to the rest of Philadelphia and that these characteristics change from year to year. Therefore, I also employ the xsynthdid package, which allows me to include covariates. I include all covariates used in the difference-in-differences and event-study models from earlier. This package also allows me to compute clustered-bootstrap standard errors, which I implement with 10,000 iterations. This provides an estimate of 4.085 with 95% CI [-8.12, -0.47].31 Figure 1.8 displays the synthetic difference-in-differences plot from this specification. The plot displays a parallel pre-trend (notably more parallel than the one naturally existing in the data) and that the average treatment effect is represented by the black arrow. In addition, the red shaded area on the bottom left of the plot represents the time weights (the λ term as described in Arkhangelsky et al. (2021)).

Upon attempting to disaggregate the effect by specifying the synthetic difference-in-differences model with each component of violent crime as a separate outcome variable, none of them are significant at the 5% level. The results for simple assault come close, though, with a point estimate of -2.92 and 95% CI [-6.35, 0.37]. The significant effect on violent crime, a composite variable, is likely driven primarily by a reduction in simple assaults combined with other crimes’ minor reductions that are not large enough to be statistically significant on their own.32

Additionally, I find no evidence of any change in property crime33 rates, with a point estimate of 1.991 and 95% CI [-3.75, 8.89].

30 As discussed in Section 1.4 of this paper.
31 Note that this confidence interval is not immediately provided by the package. These constitute the 2.5 and 97.5 percentile (the 250th and 9750th) values of the 10,000 iterations for the bootstrapped mean difference.
32 Further analysis using the synthetic control estimator and aggregation level suggested by Robbins and Davenport (2021) is provided in Appendix A. This alternative methodology strongly suggests a reduction in simple assaults and a potential mild reduction in aggravated assaults with firearms. It should be noted that the permutation method used for inference in that method is less conservative than the cluster bootstrap methodology suggested by Arkhangelsky et al. (2021).
33 This is composed of “Theft from Vehicle”, “Motor Vehicle Theft”, “Thefts”, “Burglary Residential”, “Burglary Non-Residential”, and “Receiving Stolen Property”.
1.7 Standardized Test Scores at Treated Schools

The GEAR UP grant began disbursement in the fall of 2014 with a $4.4 million grant ($14.5M for the full seven years of the grant) to the school district of Philadelphia to provide “mentoring, tutoring, college awareness, college preparation work, financial aid support, and study skills” to eight schools, two of which lay just outside the border of the Promise Zone and six of which are fully contained within the Promise Zone. According to governmental documents,\(^\text{34}\) the target cohort included 5,463 students. The Promise Neighborhoods Grant began disbursing $30 million in 2017 (spread across six years), and targeted seven schools, one of which lay just outside the Promise Zone. These grants are by far the largest grants raised in the time period studied, and they represent more than half of the total grant money released from 2014 to 2019. All of the treated schools are publicly funded (district or charter) elementary, middle, or high schools or some mix of the three.

In addition to examining the overarching effect of the many programs facilitated through the zone as one unified treatment affecting tract-level crime, I examine scores from standardized tests administered to public (non-charter) schools. Data for the Pennsylvania System of School assessment tests are available for 158 Philadelphia public schools with 3rd-8th grade students. These tests are administered every Spring: 3rd-8th grade students take the ELA and math exams while 4th and 8th grade students take the science exam. Among these schools, five were awarded a piece of the GEAR UP grant in 2014 with three of those five receiving a piece of the Promise Neighborhood grant in 2017. One school that did not receive GEAR UP but did receive the Promise Neighborhood grant is also observed and an indicator variable controls for this. Since the tests are administered in the Spring, and grant disbursement began in 2014, I specify a model of the following form

\[
Score_{it} = \sum_{t=-5}^{-2} \beta_k \times Treat_i + \sum_{t=0}^{4} \beta_k \times Treat_i + \alpha_i + \gamma_t + \mu_{it} \tag{1.3}
\]

where the pre-treatment reference year is 2014, and \(Treat\) is equal to 1 when \(t \geq 0\). \(\alpha\) is a

\(^{34}\text{https://www2.ed.gov/programs/gearup/gu-abstracts2014.pdf}\)
set of school dummy variables and $\gamma$ is a set of year dummy variables. Score is a measure of the proportion of students who scored “Proficient” or “Advanced” (as opposed to “Basic” or “Below Basic”) at school $i$ in year $t$. Standard errors are clustered at the school and year level.

Results from this model are displayed in Figure 1.10, Figure 1.11, and Figure 1.12. Among the three, the results for science suggest a significant increase in science PSSA scores after the GEAR UP grant was disbursed to those schools.

1.8 Discussion

1.8.1 Possible Mechanism

It is useful to attempt to identify the mechanism through which the Promise Zone reduces violent crime in order to better inform public policy and future Promise Zone initiatives. The primary goal of the West Philadelphia Promise Zone is to improve educational attainment. Figure 1.9 provides a rough estimate of educational attainment within and outside the Promise Zone. While the relatively large margins of error for tract-level ACS data warrant caution, the figure paints the expected picture: the Promise Zone has lower educational attainment than Philadelphia at large. At a glance, it seems as though the share of residents who lack a high school diploma fell in the post-period.

While consecutive years from the ACS five-year estimates can be used as covariates, the Census Bureau recommends against using consecutive years as outcome variables since overlapping samples might not detect any treatment effects. For example, the 2015 ACS five-year estimate of high school completion in a tract involves data collected since 2011. When compared to the 2014 estimate, there are four years of overlap in data collection. Unfortunately, one-year estimates do not cover units with fewer than 65,000 residents. For this reason, I specified a difference-in-differences model that uses the 2013 and 2019 tract-level education estimates. This allowed me to compare estimates made with information from 2009 to 2013 to estimates made using information from 2015 to 2019. Excluding the year of the establishment provides a very clear cutoff for the onset of treatment. One limitation of this method is that it is difficult to make any claim of a parallel pre-trend when I am only using two periods of data. That said, Table 1.4 displays the results of a two-period difference-in-differences model for education. I included tract fixed effects, but year fixed effects
are already captured by the Post variable, which is equal to 1 for 2019. Similarly to the graphs, Column 1 shows very limited evidence that the proportion of individuals lacking a high school diploma or GED fell by 2.9 percentage points, while a time trend accounts for 3.4 percentage points of decline as well. It would be easier to say that the high school completion rate rose than that the rate of non–high school completion fell, but unfortunately I lack the statistical power to make that claim. Regardless, the signs are in the correct direction, suggesting that the zone may be improving educational outcomes as observed by the ACS 5-year data.

That said, giving teens more guidance, purpose, and activities at or around their schools (included in the significant GEAR UP and Promise Neighborhood grants) could very well be the mechanism reducing crime in the zone. Considering that simple assaults are the primary factor underlying crime reduction, I suspect that teenagers are less likely to be getting into fights and other forms of violence after expanded activities, parental intervention, and guidance interventions are provided at the schools by GEAR UP and Promise Neighborhood. The analysis from Section 1.7 provides some evidence that test scores are rising at treated schools.

It is possible that demographic factors that affect violent crime could have changed in ways that would not be captured by tract and year fixed effects, but I do my best to control for these factors. Tract and year fixed effects determine the vast majority of the variation in violent crime, while the controls determine very little. The totality of the situation suggests that the overarching effect of the Promise Zone is to reduce crime, but the most noticeable effect is from the large grants targeted at schools.

### 1.8.2 Takeaway

The Promise Zone led to a reduction in violent crime overall, primarily through a reduction in simple assaults. In addition, the GEAR UP grant may have caused an increase in science PSSA scores. This paper finds two general trends: First, violent crime trended downward all across Philadelphia from 2010 to 2014. Second, the tracts within the Promise Zone saw continuous declines in violent crime despite the leveling out in Philadelphia at large. Combined with the causal inference methods used in this paper, the trends present a compelling case for the general effectiveness of the Promise Zone in reducing violent crime.

Many advocates for the program and groups involved with its implementation reported de-
creased crime rates in the areas where they performed their outreach. This is perfectly in line with the results from this study. In that case, not only are these results statistically significant, they are economically significant and have implications for the city as a whole. Properly capturing the counterfactual of the Promise Zone is difficult, but the Mayor’s Office of Community Empowerment and Opportunity and the many community groups and citizens involved seem to be meeting their goal of reducing violent crime in this pocket of West Philadelphia and making it a more livable place.

These results suggest that Promise Zones may be effective at reducing certain types of violent crime. Since the Promise Zone program provides federal coordination and fast-tracking of grants but does not guarantee any specific set of grants, implementation of the program may vary greatly across cities. Therefore, Promise Zone programs should be studied on an individual basis to determine the strengths and weaknesses of different cities’ approaches. The findings from Los Angeles in Kitchens and Wallace (2021) indicate that not all Promise Zone programs reduce violent crime. Given the violent-crime-reducing effects of the Philadelphia Promise Zone, policy makers interested in reducing violent crime could look to the Philadelphia Promise Zone’s implementation for guidance.
Figure 1.1: Philadelphia’s Census Tracts, with the Promise Zone border in red.
Table 1.1: Crime incidents in the Promise Zone

<table>
<thead>
<tr>
<th>Violent Crime Types</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robbery w/ Firearm</td>
<td>820</td>
</tr>
<tr>
<td>Robbery w/o Firearm</td>
<td>1121</td>
</tr>
<tr>
<td>Rape</td>
<td>334</td>
</tr>
<tr>
<td>Criminal Homicide</td>
<td>83</td>
</tr>
<tr>
<td>Arson</td>
<td>134</td>
</tr>
<tr>
<td>Aggravated Assault w/ Firearm</td>
<td>626</td>
</tr>
<tr>
<td>Aggravated Assault w/o Firearm</td>
<td>1686</td>
</tr>
<tr>
<td>Other Assaults</td>
<td>5755</td>
</tr>
<tr>
<td>Other Sex Offenses</td>
<td>336</td>
</tr>
</tbody>
</table>
Figure 1.3: The Promise Zone in OpenStreetMap
Table 1.2: Summary Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post</td>
<td>3,750</td>
<td>0.600</td>
<td>0.490</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Zone</td>
<td>3,750</td>
<td>0.021</td>
<td>0.145</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Violent</td>
<td>3,750</td>
<td>28.542</td>
<td>20.994</td>
<td>0.673</td>
<td>251.557</td>
</tr>
<tr>
<td>Simple/Other Assault</td>
<td>3,750</td>
<td>16.378</td>
<td>12.115</td>
<td>0.000</td>
<td>155.666</td>
</tr>
<tr>
<td>Agg. Assault</td>
<td>3,750</td>
<td>3.874</td>
<td>3.261</td>
<td>0.000</td>
<td>31.142</td>
</tr>
<tr>
<td>Agg. Assault Firearm</td>
<td>3,750</td>
<td>1.525</td>
<td>1.731</td>
<td>0.000</td>
<td>27.233</td>
</tr>
<tr>
<td>Robbery</td>
<td>3,750</td>
<td>2.756</td>
<td>3.019</td>
<td>0.000</td>
<td>51.059</td>
</tr>
<tr>
<td>Robbery Firearm</td>
<td>3,750</td>
<td>1.943</td>
<td>1.868</td>
<td>0.000</td>
<td>22.000</td>
</tr>
<tr>
<td>Homicide</td>
<td>3,750</td>
<td>0.200</td>
<td>0.349</td>
<td>0.000</td>
<td>4.702</td>
</tr>
<tr>
<td>Rape</td>
<td>3,750</td>
<td>0.736</td>
<td>0.797</td>
<td>0.000</td>
<td>10.000</td>
</tr>
<tr>
<td>Arson</td>
<td>3,750</td>
<td>0.311</td>
<td>0.472</td>
<td>0.000</td>
<td>5.758</td>
</tr>
<tr>
<td>Other Sex Offenses</td>
<td>3,750</td>
<td>0.817</td>
<td>1.032</td>
<td>0.000</td>
<td>19.231</td>
</tr>
<tr>
<td>Non-Violent</td>
<td>3,750</td>
<td>81.985</td>
<td>78.626</td>
<td>7.833</td>
<td>1,189.904</td>
</tr>
<tr>
<td>Property</td>
<td>3,750</td>
<td>36.071</td>
<td>41.154</td>
<td>2.373</td>
<td>721.154</td>
</tr>
<tr>
<td>Population</td>
<td>3,750</td>
<td>4,124.635</td>
<td>1,701.694</td>
<td>173</td>
<td>9,510</td>
</tr>
<tr>
<td>Children under 18</td>
<td>3,750</td>
<td>918.832</td>
<td>617.587</td>
<td>0</td>
<td>3,887</td>
</tr>
<tr>
<td>Population 25+</td>
<td>3,750</td>
<td>2,713.426</td>
<td>1,113.781</td>
<td>71</td>
<td>6,149</td>
</tr>
<tr>
<td>Population 16-64</td>
<td>3,750</td>
<td>2,786.136</td>
<td>1,135.092</td>
<td>68</td>
<td>7,375</td>
</tr>
<tr>
<td>% White</td>
<td>3,750</td>
<td>0.418</td>
<td>0.322</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>% Black</td>
<td>3,750</td>
<td>0.437</td>
<td>0.353</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>3,750</td>
<td>0.116</td>
<td>0.169</td>
<td>0.000</td>
<td>0.917</td>
</tr>
<tr>
<td>% Single mother children</td>
<td>3,750</td>
<td>0.445</td>
<td>0.263</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>% Males 15 to 29</td>
<td>3,750</td>
<td>0.123</td>
<td>0.065</td>
<td>0.022</td>
<td>0.534</td>
</tr>
<tr>
<td>% Less than high school</td>
<td>3,750</td>
<td>0.180</td>
<td>0.111</td>
<td>0.000</td>
<td>0.629</td>
</tr>
<tr>
<td>% High school</td>
<td>3,750</td>
<td>0.331</td>
<td>0.125</td>
<td>0.000</td>
<td>0.599</td>
</tr>
<tr>
<td>% Some college</td>
<td>3,750</td>
<td>0.221</td>
<td>0.075</td>
<td>0.025</td>
<td>0.694</td>
</tr>
<tr>
<td>% Bachelor</td>
<td>3,750</td>
<td>0.151</td>
<td>0.104</td>
<td>0.000</td>
<td>0.528</td>
</tr>
<tr>
<td>% Grad</td>
<td>3,750</td>
<td>0.117</td>
<td>0.132</td>
<td>0.000</td>
<td>0.763</td>
</tr>
<tr>
<td>% No work 12 months</td>
<td>3,750</td>
<td>0.326</td>
<td>0.130</td>
<td>0.038</td>
<td>0.865</td>
</tr>
</tbody>
</table>

Observations are tract-year. Crime is per 1k residents.
Figure 1.4: Violent Crime Trends
Table 1.3: Difference-in-Differences for Violent Crime

<table>
<thead>
<tr>
<th>Dependent variable: V. Crimes per 1k Tract Residents</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post x Zone</td>
<td>$-4.764^{***}$</td>
<td>$-4.764^{**}$</td>
<td>$-4.177^{***}$</td>
<td>$-4.331^{**}$</td>
</tr>
<tr>
<td></td>
<td>(0.638)</td>
<td>(1.487)</td>
<td>(1.306)</td>
<td>(1.540)</td>
</tr>
<tr>
<td></td>
<td>[0.006]***</td>
<td>[0.04]**</td>
<td>[0.04]**</td>
<td>[0.04]**</td>
</tr>
<tr>
<td>Post</td>
<td>$-3.213^{***}$</td>
<td>-1.879</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.669)</td>
<td>(1.185)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zone</td>
<td>8.680**</td>
<td>-3.495</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.773)</td>
<td>(3.650)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.777</td>
<td>-2.206</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.362)</td>
<td>(8.570)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.341***</td>
<td>0.404</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.675)</td>
<td>(8.945)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.080</td>
<td>-6.266</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.300)</td>
<td>(7.840)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Mother</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.311</td>
<td>2.201</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.498)</td>
<td>(2.776)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male 15 to 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28.488*</td>
<td>-25.943**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(16.261)</td>
<td>(9.312)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than Highschool</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.553</td>
<td>8.307</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(18.326)</td>
<td>(6.962)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.732</td>
<td>5.827</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9.846)</td>
<td>(6.303)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some College</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$-53.645^{***}$</td>
<td>6.502</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(12.669)</td>
<td>(4.691)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$-27.104$</td>
<td>10.994</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(17.983)</td>
<td>(6.752)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0002</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No work 12 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>46.500***</td>
<td>-7.214</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(11.188)</td>
<td>(7.299)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>30.345***</td>
<td>0.147</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.255)</td>
<td>(13.122)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TWFE?</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>3,750</td>
<td>3,750</td>
<td>3,750</td>
<td>3,750</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.008</td>
<td>0.886</td>
<td>0.361</td>
<td>0.888</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.007</td>
<td>0.873</td>
<td>0.359</td>
<td>0.875</td>
</tr>
<tr>
<td>Residual Std. Error</td>
<td>20.920</td>
<td>7.474</td>
<td>16.809</td>
<td>7.428</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01
Std. Errors Clustered at tract & year level
Figure 1.5: Event Study

Event study: Effect of treatment on violent crime over time
Figure 1.6: Event Study with Staggered Rollout
Figure 1.7: Disaggregated Event Study
Figure 1.8: Synthetic Difference-in-differences with Covariates
Figure 1.9: Highest Level of Education Attained by Tract Residents
Table 1.4: Two Period Difference-in-Differences for Education (2013 vs. 2019)

<table>
<thead>
<tr>
<th>Dependent variable: Highest Level of Education</th>
<th>No Diploma/GED (1)</th>
<th>High School (2)</th>
<th>Some College (3)</th>
<th>Bachelor’s (4)</th>
<th>Graduate (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post</td>
<td>-0.034***</td>
<td>-0.017***</td>
<td>-0.004</td>
<td>0.032***</td>
<td>0.023***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Post × Zone</td>
<td>-0.029*</td>
<td>0.019</td>
<td>0.007</td>
<td>0.030</td>
<td>-0.027</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.018)</td>
<td>(0.024)</td>
<td>(0.019)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Tract FE?</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Observations</td>
<td>750</td>
<td>750</td>
<td>750</td>
<td>750</td>
<td>750</td>
</tr>
<tr>
<td>R²</td>
<td>0.920</td>
<td>0.913</td>
<td>0.801</td>
<td>0.935</td>
<td>0.955</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.839</td>
<td>0.826</td>
<td>0.600</td>
<td>0.870</td>
<td>0.910</td>
</tr>
<tr>
<td>Residual Std. Error</td>
<td>0.044</td>
<td>0.053</td>
<td>0.049</td>
<td>0.038</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01
    Std. Errors Clustered at tract level.
Figure 1.10: Effect of GEAR UP grant on math PSSA.
Figure 1.11: Effect of GEAR UP grant on reading/ELA PSSA scores.
Figure 1.12: Effect of GEAR UP grant on science PSSA scores.
Chapter 2

Supervised Injection and NIMBY in Philadelphia: A Contingent Valuation Experiment

2.1 Introduction

Every year in the United States, tens of thousands of people die from drug overdoses. Shortly after World-War II, heroin, a drug synthesized from morphine, caused the first “heroin epidemic” in the United States. Two decades later, heroin returned in the veins of disaffected Vietnam veterans (Hughes and Rieche, 1995). Then, in the 1990s, drugs that tapped into the same pleasure receptors in the brain began being marketed and distributed throughout the United States en masse. These drugs, known as “prescription opioid pain-killers”, were other derivatives of morphine that began being prescribed by the millions for self-reported pain. Eventually, the American Pain Society (APS) developed standards for treating pain in Max et al. (1995) which heavily involved the use of these new drugs. The following year, in 1996, the President of the APS declared pain to be “the fifth vital sign” (Humble, 2016).

Over twenty years later, the United States has found itself in what could be known as the “third major heroin epidemic”. With millions of people addicted to some form of opioid painkiller, and hundreds of thousands of them being addicted to heroin, there seems to be more cause for alarm
than ever before (Surratt et al., 2013). According to the CDC, groups with historically low rates of heroin usage are seeing increases in usage among both sexes and all income levels. Women, privately insured individuals, wealthy individuals, and other previously less affected groups have seen substantial increases in intravenous drug (ID) use and overdose rates in the past two decades. A major issue that manifested in the late twenty-tens is the increased conditional probability of overdose (conditional on usage of heroin) related to the synthetic morphine derivative known as Fentanyl. This problem, combined with the long-standing and well-known problems of HIV and Hepatitis transmission among intravenous drug users has led “harm-reduction” advocates to propose an alternative solution that avoids the use of law enforcement. One proposed solution is the “Supervised Injection Facility”,\(^1\) or SIF. These facilities greatly expand on the services provided by Needle Exchange Programs. While they provide clean needles, they also have trained medical staff providing overdose reversal and strict monitoring services to promote the safety of the on-site users. At an SIF, the sharing of needles, exchange of drugs or money, and any “crime” other than injection of one’s own drugs is strictly forbidden and prevented by staff. In addition, SIFs generally collect non-identifiable information that can be useful to researchers and provide information on drug counseling and abatement resources to all users of the facility.

At this point in time, dozens of SIFs operate around the world and are recognized legally by their respective governments; in the United States, it is understood that there are unlicensed SIFs operating unknown to local law enforcement and one legally operating in New York as of 2022. In 2020, after a months-long legal battle with the state of Pennsylvania, Safehouse (a harm-reduction focused charitable organization) was slated to open the nation’s first authorized SIF in the neighborhood of Kensington, a heavily drug-stricken neighborhood in Philadelphia. However, residents of South Philadelphia (not Kensington, which is in North Philadelphia) were blindsided when they found that the site would open “next week” in their area without any round-table discussion involving the residents of the area. Public opinion, which had been relatively positive in Kensington, shifted rapidly as the immediacy of the information shock and lack of community debate hit local residents in South Philadelphia, far from Kensington (Feldman and Blumgart, 2020). During a press conference announcing the opening of the facility, one South Philadelphia

\(^1\)Also called safe injection sites, safe consumption sites, and supervised consumption spaces.
resident interrupted the Vice President of Safehouse with the following statement:  

“You blindsided us. You don’t sit there and live in that community. You don’t take your kids to the day care like I do. I care about what my children have to see at 6 and 10 years old that I have to explain hard drug addiction. This is unacceptable. And you were a sneak about it.”

Ultimately, the landlord for the proposed South Philadelphia site backed out given the immense community backlash and the plans were put on hold indefinitely (Feldman and Blumgart, 2020). This leads to an interesting question: how large is the perceived cost to residents of such a facility in their area? To the quoted resident, the cost seems tremendous. Without a contingent valuation (CV) method survey experiment, it is difficult to observe the level of discomfort a resident might feel from having to “explain hard drug addiction” to his or her children as they walk by a site like this in their daily activities. This provides a novel subject for a CV method experiment, which have most often previously been used to study environmental amenities such as windmills, greenery, and pollution.

There is a body of research studying the efficacy of these sites in a more general sense, (see Pinkerton (2011); Marshall et al. (2011); DeBeck et al. (2011)), along with systematic literature reviews (see Potier et al. (2014)), which find generally positive effects from these sites such as decreased disease transmission, fewer overdoses, fewer discarded needles and reduced street crime. Even a brief glance at the literature would suggest to many that negative externalities of heroin and other intravenous drug usage are internalized by the SIF when established. If this research is correct, then the residents of South Philadelphia (and Kensington) would have little to worry about from the opening of such a site. Kral and Davidson (2017) spent years studying an unsanctioned SIF in an undisclosed location in the U.S. They made several observations of interest to the efficacy of sites such as these:

• More than 90% of users reported they would have otherwise been injecting in public restrooms, the street, or parking lots.

• 67% reported recent unsafe needle disposal practices, which were curbed when using the facility.

• The rate of overdose is about the same as the standard, but all instances of overdose were averted by facility staff. No one died.

Unfortunately, convincing residents that this is true is a difficult task and it is unclear whether providing this type of information would convince them to accept the site in their area. Additionally, the benefits to other people, namely drug users, may not matter very much to a drug-free resident, especially one with children. The purpose of this paper is to discuss the economics of SIFs and study the challenges of creating an SIF in the United States through a contingent valuation experiment mimicking this specific case in Philadelphia. By experimentally studying individuals’ “willingness to accept” (WTA) the opening of a site near them, public policy and organizational strategies can be guided toward opening SIFs in cooperation with local communities. This paper attempts to uncover the effectiveness of an informational treatment, the heterogeneity with which an informational treatment may affect different demographic groups, and the perceived cost to individuals of a site like this opening in their vicinity. While there is ample research studying the effects of SIFs, there is (to my knowledge) only one other paper that uses a non-market valuation technique (in this case, discrete choice as opposed to CV) to study SIFs, Berrigan and Zucchelli (2022).

2.2 The Economics of Drug Abuse

What makes addiction unique in relation to other forms of consumption is that individuals do not necessarily choose to be addicted; once physical addiction occurs, however, the cost of abstinence in the short-term becomes very high. The “adjacent complementarity” of addiction was first developed as a theory by Boyer in 1983, then Iannaccone in 1986. In their models, they show how the present marginal utility of consumption is related to past consumption of the good. In this case, intravenous drugs are the good, and the marginal utility gets lower as tolerance is built through consistent consumption.

Becker and Murphy (1988) developed the “Theory of Rational Addiction”, which builds upon adjacent complementarity, arguing that the current need for drug consumption is a monotonically increasing function of previous drug use, particularly for “hard” drugs like heroin.
Becker and Murphy (1988) present the utility function

$$u(t) = u[y(t), c(t), S(t)] \quad (2.1)$$

where $c$ is consumption of an addictive (adjacently complementary) good and $y$ is a non-addictive numeraire good that only affects utility in the period it is consumed. Past consumption of $c$ affects current utility through a stock of “consumption capital” $S$. $S$ is analogous to a level of addiction: higher $c$ in the past causes larger $S$, which necessitates further larger levels of $c$ to prevent a penalty to utility. In other words, they argue that $\partial S_t / \partial c_{t-1} > 0$ and $\partial u / \partial S < 0$.

Building on this model, we can assume that an addicted individual receives positive utility from heroin and consumption of other goods such as food and shelter, while they receive negative utility from arrest, overdose, or withdrawal. Someone in a deep stage of addiction would likely be less concerned with arrest or overdose than someone who is in an early stage of their addiction and consumes less heroin (Har, 2011). In the described situation, the risks associated with heroin use will, on the margin, deter some individuals from partaking. Repeated use dulls an individual’s response to the associated risks, which could explain why law enforcement has failed to curtail decades of drug abuse in the United States.

Withdrawal may occur when $c_t = 0$ and $S_t > 0$; arrest often leads to withdrawal as many jails do not provide Suboxone treatment to addicts (Mitchell et al., 2009). The risk of heroin tainted with Fentanyl adds a new complication to this model: suppose there is some additional parameter $\omega$ such that

$$u(t) = u[y(t), c(t), S(t), S(t), \omega(S(t), c(t))] \quad (2.2)$$

where $\omega$ is an expected cost of arrest or overdose as a function of addiction level $S_t$ and immediate heroin usage $c_t$. While using at an SIF, $\omega$ can be assumed to be zero, thus increasing the marginal utility of consumption. Considering the marginal utility of heroin consumption, we can derive demand; by increasing the marginal utility of heroin use by removing some level of risk, we are increasing the demand for heroin. Thus spawns a possible moral hazard with creating a safe space for individuals to use drugs. Removal of these risks may incentivize higher consumption of heroin,
all else being equal.

A basic understanding of externalities tells us that reducing the private marginal cost of something with a near-zero socially optimal level of consumption is generally bad for society. On the other hand, an SIF is technically internalizing many of the costs of usage by handling the problem privately. One could argue that, despite the negative externalities of heroin consumption, there is an internalization of negative externalities when supplying harm-reduction services. Safehouse Philly is fully privately funded and would be saving both local hospitals and, ultimately, taxpayers a hefty bill in providing naloxone and monitoring services. The question locals would have to contend with is whether the internalization of costs by an SIF outweighs the possible uptick in risky drug behavior due to the aforementioned change in marginal costs and benefits. Additionally, SIFs requiring disposal of needle trash on-site could prevent littering of dangerous needles. Locals, such as the aforementioned mother of children, likely imagine gatherings of drug-addicted individuals taking place directly outside this facility with needle trash littering the sidewalk. A major step in getting a local community to accept the opening of an SIF is to convince them that the negative externalities of drug use will be internalized by the facility, and not spill over into the community.

Consider a scenario where the user of the drug is both the demander and the supplier, experiencing a marginal benefit from use (depicted by their demand curve) and a private marginal cost in terms of the risk of overdose or arrest. By reducing the risk with an SIF, we are reducing the private marginal cost to the user and shifting their personal supply curve to the right, increasing the quantity consumed. Recovery Center of America estimates that productivity losses from illicit drug use tally up to hundreds of billions of dollars before the cost of healthcare and legal enforcement are even considered. With this in mind, there is an inelastic, high “social marginal cost” curve implying a near-zero optimal quantity of heroin consumption. The social marginal cost curve includes the risk to the individual (the private marginal cost) plus the external harms caused by heroin use such as strain on the healthcare system, discomfort to nearby residents, needle trash, opportunity cost for first responders, and legal enforcement.

On the other hand, Safehouse is privately funded and would be internalizing some of the social costs associated with heroin abuse. This would reduce the negative externality associated with heroin use. The crux of the political debate over SIFs comes down to the following issue: either Safehouse will internalize the negative externalities and lower the social costs associated with heroin
abuse, or it will encourage more risky behavior and promote addiction. Economic theory suggests that both are true; however, if the reduction in external cost is larger in magnitude than the reduction in private cost, then the welfare loss from heroin use is reduced. The randomized part of this paper’s experiment, where respondents are randomly given extra information about the lack of negative externalities from an SIF, will prime respondents with a belief about a reduction in negative externalities.

2.3 Literary Context

2.3.1 NIMBY

The “not-in-my-backyard” phenomenon is one that has plagued the provision of public goods and common resources for many years. Many public goods and common resources, such as fire stations, landfills, prisons, and homeless shelters require land. In requiring land to operate, they often end up in someone’s “backyard”, so to speak. Society benefits as a whole from certain goods that impose costs on households in their direct vicinity. The dilemma faced by those in the vicinity has also been referred to as the “volunteer’s dilemma” (Diekmann, 1985).

A Supervised Injection Facility arguably falls within this category because it keeps “hard” drug use off the streets of Philadelphia while simultaneously concentrating drug users in a small area directly bordering many residents’ living spaces. Despite the fact that the literature suggests that facilities such as these actually reduce negative externalities related to drug use, it is the perception of local residents that matters for the sake of this experiment. Perceiving that the neighborhood is more dangerous due to this facility can cause psychological turmoil and is a cost incurred by nearby residents, whether or not it is empirically true.

NIMBY is not strictly the act of saying “I do not want this in my backyard”, but saying that while generally supporting its existence in “someone else’s backyard”. Most of us are, at the very least, NIMBYs in the sense that we believe that landfills are a necessary common resource, but do not want them placed anywhere near our homes. This phenomenon would embody itself in this experiment as people supporting the existence of an SIF in an area they never interact with, but

\footnote{Which is the argument in the preceding paragraphs, namely that an SIF internalizes the externality, reducing external cost, and makes it less personally costly to use drugs given the reduced risk of death.}
opposing it when it is proposed close to their home. Even if NIMBY sentiment is not found in this experiment, (such as in Boyle et al. (2019)), it is still useful to understand what a resident’s perceived cost is of a nearby SIF.

NIMBY is a well-studied phenomenon in regard to (dis)amenities such as wind energy development and social services for homelessness. According to Linton et al. (2013), homelessness and intravenous drug use are highly correlated; for that reason, NIMBY research regarding services for the homeless (which are much more common than services for drug users) is highly relevant. What papers such as these find is that citizens who would otherwise be unperturbed by a social program often do not want it “in their backyard”. In the case of homeless shelters, state and federal governments have often had to step in and override local municipalities to get shelters made (Oakley, 2002). Yang and Beletsky (2020) find that there is still immense social stigma related to SIFs, which they attribute to the decades-long “War on Drugs”. The authors argue that, even when legal authorization is given, public understanding and support are both crucial in getting these sites established. This is, of course, an empirically testable claim that will be discussed in this experiment. The aforementioned paper is the only other paper currently in the literature to mention Safehouse by name.

Takahashi (1997) also studies how social stigma leads to the rejection of community services, specifically for homelessness and HIV/AIDS. Once again, there is a non-trivial intersection between individuals affected by homelessness, HIV, and intravenous drug addiction. It discusses three primary factors of stigma: non-productivity (“these people are lazy”), dangerousness (“they may rob or terrorize the locals”), and personal culpability (“it’s their fault they’re like this”).

### 2.3.2 Contingent Valuation and Public Goods Provision

For the experiment, one paper I follow is Grootwuis et al. (2008), which sits at the intersection of studying NIMBY and “willingness to accept” of a public good. The authors study the determinants of NIMBY in regards to building windmills that block scenic views for local residents. They use a willingness to accept framework in a survey form, and find that individuals who view wind energy as a “clean source of power” have a lower WTA, while individuals who do not hold that view have a higher WTA. A much more recent application of the “NIMBYism with windmills” idea is Boyle

---

4 The amount of payment required to accept a disamenity.
et al. (2019), which finds no NIMBY effect in siting wind farms. They find that the majority of respondents who support wind power will accept the windmills without compensation, while those who oppose wind power generally oppose windmills in their viewshed regardless of compensation. This is significant in that it demonstrates a total rejection of NIMBY: people who support this particular good are generally okay with it being in their backyard.

There is also a rich literature in experimental designs involving altruism and empathy. This experiment attempts to uncover an “impure altruism” preference in the respondents, an idea pioneered by Andreoni (1990) in the “warm-glow” model. The warm-glow model posits that as opposed to people being completely selfless in their altruism, there may be a selfish pleasure gained through altruistic acts. Andreoni et al. (2017) devised an experiment where they placed Salvation Army bell-ringers at various exits of a supermarket, varying the ease with which a shopper could avoid the bell-ringer. Upon adding a verbal ask, both avoidance and giving became more common. In other words, people may try to avoid empathetic situations to reduce the costs they have to incur to help others. People feel good (or mitigate guilt) when they help others, but they are also selfish.

This paper finds itself in a unique niche in the NIMBY and CV method literature, since it studies a type of “disamenity” that is relatively new and unknown to most people in the United States. To my knowledge, the only other paper in the literature which uses an experimental design to study willingness to accept a Supervised Injection Facility is the recent Berrigan and Zucchelli (2022), which examines this issue in Canada, where SIFs are legal. Along with their research, this paper sits on the cutting edge of a relatively unstudied phenomenon. This paper complements theirs by employing the CV method as opposed to the discrete-choice method.

2.4 Study Design

This study aims to quantify the perceived costs to locals in regard to the establishment of an SIF. I also seek to understand how an informational nudge priming the benefits of an SIF may affect someone’s willingness to accept.
2.4.1 The Goals

There are several questions this paper seeks to answer, and they are as follows:

- How does the distance and/or perceived interaction rate with the site affect people’s willingness to accept?
- How does pre-existing familiarity of SIFs affect someone’s WTA?
- How does presentation of a blurb about peer-reviewed research regarding the efficacy and community benefits of an SIF affect someone’s openness to it?
- How do demographic and household characteristics affect someone’s openness?
- How do the above factors interact with a person’s willingness to accept, measured by a subsidy that the supposed SIF would pay to them as compensation?

2.4.2 The Subject Pool

To study this phenomenon, I surveyed several hundred Philadelphians using the Dynata online panel and the Qualtrics survey platform during August of 2021. I followed the established standards for CV method surveys laid out in papers such as Boyle et al. (1985), Whitehead (2006), Groothuis et al. (2008) and Boyle et al. (2019). Dynata is considered an “actively managed” data collection service, which means they “recruit and verify the identity of panel members then use statistical techniques to draw samples” (Arndt et al., 2021). While respondents are kept fully anonymous to the researcher, Dynata prevents double-dipping by respondents and allows for more accurate targeting of geographic areas (in this case, a relatively diverse sample of Philadelphians).

Screening Criteria

There is a general consensus in the survey literature that the quality of screening can greatly influence statistical results of survey experiments (Huang et al., 2012; DeSimone and Harms, 2018; Arndt et al., 2021). There are several techniques used to screen out low-quality respondents who may not be motivated to provide accurate responses. For example, some respondents, known as “speeders”, move quickly through the survey to receive their reward and do not fully absorb the
information being given to them. Even if a respondent is not a speeder, they still may not be fully absorbing the information and may contradict themselves at various points in the survey. One method of screening involves providing opportunities for respondents to logically contradict themselves in ways they would not had they been paying attention (Leland Wilkinson and the Task Force on Statistical Inference, 1999; Huang et al., 2012).

Several checks of consistency and understanding filtered out unreliable respondents in this experiment. Respondents were asked about their age and income early in the survey, choosing an age bracket and answering whether their income was above or below $70,000. Later on in the survey, respondents are asked to slide a a bar to their year of birth and to choose a specific income bracket. A contradiction here shows that, either early in the survey or later in the survey, the respondent is not paying close attention or answering dishonestly. If these did not align, they were removed (e.g. a respondent answers that they are 25-34 but then chooses 1999 as their birth year, meaning they were 21 or 22). In the case of this experiment, if a respondent opposes an SIF opening at some distance to them, but supports it when it is on their block, I assume they are not a trustworthy respondent and that their response may be biasing the results. This is based on the assumption that respondents become less likely to show support as the interaction rate of the SIF increases and the distance lessens, which is consistently true with the vast majority of respondents.

Another problem can arise when a respondent who is treated with information does not actually read the information. Any respondent in the treated group who responded “No, I did not read it.” when asked whether they read the treatment blurb, were counted as not being treated, but only removed completely if they failed one of the other consistency or speed checks. When using a between-subject design such as this, the randomized treatment is the most important of the study, from which claims about causality can actually be made. Therefore, it is crucial that I assure the “treated” group is actually treated, or else the results could be meaningless. The second to last debriefing question at the end of the survey asks respondents, using a Likert Scale, to rate their understanding of the hypothetical situations offered in the survey. If they answered “Strongly Disagree” or “Somewhat Disagree”, they were removed. This is one of the only places in the survey

---

5In additional robustness checks, an additional indicator is added to account for these individuals as their own group.
6For example, I cannot make any strong claims about demographic determinants of support since there is endogeneity in selection into Dynata. However, the randomly assigned treatment does not experience this problem and we can make causal claims regarding it.
where respondents have the ability to skip the question. With that in mind, any respondent who skipped this question is also considered to be unreliable, (consciously skipping a question as soon as they have the opportunity demonstrates lack of motivation to answer), and is removed from the results.

This screening process results in 412 responses before any duration criteria is applied for screening. As Arndt et al. (2021) points out, survey speeds approximating a normal distribution can use a relatively standardized method for removing outliers; the problem is that survey speeds among respondents often resemble a Chi-Square distribution, as shown by Figure 2.1.

![Figure 2.1: Distribution of Durations](image)

The survey was conducted using the Qualtrics survey platform. According to Qualtrics, the
survey would take approximately 12.7 minutes to complete. Following Arndt et al. (2021), I consider anyone taking less than 20% of that time, two minutes and thirty seconds, to be speeders. Speeders are suspected to be rushing through to receive their monetary reward, which Dynata claims is approximately $2 per response. While they are substantially less common than speeders, I consider anyone taking more than 1500 seconds to be a “slowpoke”. While some of these individuals may have opened the first page of the survey, gotten sidetracked, and returned, it is also possible that they were sidetracked part-way through the survey. If that is the case, then they are not giving their full attention. Suppose they get sidetracked some time after the information treatment to where they return 20 minutes later; without the information fresh in their heads, they are essentially of a different treatment status from the other treated respondents, possibly biasing their answers. Therefore, speeders and slowpokes are removed from the sample that will be analyzed for the sake of this project, resulting in the slightly more attractive distribution shown in Figure 2.2.

2.4.3 The Design

The first thing a respondent sees when they open the survey is the page depicted in Figure 2.3. This page assures respondents that their truthful responses will be kept confidential and that they are helping further impactful research for public policy. Even though Dynata qualifies the survey to the respondents as taking up to 15 minutes –this is the amount of time they are compensated for– preliminary testing by friends, colleagues, and myself suggested that the survey would take approximately 5 to 8 minutes, which is demonstrated by Figure 2.1. The goal of this nudge is to encourage respondents to take their time, since the survey will take substantially less time than they are being paid to provide.

First, respondents are screened by being asked if they are from Philadelphia, which is further confirmed by a question about their zip code. Respondents who respond “No” or provide a zip code outside of Philadelphia are screened out immediately. They are then asked to indicate their age. If they respond “Under 18” to the age question, they are screened out from the survey immediately. They are then asked whether their household income in 2020, before taxes, was above or below $70,000. This is part of the screening process described above, as it is later cross-checked at the end of the survey.

Then, a series of preliminary questions are asked of the participants. The first is as follows:
Have you ever witnessed needle trash or suspected heroin use in public in Philadelphia?

With the options being “Yes, very recently”, “Yes, but not recently”, “No, never”, and “I am not sure”.

The second and third preliminary questions are as follows:

_Are you familiar with the existence of Methadone Clinics and Needle Exchange Programs?_

_Are you familiar with the existence of Supervised Injection Facilities?_

While the answers to these questions are not confounders, as they have no effect on treatment assignment, they could serve as partial mediator variables on the causal pathway from treatment to outcome.
Then, half the participants are randomly assigned to be given the following information.

*A rich body of peer-reviewed research by both academic and governmental organizations suggests that Supervised Injection Facilities reduce needle trash, street crime, disease transmission, and overdose levels in surrounding areas. There is little to no evidence that these facilities increase crime or drug use in surrounding areas.*

This information is essentially a paraphrase of the findings of the aforementioned Potier et al. (2014), Kral and Davidson (2017) and others. The purpose of this is to test how additional information about research may change an individual’s willingness to accept the facility. This behavioral nudge attempts to prime the “impure altruism” preference in respondents by attempting to frame the benefits of the site as improving conditions in their immediate area. Instead of telling them that the site will simply reduce disease and overdose levels, (problems that are likely irrelevant to respondents, but relevant to users of the site), the nudge additionally tells them that there is evidence that the site reduces needle trash and street crime. Of course, the latter may still be irrelevant and difficult to imagine for someone living in a high income area of Philadelphia. A selfish economic actor living in a high income area that does not experience needle trash or drug-related street crime should, in theory, be less responsive to this type of nudge than someone living
in Kensington, for example, where these problems are much more salient.

All participants are then asked the following attitudinal question:

A “Supervised Injection Facility” would allow intravenous drug users to use drugs under medical supervision. Any other illicit activity, such as exchanging (selling) of drugs would be strictly forbidden at this facility.

The facility is run by a private charity. There are approximately 100 of these facilities across Canada and Europe, but none in the US due to legal issues. In your opinion, should a facility like this be legally allowed to operate in the US?

The options are given as a Likert Scale from Definitely Yes to Definitely Not.

Then, the questions move into NIMBYism, asking the following question:

Suppose a non-profit would like to open a “Supervised Injection Facility” in Philadelphia. However, the non-profit wants to know if you would be willing to accept this facility in your community.

The Supervised Injection Facility would be several miles away from your neighborhood. You would never see it or drive by it in your regular activities.

Would you support the facility opening?

For both the information treated group and the control group, there are three different levels for each of the attributes:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Levels</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>“Several miles away from neighborhood”</td>
<td>Far</td>
</tr>
<tr>
<td></td>
<td>“Within walking distance”</td>
<td>Near</td>
</tr>
<tr>
<td></td>
<td>“On your block, visible from your residence”</td>
<td>Block</td>
</tr>
<tr>
<td>Interaction Rate</td>
<td>“Never”</td>
<td>Never</td>
</tr>
<tr>
<td></td>
<td>“When working or doing regular activities”</td>
<td>Frequent</td>
</tr>
<tr>
<td></td>
<td>“Any time leaving residence”</td>
<td>Always</td>
</tr>
</tbody>
</table>
• Far/Never
• Far/Frequent
• Near/Never
• Near/Frequent
• Block/Always

The options are a Likert Scale ranging from Strongly Support to Strongly Oppose. Upon answering, a follow-up question was asked in the order described by Figure 2.4. In the case that a respondent answered “Strongly Oppose” to Far/Never or Near/Frequent, questions are skipped to save time for the respondents and reduce the likelihood of rushing.\footnote{This rests on an assumption I am making that someone who strongly opposes Far/Never would rationally strongly oppose anything closer or more frequently interacted with. The change from Far/Frequent to Near/Never is ambiguous. Once again, if someone strongly opposes Near/Frequent, I assume I will not receive any new information from them for Block/Always.}

\begin{itemize}
  \item Far/Never
  \item Far/Frequent
  \item Near/Never
  \item Near/Frequent
  \item Block/Always
\end{itemize}

Figure 2.4: Order of the Openness Questions
Upon completion of this line of questioning, the respondents are then given the same prompt as the fifth level, with the addition of a monetary compensation offer:

*Suppose the Supervised Injection Facility is planning to open on your block and would be visible from your residence. You would see it or walk by it any time you leave your residence. Suppose the Supervised Injection Facility is willing to financially compensate anyone on the same block.*

*If they offered you a monthly stipend of $X, would you accept the facility’s opening on your block?*

$X starts at $200, $500, or $3000 with even probabilities. Anchoring bias is heavily discussed in the CV method literature (see Green et al. (1998), Whitehead (2000), Flachaire and Hollard (2006), Flachaire et al. (2007)); some of the proposed methods for abating this bias involve asking follow-up valuation questions, varying the starting point or allowing respondents to write out any value they desire. I employ the first two of the aforementioned techniques. One of the theories behind why anchoring bias occurs in these types of surveys is that, in the case of bidding for a good, a respondent may consciously or unconsciously believe that the first value they see is the “true” value for something. This is less applicable when we are asking them to subjectively value a personal cost as opposed to a good which may have an “objective” market price, but is nonetheless worth controlling for. Green et al. (1998) discusses the problem of starting-point bias created by researcher choice creating inconsistency across CV research; this is especially relevant when considering how Berrigan and Zucchelli (2022) used one-time lump-sum offers of $0, $25,000, and $50,000 while I start my respondents at a monthly stipend of $200, $500, or $3000. As an additional robustness check, I include a model with starting-point fixed effects in the appendix.

A Likert Scale is not provided for these questions; instead, the respondents’ only options are “Yes” or “No”, essentially compelling them to “pick a side”. If the respondent answers “Yes”, they are iteratively asked the same question with one hundred fewer dollars (except in the case of the $3000 starting point, where the next option downward is $1000) in each iteration until they say “No”. If they say “No” to the first amount offered, the amount is iterated up by $100 until the respondent says “Yes” (except in the case that the respondent is offered $3000, where there is no higher option). The goal of this line of questioning is to estimate a minimum amount each
respondent would accept. The starting amount is randomly distributed among $200, $500, and $3000. The other values after the initial question range from $100 to $1000 in increments of $100. If a respondent answered “Strongly Support” or “Somewhat Support” to Block/Always before the valuation section began, they are considered to have a minimum WTA of $0.

After the valuation portion of the survey, more questions are posed to the respondent. They are asked whether a friend or family member has ever been addicted to opioids or intravenous drugs (this was held until after the main questions to prevent a possible emotional nudge), their year of birth, their racial identity, their gender, their income bracket, their home ownership status (own, rent, or live with relatives), size of their household, number of children, whether they live or work in Kensington, and their highest level of education.

Finally, respondents are asked debriefing questions asking whether they “understand all of the information presented about the hypothetical situations provided throughout this survey” and whether they “believe the results of this survey will be shared with decision makers”. The full survey is available in the appendix.

2.5 Estimation and Results

2.5.1 Control variables collected from the survey

The wide range of demographic and attitudinal questions in the survey allows for a large set of controls. They account for many potential mediator variables that, when included, may make treatment point estimates more precise: gender, race, income, age, education, housing situation, addicted family member, witnessing needle trash, familiarity with needle exchange, familiarity with SIFs, living or working in Kensington, household size, and number of children under 18. These factor variables, with one removed in each case as a reference category, results in 57 control variables, as shown on the next page. Their frequency distributions are displayed in Figure 2.5.

2.5.2 Treatment Effect on Support

I specify an ordered logit model to measure the marginal effect of the information treatment on the outcome. I also transform my data to a “long” format, where there are five observations per
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>No. of Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male, Female, Non-binary, Prefer not to answer</td>
<td>4</td>
</tr>
<tr>
<td>Race</td>
<td>Respondents' combinations resulted in 16 unique answers concatenated down to: Black Only, White Only, Asian Only, Hispanic Only, Mixed, American Indian/Other/Prefer not to answer.</td>
<td>6</td>
</tr>
<tr>
<td>Household Income</td>
<td>Less than $10,000, $10k-$19.999k, $20k-$34.999k, $35k-$49.999k, $50k-$74.999k, $75k-$99.999k, $100k-$149.999k, $150k-$199.999k, $200k+, Prefer not to answer.</td>
<td>10</td>
</tr>
<tr>
<td>Age Bracket</td>
<td>18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+</td>
<td>8</td>
</tr>
<tr>
<td>Education Level</td>
<td>Less than high school, High School Diploma or GED, Some college but no degree, Associate's Degree (2yr), Bachelor's Degree (4yr), Master's Degree, Professional Degree (JD,MD), Doctoral Degree (PhD), Prefer not to answer.</td>
<td>9</td>
</tr>
<tr>
<td>Housing</td>
<td>Own, Rent, Live with parents/relatives, Prefer not to answer.</td>
<td>4</td>
</tr>
<tr>
<td>Addicted Family</td>
<td>“Has a friend or a family member ever been addicted to prescription opioids, narcotics, or other intravenous drugs?”: Yes, No, Prefer not to answer.</td>
<td>3</td>
</tr>
<tr>
<td>Seen Trash</td>
<td>“Have you ever witnessed needle trash or suspected heroin use in public in Philadelphia?”: “Yes, very recently”, “Yes, but not recently”, “No, never”, “I am not sure”</td>
<td>4</td>
</tr>
<tr>
<td>FamiliarNEP</td>
<td>“Are you familiar with the existence of Methadone Clinics and Needle Exchange Programs?”: Yes, No</td>
<td>2</td>
</tr>
<tr>
<td>FamiliarSIF</td>
<td>“Are you familiar with the existence of Supervised Injection Facilities?”: Yes, No</td>
<td>2</td>
</tr>
<tr>
<td>Kensington</td>
<td>“Do you live or work in the neighborhood of Kensington?”: Yes, No, I am not sure, Prefer not to answer.</td>
<td>4</td>
</tr>
<tr>
<td>Household Size</td>
<td>“How many individuals live in your household other than yourself?”: 0, 1, 2, 3, 4, 5+, Prefer not to answer</td>
<td>7</td>
</tr>
<tr>
<td>No. of Children</td>
<td>“How many children do you have who are under the age of 18?”: 0, 1, 2, 3, 4, 5+, Prefer not to answer</td>
<td>7</td>
</tr>
</tbody>
</table>
respondent, one for each answer\textsuperscript{8} they gave. It should be noted that, within the Likert Scale questions regarding the facility itself, respondents who answer “Strongly Disagree” are not shown follow-up questions in the scenario described by Figure 2.4. In other words, the structure I chose for my Qualtrics survey creates “attrition” in the respondent pool if they answer “Strongly Disagree” at certain points. Therefore, I perform my analysis on two datasets: a balanced panel of \(n = 382 \times 5\), where automatically skipped responses are considered “Disagree”, and an unbalanced panel where \(n < 382 \times 5\), where some respondents retain fewer than five observations due to automatically skipped questions. The ordered logit model takes on the form\textsuperscript{9}

\[
\text{logit}(p_{\text{StronglySupport}}) = \ln \frac{p_{\text{StronglySupport}}}{p_{\text{SomewhatSupport}} + p_{\text{Neither}} + p_{\text{SomewhatOppose}} + p_{\text{StronglyOppose}}}
= \beta_0 + \beta_1 \text{Far/Never} + \beta_2 \text{Far/Frequent} + \beta_3 \text{Near/Never} + \beta_4 \text{Near/Frequent} + \beta_5 \text{Treated} + \sum \beta_{n+1} X_{n+1} + \epsilon
\]  

(2.3)

where \(X\) is a vector of the aforementioned categorical control variables. The odds ratios and p-values from this ordered logit regression are displayed in Table 2.3. With odds ratios ranging from 1.174 to 1.3, this suggests that treated individuals are 17.4% to 30% more likely to respond with “Strongly Support” (as opposed to not) than individuals who are not treated with the extra information.

2.5.3 Heterogeneous Treatment Effects across Demographics

While the average treatment effect is positive, as demonstrated by the results, it is worthwhile to understand whether some groups were affected by treatment differently than others to help shine light on which subgroups may be most affected by messaging. I perform a heterogeneity analysis where I interact the treatment variable with the categorical variables for age, income, living situation (own, rent, live with relatives), whether they have had an addicted family member, working or living in Kensington, and education level following the model from Column (1) and (3)

\textsuperscript{8}Far/Never, Far/Frequent, Near/Never, Near/Frequent, and Block/Always.

\textsuperscript{9}Their answer, which is an ordered factor with the level of interest being “Strongly support”, is then regressed using the \textit{polr} function from the \textit{MASS} package in R on the treatment binary variable, all or none of the controls, and a set of dummy variables indicating which question they are answering, with “Block/Always” excluded as a reference category.
of Table 2.3. That being said, the many categories for age and income (some of which have fewer than ten respondents each) would result in low statistical power for these interactions. For that reason, I group income into “low” and “high” based on the “above/below” $70,000 question asked earlier in the survey. I categorize 18-34 as “young”, 35-54 as “middle-aged”, and 55+ as “older”. In each of these regressions, the confidence intervals for the interaction term point estimates include zero. Therefore, I am unable to claim that there is any kind of treatment effect heterogeneity. Treatment seems to uniformly affect the respondents across the seven categories examined.

### 2.5.4 Contingent Valuation

For the valuation, I consider the 86 respondents who supported the site on their block as having a “Willingness to accept” (WTA) of 0. In other words, these people do not require compensation for the site to be on their block.\(^{11}\)

Hanemann (1984) pioneered an equation for calculating mean and median WTP and WTA from the results of logit and probit models, suggesting that median is the superior measure. This method has become standard and is used throughout the literature (see Buckland et al. (1999); Buckley et al. (2008); Li et al. (2018); Chandra et al. (2020)). Following, Hanemann (1984), I estimate a repeated logit model specified as

\[
\ln \left( \frac{Pr(Yes)_{ik}}{1 - Pr(Yes)_{ik}} \right) = \beta_0 + \beta_1 DollarAmount_k + \sum \beta_{n+1} X_{n+1} + \epsilon
\]  

(2.4)

where \( \sum \beta_{n+1} X_{n+1} \) is a set of \( n \) control variables and their respective coefficients, \( i \) is the respondent, and \( k \) is the valuation question being asked.

Consider that the median WTA among respondents is the value for which the probability of a “Yes” is exactly 0.5.\(^{12}\) Furthermore, consider the functional form of the dependent variable in a logit model, namely that if \( Pr(Yes) = 0.5 \), then:

\[
\ln \left( \frac{Pr(Yes)}{1 - Pr(Yes)} \right) = 0
\]

(2.5)

This implies that the median WTA (or WTP) can be solved for such that

---

\(^{10}\) I choose not to include all other controls since treatment is randomly assigned.  
\(^{11}\) I relax this assumption in the appendix and still draw similar conclusions.  
\(^{12}\) This intuitive conceptualization comes from Buckley et al. (2008).
\[ \hat{\beta}_0 + \hat{\beta}_1 WTA + \sum \hat{\beta}_{n+1} X_{n+1} + \epsilon = 0 \]  

(2.6)

and median willingness to accept can then be measured as

\[ WTA = (\frac{-\hat{\beta}_0 - \sum \hat{\beta}_{n+1} X_{n+1} - \epsilon}{\hat{\beta}_1}) \]  

(2.7)

where whichever explanatory variables included, such as Income, along with the error term \( \epsilon \) take on their mean values. While the mean values of significant demographic covariates within the data allow us to solve for median WTA among the respondents, using the actual means for the public at-large within Philadelphia could provide a more generalizable estimate of median WTA.

While many respondents have bounded answers (for example, “No” to $300 but “Yes” to $200, bounding their WTA), others always respond no, even when offered large sums such as $1000 or $3000 per month as their hypothetical compensation. For a more informative analysis, it is important to consider different ways of conceptualizing their valuation. Therefore, I report several sets of results and WTA calculations, namely repeated measures logit models with their answer regressed on:

1. Offer value only.
2. Offer value and income categories.
3. Offer value, income, and treatment status.
4. Offer value, income, and treatment status with respondents shown $3000 removed completely.
5. Offer value, income, and treatment status with “Always No” respondents removed completely.

Reported in Table 2.5, the first three models are very easy to interpret. To calculate the median willingness-to-accept, we take the constant and divide by the coefficient on Offer. Model 3, which includes both controls for income and whether a respondent was treated, estimates a median WTA of $2588.

When respondents who always respond “No” are removed, the estimate is much lower. Eight respondents gave a “prefer not to answer” response; adjusting for the significance on this variable
results in an estimate of $135, while not adjusting for it provides an estimate of $113. It is notable that when the “Always No” respondents are removed, the median WTA falls by an order of magnitude. This shows the deep disconnect between individuals who are relatively open to the idea of a facility compared to those who will not accept it under any circumstances.\textsuperscript{13}

This is not necessarily incompatible with the results from Berrigan and Zucchelli (2022), who found a WTA of $11,500 CAD, which converts to roughly $8865 USD. The authors in that paper measured a hypothetical one-time lump sum payment while my paper poses the question as a monthly payment. This could be a case of hyperbolic discounting on the part of the respondents, who highly value the first few periods of payment but value later periods less-so (Thaler, 1981). Stevens et al. (1997) examined respondent sensitivity to payment schedules, namely one-time payment versus periodic payment, and found a very high implicit discount factor in their behavior. Lew (2018) reached a similar conclusion: future payments are discounted at a very high rate.

It should also be noted that there is literature examining the difference in valuations between WTA and WTP formats, generally suggesting that WTAs are larger than WTPs given the same intervention. Horowitz and McConnell (2002) find that this holds throughout the literature, and is most pronounced for situations involving non-market goods. The further a “good” was from being a typical good one could buy, the more WTA and WTP diverged. That being said, in the case of siting a waste site they found a smaller divergence than many of the other scenarios tested, a finding they attribute to the respondents’ perception of having a property right to decide the site’s existence. This is relevant to my survey, given that I present a “property right” about the SIFs existence to the respondents.

\subsection*{2.5.5 Deriving Supply Curves of Respondents}

In economic theory, a supply curve is derived from the minimum price a supplier will accept for a good or service that they provide. In theory of the firm, it is generally assumed that the minimum price a firm will accept, (often called “Willingness-to-Sell” or “Willingness-to-Accept”) is derived from the marginal cost to that firm of providing an extra unit of a good. In this case, instead of

\textsuperscript{13}It also demonstrates the bimodality of the data under the assumption that those who supported a facility on their block before being offered money have a WTA of $0. Upon being offered money, some of these respondents developed a bounded WTA above $0. An additional run of the model with this assumption relaxed is provided in the Appendix. The results are not very sensitive to this change.
firms we have Philadelphian respondents to the survey. Instead of a market price faced by them as sellers of a good, we have a potential monetary offer made to them to provide the "good" of "allowing an SIF on their block". By plotting the proportion of respondents answering "Yes" in relation to iteratively higher monetary offers, we can observe something akin to a supply curve (the only difference being that instead of discrete quantities we observe proportions of "Yes" responses, which could be mapped to quantities given some assumption about how many individuals would actually live on the same block as the SIF).

Two general behaviors are standard with supply curves. The first is that they should slope upward: an increase in the monetary offer should cause a higher proportion of respondents to answer "Yes", holding all else equal. The second is that anything which fundamentally reduces the cost (or increases the benefit) of an SIF in the mind of a respondent should shift the curve downward; a downward-shifted supply curve would show a higher proportion of "Yes" at any given money offer (or, a lower level of money required at any given proportion of "Yes").

The supply curves in Figure 2.6 depict the proportion of respondents who answered "Yes" to varying money offers regarding an SIF on their block, grouped by treatment status. As was implied from the ordered logit results, support is higher among respondents treated with the extra information; this is demonstrated by the "Treated" supply curve being to the right of the "Untreated" supply curve for all values. Figure 2.7 shows a mild downward shift of the supply curve for individuals who answered that they have a family member or friend who is or was addicted to drugs. The same is true in Figure 2.8, where respondents who were already familiar with SIFs require slightly less compensation. This is notably very similar to the shift in 2.6, suggesting that prior knowledge of SIFs may have been positive like the information treatment provided in the survey. Figure 2.9 suggests that older individuals require more compensation than younger individuals; this could be related to higher levels of income or more conservative viewpoints that are associated with age. Perhaps the most intriguing of the supply curves is Figure 2.10, which shows that half of the respondents who reported living or working in Kensington expressed support for the facility before being offered any monetary compensation. This is not surprising, as individuals who live or work in Kensington are likely well aware of the seriousness of public usage and discarding

14 It should be noted that amounts between $1000 and $3000 were not offered, and the portion of the curves on that range of y-values is a straight line connecting two observations.

15 I have been informed that "old" is merely a mindset, and I apologize to anyone 55 or older reading this.
of heroin needles. Individuals from Kensington are notably much more pro-SIF than any other subgroup in the survey, including the individuals treated with extra information.

2.6 Conclusion and Takeaway

These results provide a few relevant takeaways. First, a simple information treatment describing a research consensus about the safety and efficacy of these facilities provides a modest boost to respondents’ openness to an SIF (a 17 to 30% increase in probability of “Strongly Support”). If local residents can be shown convincing, accurate information explaining that an SIF will likely reduce the problem of open-air drug usage and needle trash (thus providing benefit to the non drug using resident directly), they will be less likely to protest. More outreach involving proliferation of this information to local residents could reduce public backlash. Despite this, the treatment does not seem to have a significant effect on WTA itself.

Second, the perceived cost to residents of an SIF on their block is tremendously high, with the majority of respondents who opposed a site on their block remaining steadfast in their opposition even when offered a hypothetical $1000 per month. Calculations suggest that the median WTA for an SIF on one’s own block is somewhere between $1800 and $2900 per month, but this is likely biased upward by a large number of respondents steadfastly opposing the site regardless of a monetary offer. With respondents who are “less stubborn”, so to speak, the WTA falls drastically: closer to $100 per month. Third, approximately half of the residents who support an SIF far away oppose its opening on their block, demonstrating immense “not-in-my-backyard” (NIMBY) sentiment. This creates a potential catch-22: people do not want SIFs in their area of residency or work, but an SIF in a more remote location would potentially be inaccessible to the individuals it attempts to serve.

Additionally, respondents who identified as living or working in Kensington, (the originally planned location of the SIF), were substantially more likely to support the facility opening on their block, suggesting that the same backlash would not have occurred had the plans not changed to South Philadelphia. While there could be something fundamentally different about Dynata respondents from Kensington and the public at-large in that area, it is certainly notable how much more supportive respondents from Kensington were. Moving forward, if Safehouse still wishes to
open a facility in Philadelphia, they need to recognize the immense fear that many residents have about the possibility of an SIF on their block. Safehouse should push for their facility not in South Philadelphia, but Kensington, at the heart of the problem. Here, they may find themselves welcomed with open arms into a community that has little to lose from cutting-edge harm-reduction services.

Table 2.1: Frequency table: responses with attrition.

<table>
<thead>
<tr>
<th></th>
<th>Oppose</th>
<th>Neither</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Far/Never</td>
<td>131</td>
<td>60</td>
<td>191</td>
</tr>
<tr>
<td>Far/Frequent</td>
<td>84</td>
<td>63</td>
<td>157</td>
</tr>
<tr>
<td>Near/Never</td>
<td>82</td>
<td>64</td>
<td>158</td>
</tr>
<tr>
<td>Near/Frequent</td>
<td>104</td>
<td>63</td>
<td>137</td>
</tr>
<tr>
<td>Block/Always</td>
<td>118</td>
<td>53</td>
<td>86</td>
</tr>
</tbody>
</table>

Table 2.2: Frequency table: responses without attrition.

<table>
<thead>
<tr>
<th></th>
<th>Oppose</th>
<th>Neither</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Far/Never</td>
<td>131</td>
<td>60</td>
<td>191</td>
</tr>
<tr>
<td>Far/Frequent</td>
<td>162</td>
<td>63</td>
<td>157</td>
</tr>
<tr>
<td>Near/Never</td>
<td>160</td>
<td>64</td>
<td>158</td>
</tr>
<tr>
<td>Near/Frequent</td>
<td>182</td>
<td>63</td>
<td>137</td>
</tr>
<tr>
<td>Block/Always</td>
<td>243</td>
<td>53</td>
<td>86</td>
</tr>
</tbody>
</table>
Figure 2.5: Frequency Distribution of Control Variables
Table 2.3: Effect of Information Treatment on Support

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated</td>
<td>1.189**</td>
<td>1.300***</td>
<td>1.174*</td>
<td>1.289**</td>
</tr>
<tr>
<td></td>
<td>p = 0.012</td>
<td>p = 0.005</td>
<td>p = 0.077</td>
<td>p = 0.014</td>
</tr>
<tr>
<td>Far/Never</td>
<td>3.140***</td>
<td>4.195***</td>
<td>1.575***</td>
<td>2.315***</td>
</tr>
<tr>
<td></td>
<td>p = 0.000</td>
<td>p = 0.000</td>
<td>p = 0.002</td>
<td>p = 0.00000</td>
</tr>
<tr>
<td>Far/Frequent</td>
<td>2.245***</td>
<td>2.696***</td>
<td>1.919***</td>
<td>2.490***</td>
</tr>
<tr>
<td></td>
<td>p = 0.000</td>
<td>p = 0.000</td>
<td>p = 0.0002</td>
<td>p = 0.000</td>
</tr>
<tr>
<td>Never/Near</td>
<td>2.292***</td>
<td>2.820***</td>
<td>2.032***</td>
<td>2.687***</td>
</tr>
<tr>
<td></td>
<td>p = 0.000</td>
<td>p = 0.000</td>
<td>p = 0.00001</td>
<td>p = 0.000</td>
</tr>
<tr>
<td>Frequent/Near</td>
<td>1.897***</td>
<td>2.194***</td>
<td>1.561***</td>
<td>1.928***</td>
</tr>
<tr>
<td></td>
<td>p = 0.00001</td>
<td>p = 0.00000</td>
<td>p = 0.004</td>
<td>p = 0.00003</td>
</tr>
<tr>
<td>Controls?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Attrition?</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>1,910</td>
<td>1,910</td>
<td>1,551</td>
<td>1,551</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 2.6: Supply curves by treatment status
Figure 2.7: Supply curves by addicted family member status

Figure 2.8: Supply curves by familiarity with SIFs
Figure 2.9: Supply curves grouped by age: 55 or older vs. younger than 55.

Figure 2.10: Supply curves grouped by whether respondent lives/works in Kensington.
Table 2.4: Responses to each money offer exposure.

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
<th>3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>296</td>
<td>13</td>
<td>84</td>
<td>87</td>
<td>128</td>
<td>123</td>
<td>126</td>
<td>123</td>
<td>119</td>
<td>100</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>86</td>
<td>25</td>
<td>38</td>
<td>25</td>
<td>36</td>
<td>56</td>
<td>20</td>
<td>18</td>
<td>25</td>
<td>30</td>
<td>53</td>
<td>38</td>
</tr>
</tbody>
</table>

Table 2.5: Repeated Measures Logit for WTA

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Answer = “Yes”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline (1)</td>
</tr>
<tr>
<td></td>
<td>Income (2)</td>
</tr>
<tr>
<td>Income &amp; Treatment</td>
<td>(3)</td>
</tr>
<tr>
<td>Always No Removed</td>
<td>(4)</td>
</tr>
<tr>
<td>First Response Only</td>
<td>(5)</td>
</tr>
</tbody>
</table>

| Offer               | 0.00063***     |
|                     | (0.00014)      |
|                     | 0.00068***     |
|                     | (0.00015)      |
|                     | 0.00069***     |
|                     | (0.00015)      |
|                     | 0.00439***     |
|                     | (0.00079)      |
|                     | 0.00016        |
|                     | (0.00010)      |
| Constant            | −1.81276***    |
|                     | (0.20463)      |
|                     | −1.63936***    |
|                     | (0.61462)      |
|                     | −1.78558***    |
|                     | (0.65089)      |
|                     | −0.49476       |
|                     | (0.72125)      |
|                     | −0.24126       |
|                     | (0.45773)      |

| Observations        | 1,793          |
| Log Likelihood      | −791           |
| Akaike Inf. Crit.   | 1,588          |
| Bayesian Inf. Crit. | 1,605          |

Notes: *p<0.1; **p<0.05; ***p<0.01. Models (2) to (5) contain controls for income, models (3) to (5) contain the treatment variable.
Chapter 3

Soda Taxes and Fecundability: Evidence from the Philadelphia Sugar-Sweetened Beverage Tax

3.1 Introduction

In 1876, Root Beer became the first mass-produced soda available for public sale in the United States. Then, in 1886, Dr. John Pemberton created Coca-Cola by combining the African kola nut and South American cocaine (The Editors of Encyclopedia Britannica, 2018). While science had not yet discovered the health risks of sugar, the American lifestyle at that time did not leave much room for overconsumption. In the 1950s, however, medical science began to pick up on signs that sugar could lead to coronary heart disease. In response, the Sugar Research Foundation sponsored a series of research papers in the 1960s and 1970s linking dietary fat and cholesterol to heart disease, while downplaying the evidence of the negative health effects of sugar (Kearns et al., 2016). Along with this research came general guidelines from the government which shaped American culture to limit fat but not sugar. It would not be until 1994 that studies linking soda to weight gain would be published. It would then be another decade until soda intake would be linked to Type 2 Diabetes, a fact that may be considered common knowledge to an average American today.
Malik et al. (2006) and Vartanian et al. (2007) provide systematic reviews of decades of literature related to “sugar-sweetened beverage” (SSB) intake, and a general research consensus exists which posits that the prevalence of SSBs in the American diet are clearly contributing to the overall “obesity epidemic” in the United States. The reality of the harms of sugar in sweetened beverages like soda combined with a rising epidemic of obesity in the United States led to a series of taxes on SSBs, colloquially known as “soda taxes”, in the early 21st century. By artificially increasing the price of soda, governments sought to curb its consumption in an effort to combat obesity. While a handful of states impose excise taxes on SSBs and have for decades, only recently have cities begun taxing them. In 2015, Berkeley became the first city in the country to implement an SSB tax, which Falbe et al. (2016) and Silver et al. (2017) conclude successfully reduced consumption of SSBs, but not by a large margin.

With information like this in mind, Mayor Jim Kenney of Philadelphia had an SSB tax approved by city council in June of 2016 with an effective date of January 1st, 2017. The primary goal of the tax, aside from raising revenue, is to reduce consumption of SSBs and make Philadelphians healthier. At 1.5 cents per ounce, a two-liter bottle has an excise tax of one dollar. A quick Google search about the soda tax finds a plethora of articles with titles along the lines of “soda tax causes plunge in sales”, indicating that perhaps part of the mayor’s goal was met. However, some research has concluded that people who could afford to travel further traveled outside the city to buy soda while lower-income Philadelphians took on the burden of the tax due to their relatively inelastic demand for within-city soda (Seiler et al., 2020). Philadelphia and Berkeley are not the only cities to implement such taxes, either. The cities of Albany, Boulder, Oakland, Seattle, San Francisco, and Washington D.C. have also implemented soda taxes in the past decade, with Berkeley having led the way in 2015.

The effects of Philadelphia’s tax have been relatively well-studied over the past few years, and there seems to be a general consensus that soda prices rose for consumers significantly in response to it, suggesting that demand is less elastic than supply (Bleich et al., 2020; Cawley et al., 2019; Roberto et al., 2019; Seiler et al., 2020). Roberto et al. (2019) finds that sales of taxed beverages fell by 1.3 billion ounces in Philadelphia, with a rise of approximately 300 million ounces outside the city, suggesting that some individuals travelled outside the city (or delayed purchases for when they were outside the city) to purchase soda without the tax. The phenomenon of shopping outside
the city is supported by the findings of Seiler et al. (2020), which argues that the net reduction in sales when accounting for surrounding area sales is only 22%, as opposed to an approximate 46% reduction in sales within the city. They state that “there is no significant reduction in calorie intake, [the tax burden] affects low income households more severely, and is limited in its ability to raise revenue”. On the other hand, Zhong et al. (2020) argues that there was “no major overall impact” in consumption of SSBs within the city a year after the tax was placed. One of the reasons why Zhong et al. (2020) seems to contradict some of the other research is that they performed phone interviews asking about consumption as opposed to studying aggregate sales data. One potential strength of phone interviews is that they examine consumption itself as opposed to sales at stores. Cawley et al. (2022) examines an SSB tax on the island of Mauritius and finds that the tax reduced SSB consumption among boys by 11%, although these effects may not be analogous to the United States.

With that in mind, the purpose of this paper is not to study the elasticity of soda demand in Philadelphia, nor will it be to study the incidence of the tax. There is evidence in the medical literature that sugar-sweetened beverage consumption reduces both male and female fertility, thus reducing the probability that a birthing-aged couple would conceive (known as “fecundability”) (Hatch et al., 2012, 2018). Research supporting this connection controls for consumption of fruit juice, diet soda, and other dietary sugar intake, still finding an effect specific to soda.

Since Hatch et al. (2018) controls for other dietary factors and finds that fruit juice and diet soda do not reduce fecundability, there may be something unique about soda consumption –namely the ease with which it allows a person to consume large amounts of high fructose corn syrup– that reduces fecundability in a way that substitution to other “junk” foods may not cause. This is an important research question because it constitutes the first natural experiment to bridge the gap between what is known about soda’s effect on fecundability and the possible unintended consequence of a soda-consumption reducing policy that may have had limited success in other areas of health, such as obesity. A finding that SSB taxes increase birth rates would vindicate the limited research that has been done connecting soda to fecundability (by introducing an exogenous shock to soda consumption) and would offer a new perspective through which policymakers can analyze the effect of an SSB tax. Declining birth rates in the United States are often attributed to higher rates of contraceptive use and a conscious choice by couples to delay pregnancy (Mather
et al., 2021; Chapman, 2022); however, policies that increase fecundability within the population could marginally buffer against the decline in birth rates that the country faces.

It is conceivable that individuals demand a specific amount of calories, sugary beverages, or junk food in a given day. To fill their needs, part of their consumption bundle may include soda, which fills a niche within a subset of “junk foods”. Suppose sweetened beverages are part of a composite good called “junk food” that provides utility to a consumer: a price increase causing fewer purchases of soda may necessitate a substitution to other types of junk food or “naturally sweetened” beverages that are unaffected by the tax. If obesity fails to decline despite the implementation of an SSB tax, it lends credence to the idea that individuals substitute to other junk foods. In fact, Lozano-Rojas and Carlin (2022) finds that the reduction in sugar intake from reduced soda consumption in Philadelphia is partially offset by increased purchases of other sugary snacks.

3.2 Literary Context

3.2.1 SSBs and infertility

Hatch et al. (2018) performs a prospective cohort study examining approximately 4000 women planning pregnancy and approximately 1000 of their male partners. They find that females consuming seven or more sugar-sweetened beverages per week are 19% less likely to conceive; when narrowing down specifically to sugar-sweetened sodas, 2-6 sodas per week has a marginally significant effect of a 13% reduction in fecundability, and seven or more has an effect of 25% reduced fecundability. The effects of diet soda, fruit juice, and sports drinks were found to be insignificant. Results were similar for the male partners who participated in the study and soda intake among couples had an additive rather than an interactive effect on fecundability. Overall, their results were robust to a plethora of counter-considerations and confounders that the authors discuss in detail. Hatch et al. (2018) discusses the potential mechanism as well: sugar intake increases insulin resistance (see Musselman et al. (2011)), which increases oxidative stress (see Park et al. (2009) and Chiu et al. (2014)), which may reduce semen quality in males (see Agarwal et al. (2014)) and ovulatory function in females (see Ruder et al. (2009)). Additionally, American sodas are sweetened with high-fructose corn syrup, which increases insulin resistance more than other sugars (Bocarsly et al., 2010).
3.2.2 The effect of soda taxes on demand and population health

Fletcher et al. (2010b) provides the first empirical analysis of soft drink taxation and its effect on obesity. Using Behavior Risk Factor Surveillance System (BRFSS), the authors study soda taxes at the state level from 1990 to 2006. They find that a one percent increase in state soda taxes reduces body mass index by 0.003 points. According to these results, even a large increase in soda taxes may have a minuscule effect on the BMI of the population. Fletcher (2011) hypothesizes the possibility that soda taxes could involve a substitution effect, thus motivating why the results of Fletcher et al. (2010a) differ from much of the research regarding cigarette taxes. With cigarettes, substitution is limited by nicotine delivery that only a cigarette can provide; in the case of soft drinks, individuals can meet their cravings for sugary beverages by drinking other high-calorie beverages, such as juice (which is very sugary, but avoids the “sugar-sweetened beverage” tax.) Fletcher et al. (2010a) tests this idea directly, and uses NHANES data to find that while soda consumption is reduced by taxes, caloric reductions are completely offset by substitutions to milk and juice.

A large body of research exists examining Philadelphia’s tax. Similar to Fletcher et al. (2010a), Lozano-Rojas and Carlin (2022) examines the substitution to other junk foods (but in Philadelphia specifically), finding that the reduction in sugar intake from SSB taxes is offset 19% by increased consumption of other junk foods. Cawley et al. (2019) found that consumption of the affected beverages fell by about nine ounces per shopping trip in stores within the city compared to stores outside the city. They also found that adult monthly soda consumption frequency fell by about ten units, with a large reduction in sugar consumption for African-American adults. Additionally, they found a reduction in sugar consumption from sweetened beverages among children who had high consumption levels before the tax.

Contrary to Cawley et al. (2019), Seiler et al. (2020) fail to find a statistically significant reduction in calories and sugar post-SSB tax, but still find the expected sign. Calories and sugar both fall by about 15% but with p-values of .07 and .09 respectively. The authors also found that more than half the reduction in sales in the city was offset by cross-shopping to stores located immediately outside the city. They also argue that lower-income individuals who cannot afford to go outside of the city experience a lower elasticity of demand with respect to soda and simply bear the brunt of the tax.
Flynn (2023) examines the effect of SSB taxes on soda consumption and BMI among high schools aged children using the Youth Risk Behavioral Surveillance System and self-reported SSB consumption. The author points out that one limitation of scanner data is that it does not account for who is consuming the good (parents do the shopping) nor whether the good was consumed at all (the good could sit in the pantry unopened); another consideration could be that a student may consume soda at school from a soda fountain without it showing up in store scanner data. The author finds that the tax reduced soda consumption in Philadelphia and BMI in Philadelphia, San Francisco, and Oakland. Lin et al. (2022) also examines high school students, finding an approximate half can per week reduction in soda consumption after the implementation of city-level SSB taxes.

3.3 Economic Theory

3.3.1 Income and Substitution Effect

When the price of a good increases, such as from a tax, the law of demand states that consumers will buy fewer units of the good. Price and quantity demanded are inversely related, thus giving us downward sloping demand curves. However, there are two “forces” acting on consumers in response to a price increase: the income effect and the substitution effect. The substitution effect manifests from the change in relative price between some good and other goods. When the price of a sugar-sweetened beverage (SSB) increases, for example, the relative price of non-SSB goods (in particular, other beverages) becomes smaller in relation to the SSB. Therefore, even if we compensate a consumer with additional income after the price of an SSB rises, we expect some downward pressure on their quantity demanded given that the tradeoff between SSBs and other goods is now steeper (more of other goods need to be given up to purchase an SSB now that the relative price is higher).

The other factor affecting a person’s quantity demanded in response to a price increase is the income effect. Separate from the substitution effect, the income effect is the change in quantity demanded caused not by a change in relative price compared to other goods, but by a change in the consumer’s purchasing power. Imagine a consumer who spends all of her income on SSBs: if the price of SSBs doubles, her real income has essentially been cut in half no differently than if there
were a 100% rate of inflation. At most, she can only purchase half the SSBs she had prior to the price increase. The income effect is, therefore, proportional to how much of a consumer’s income a good takes up. A good that constitutes a larger portion of a person’s income will experience a larger income effect from a price change. For a normal good where reductions in income cause reductions in demand, a larger income effect in response to a price increase causes a larger reduction in quantity demanded of the good (all else equal). However, if a good is inferior, meaning that increases in income lead to reduced demand for the good, then the income effect moves in the opposite direction. Recall that the income effect in response to a price increase is derived from the consumer becoming effectively less wealthy (thus why the larger share of income the good constitutes leading to a more significant “less wealthy” effect); inferior goods are demanded more when a consumer becomes less wealthy. If a good is inferior, a larger income effect would result in a smaller price elasticity of demand since a price increase would cause a positive income effect (cancelling out some of the substitution effect), where upward pressure is placed on quantity demanded due to the consumer effectively becoming poorer.

Some research has established that soft drinks are a normal good by measuring their income elasticity of demand and finding a positive number (see Chacon et al. (2018)), meaning that demand for them increases when incomes rise and falls when incomes fall. However, most research examining elasticity of demand for soft drinks does not examine income elasticity. That being said, research on price elasticity finds larger price elasticity of demand for lower-income individuals. This lends credence to the idea of soft drinks as a normal good since a larger income effect (soft drinks taking up a larger portion of a low-income consumer’s income) results in a higher price elasticity of demand. If price elasticity of demand were smaller for low-income consumers, it would suggest that soft drinks are an inferior good.

From this, it is clear that if SSBs such as soft drinks take up a larger portion of an individual’s income, we would expect that her price elasticity of demand would be higher in response to a price increase. This fundamental theory rings true in the empirical literature, where it is found that young people and low-income individuals have a higher price elasticity of demand in response to soda taxes, reducing their consumption by a larger magnitude than other groups. For that reason, we would expect any side effect of reduced SSB consumption to be concentrated more in the groups with lower income, such as teenagers and African-Americans. This is essentially the finding of
Flynn (2023), which finds that the Philadelphia SSB tax reduced consumption in Philadelphia, especially for non-white females. Hatch et al. (2018) corroborates this as well: in their extensive surveying, they find that women who drink more soda were more likely to be young and less likely to be white.

It should be noted, however, that price elasticity of demand affects the pass-through rate of the tax (how much of the tax is paid by consumers in the form of higher payment at the cash register). There is a general research consensus (Cawley et al. (2019); Silver et al. (2017); Powell (2018); Leider and Powell (2022); Cawley et al. (2020)) that the majority of the tax is bore by the consumers and very little by the sellers, suggesting that supply is substantially more elastic than demand. Suppliers can more easily change their decisions about how much soda to provide in response to price changes than consumers can.

3.3.2 The SSB Tax to Increased Fecundability Mechanism

With literature establishing that heavy SSB consumption reduces fertility and literature leaning toward (but not finding a consensus) that SSB taxes reduce SSB consumption significantly, it is important to discuss to what extent any effect would be found. It is clear from Hatch et al. (2018) that one or more SSBs per day is enough to reduce fertility, while one SSB per week is not. The findings of Lin et al. (2022) (reduction of .48 SSBs per week) and Flynn (2023) (a reduction of 1.3 SSBs per week) are small in comparison to the findings of Hatch et al. (2018). Therefore, SSB taxes may not be large enough to cause a noticeable effect on fecundability.

3.4 Data

Data used includes county-level Natality data from the CDC’s Restricted-Use Vital Statistics Data, which I combined with ACS 1-year census and demographic measures to develop a measure of births per child-bearing age female within a county. Specifically, the number of births recorded is divided by the population of females 15 to 44 to create a “rate” variable of births. Conveniently, the city of Philadelphia is coextensive with Philadelphia county, as are San Francisco and Washington D.C. (with their respective counties). Since I can only observe births and females by county, I exclude soda-taxed cities that are not coextensive with their counties. First, I will compare Philadelphia
to all non-soda taxed counties of 500,000 or more residents, then I will compare Philadelphia, San Francisco, and Washington D.C. to all non-soda taxed counties of 500,000 or more. Additionally, I provide analyses using all women’s pregnancies aged 15-44, teen pregnancies 15-19, and all black women’s pregnancies aged 15-44. This choice is informed by the finding of Flynn (2023) that black youth experienced the largest consumption reduction of SSBs from the tax in Philadelphia and Hatch et al. (2018)’s finding of SSB consumption being higher among non-white youth.

When measuring births among “all women”, I measure births per year per thousand women ages 15-44 in a county. For births among a specific subgroup, I measure births per year per thousand women in that subgroup (e.g. black women aged 15-44 or women and girls aged 15-19) in a county. For example, the county of Anchorage, Alaska had 58,047 women aged 15-44 in 2005 according to the American Community Survey. 3768 of them identified as black, so a measure analyzing births to black women in that county would be proportional to black women (and only measured among women whose CDC record identifies them as black), not all women. Summary statistics for birth rates are displayed in Table 3.1.1

### 3.5 Methodology & Results

I employ a model of the following form:

\[
BirthRate_{it} = \beta(SodaTax_{it}) + \alpha_i + \gamma_t + \mu_{it}
\]  

(3.1)

Here, \(SodaTax_{it}\) is equal to 1 for any city \(i\) that has an SSB tax in place at time \(t\), and equal to 0 for all never taxed cities and cities that are not yet taxed at time \(t\). \(BirthRate_{it}\) is a measure of births per birthing-aged woman at time \(t\) whose residence is in city \(i\). \(\alpha_i\) is a set of city dummy variables (city fixed effects) which account for unchanging characteristics of cities that could affect birth rates. \(\gamma_t\) is a set of year dummy variables (year fixed effects) which account for universal shocks and trends in birth rates across cities.

Bouttell et al. (2018) argues that Synthetic Control is an excellent and underused tool in studying population-level health interventions. When a treatment is limited to one group out of

---

1One county with missing data was removed.
many, a basic differences-in-differences model may be insufficient at showing causal inference regarding an intervention. A newer method that makes improvements on both synthetic control and difference-in-differences, and is applicable in situations where either of the former are, is synthetic difference-in-differences from Arkhangelsky et al. (2021). This method can be understood as a modification of either or a combination of the two. From a difference-in-differences standpoint, synthetic difference-in-differences gives preferential weight to pre-treatment years where eventually treated and control are more similar (as opposed to equal weights in all years); from a synthetic control standpoint, synthetic difference-in-differences minimizes the difference between the pre and post-treatment outcomes for the control group while allowing an intercept adjustment to keep the levels between control and treated separate (as opposed to imposing an identical level). Additionally, inference is performed using regression and standard errors as opposed to the placebo test used by synthetic control.

Since soda taxes were rolled out in different cities at different times, I opt to first analyze Philadelphia on its own as if it were the only treated unit. This allows me to apply the classic difference-in-differences, synthetic control, and synthetic difference-in-differences estimators for one treatment start time and unit. Since the tax went into effect on January 1st of 2017 and it takes a typical birth nine months to occur after conception, I consider the following year to be the treatment start time.\(^2\) For Philadelphia, I examine all births to mothers 15-44, then teen mothers 15-19 and black mothers 15-44. Figures 3.1, 3.2, and 3.3 show synthetic control, difference-in-differences, and synthetic difference-in-differences estimates for all three of these groups. In all cases, the y-axis for these figures is the number of births per thousand women of that subgroup. In other words, births per thousand women ages 15-44, births per thousand teen girls ages 15-19, births per thousand black women ages 15-44. No results are significant at the 5% level, suggesting there is no evidence that the Philadelphia soda tax caused increased fecundability within the population. I repeated this analysis limiting the control counties down to the four used in the “baseline control group” from Flynn (2023): Brooklyn, Queens, the Bronx, and Los Angeles.\(^3\)

---
\(^2\)Note for dissertation committee. Perhaps I should sacrifice my yearly population controls and examine birth rates monthly to see if there is a change from September/October of 2016 to Sept/October of 2017, since that is the earliest we would expect effects to start? (Excluding premature birth, of course.)

\(^3\)Flynn (2023) found control units where a soda tax was publicly debated and other characteristics are similar to Philadelphia. This closes a potential backdoor on the causal pathway from the tax to people’s behavior since public discussion of the SSB tax could affect people’s consumption and it the likelihood of a soda tax coming to pass. It is a potential confounder.
exception, all are coextensive with their counties. Once again the results are insignificant for all three techniques and subgroups, as shown by Figures 3.4, 3.5, and 3.6.

Due to the staggered nature of the soda tax rollout across several treated cities and a lack of parallel trends (making event study implausible), I implement the new Staggered Synthetic Difference-in-Differences estimator from Porreca (2022), which expands the Synthetic Difference-in-Differences estimator from Arkhangelsky et al. (2021) to a staggered rollout scenario. For all mothers, the point estimate is 0.09 with 95% CI [-1.08, 1.26]. For teen mothers, the point estimate is -0.056 with 95% CI [-1.23, 1.12]. For black mothers, the point estimate is -3.97 with 95% CI [-5.15, -2.80]. Once again, I find no evidence that SSB taxes increase fecundability in treated cities, even when allowing all cities to be accounted for with their varied treatment times.

3.6 Discussion

In all cases, I fail to reject the null hypothesis that the SSB taxes have no effect on birth rates. A null result is difficult to glean information from because it leaves three potential avenues of explanation open for debate: SSB taxes do not affect demand enough to have a noticeable effect on fecundability, soda does not effect fecundability in the first place and the medical literature is flawed, or the model is misspecified and I am observing a false negative due to a mistake on my part. The rigorous causal inference of Hatch et al. (2018) leads me to believe that soda does, in fact, have a direct negative causal effect on fecundability. That being said, if two or more sodas per week is enough to reduce fecundability, the mild reductions in soda consumption from the SSB tax are likely not enough to cause a noticeable effect. While couples struggling to conceive should still be advised to reduce SSB intake, it is clear that increased rates of conception are not an unintended consequence of sugar-sweetened beverage taxes.

Note to committee members reading this: I honestly don't know how to explain this number. I can double check and triple check the code to see if there are errors, but I imagine there is something else happening and/or some kind of Type 1 error.
Table 3.1: Summary Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth Rate (all)</td>
<td>1,845</td>
<td>63.735</td>
<td>9.767</td>
<td>0.280</td>
<td>97.321</td>
</tr>
<tr>
<td>Birth Rate (teen mother)</td>
<td>1,830</td>
<td>26.733</td>
<td>15.491</td>
<td>0.061</td>
<td>103.359</td>
</tr>
<tr>
<td>Birth Rate (black mother)</td>
<td>1,845</td>
<td>74.986</td>
<td>21.177</td>
<td>0.299</td>
<td>311.734</td>
</tr>
</tbody>
</table>

Figure 3.1: All mothers (15-44), Philadelphia as treated
Figure 3.2: Teen mothers (15-19), Philadelphia as treated

Figure 3.3: Black mothers (15-44), Philadelphia as treated
Figure 3.4: All mothers (15-44) (Control group limited to Flynn (2023) baseline.), Philadelphia as treated.
Figure 3.5: Teen mothers (15-19) (Control group limited to Flynn (2023) baseline.), Philadelphia as treated
Figure 3.6: Black mothers (15-44) (Control group limited to Flynn (2023) baseline.), Philadelphia as treated
References


Sage, A., Langen, M., and Van de Minne, A. (2021). Where is the opportunity in opportunity zones?


A  Additional Explanations, Tables and Figures for Chapter 1

A.1  Expanded Explanation of ACS 5-year Demographic Variables

For the race variables, ‘B02001_002’ and ‘B02001_003’ estimate the number of White and Black individuals, respectively. ‘B01001_001’ (Pop) is an estimate of the total tract population.

For the Hispanic variable, I used ‘B03002_012’. This variable is the estimate of the number of individuals who are of Hispanic or Latino origin. This is not a “racial” variable; individuals in the “White” or “Black” groups may also be Hispanic.

For the Income variable, I used ‘B19301_001’. This variable is the estimate of the tract-level per capita income earned in the last 12 months in the given year’s inflation-adjusted dollars. In other words, this is real per capita income.

For the Male15to21 variable, I used ‘B01001_006’ (15 to 17yrs), ‘B01001_007’ (18 and 19yrs), ‘B01001_008’ (20yrs), and ‘B01001_009’ (21yrs). Each of these variables estimates the number of males of a certain age within the tract. Summing these together by tract-year provides a measure of Males 15 to 21.

For the Single Mother variable, I used ‘B09005_005’. This variable used to be the estimate of children under 18 living in a home with a female householder who has no husband present. Recently, the language was changed from husband to spouse/partner. Regardless, this is the variable for determining the number of children living in homes with a single mother. For the “rate”, I divide this by ‘B09005_001’, which is the number of children under 18 in the tract.

For the educational attainment variables, I used ‘B06009_002’ through ‘B06009_06’. This variable is the estimate for the number of individuals 25 and older categorized by their highest level of educational attainment. No high school diploma or GED, diploma or GED, some college, bachelor’s degree, and graduate or professional degree. The “rate” variable for these is divided by ‘B06009_001’, which is the population 25 and older; this is not to be confused with ’B01001_001’, which is the total population.

For individuals who have not worked in the last 12 months, I used ‘B23022_025’ and ‘B23022_049’, which are males and females, respectively, within the “universe” of 16 to 64 year olds. For the “rate” variable, I divided by ‘B23022_001’, which is the number of individuals 16 to 64.
A.2 Regarding Causal Inference and Standard Errors

One of the problems with standard error clustering is that there is no consensus over either the proper degrees of freedom to use, or the type of heteroskedasticity-robust standard errors a project such as this should be employing. This is even demonstrated by the idiosyncracies among default degrees of freedom and heteroskedasticity calculations done in R and Stata (this project uses R). It is reasonable to expect that regressors and errors would be correlated within tracts. This is why Austin and Small (2014) and Abadie and Spiess (2022) both argue that, when building a model on propensity score matched data, cluster-robust standard errors at the pair membership level should be employed. Since, after matching, each tract is very similar to its partner in the subclass, we would expect the errors to be correlated within each pair, as opposed to more specifically in each tract. Ho et al. (2007) demonstrate that matching can be used to process data, after which you can use standard regression techniques you “would have used anyway”. That being said, errors are ultimately likely correlated across tract borders. On one hand, it could be argued that since tracts themselves are not particularly chosen for treatment, and it is in fact a square area of Philadelphia that received this designation based on factors inherent to the area as a whole, the standard errors provided by clustering at the tract (or pair) level may be too large. Clustering alternatively on the binary variable of “Zone” would cause a regression with many thousands of observations to have 1 degree of freedom\(^5\), making it nearly impossible to discern significance on the treatment effect.

There is a general consensus that clustering at the tract level may give overly conservative estimates of the standard errors if errors are not correlated within tracts. Abadie et al. (2017) demonstrate that if treatment is not cluster-specific, then clustered standard errors may be too conservative, making rejection of the null unattainable even if there is a “real” treatment effect. After including tract fixed effects and tract-level demographic controls, it may not be the case that the errors need a tract-level correction.

Another problem with clustering is that it can be unreliable in instances where there are fewer than fifty clusters in the dataset. In this case, there are only eight clusters. According to Cameron and Miller (2015), when there are more regressors than clusters – these models have more than eight demographic controls and hundreds more when FEs are considered– the cluster-robust variance

\(^5\)Generally speaking, while non-clustered SEs use n-k degrees of freedom to determine p-value, clustered SEs with G clusters result in G-1 (or G-k) degrees of freedom, resulting in larger p-values. Cameron and Miller (2015)
matrix becomes rank-deficient. Despite this, individual analysis of regression coefficients can still be performed. Cameron and Miller (2015) claims that having few clusters does not necessarily make regression coefficients imprecise if there are many observations per cluster, but the standard errors are likely to become biased downward. For the most robust standard errors when there are fewer than ten treated clusters, they recommend following the Webb six-point method of wild cluster bootstrapped standard errors from Webb (2014). Even in the full sample regression with 375 clusters, having only 8 clusters being treated also falls under the “too few clusters” problem as explained by Cameron and Miller (2015). Therefore wild cluster bootstrap standard errors are utilized in those models as well wherever significance is found using non bootstrapped errors. The p-values are shown beneath the coefficients of interest in square brackets, when applicable. From Cameron and Miller (2015), I also choose to two-way cluster my standard errors, at both the tract and year level, since error correlation not only is likely to occur within tracts, but within specific years in Philadelphia as well.

In any case, I tested with standard heteroskedasticity robust and cluster-robust standard errors on all models (the latter of which are displayed in the tables by default), and included Webb Six-Point Wild Cluster Bootstrapped standard errors on any fully specified model showing significance on the treatment variable. In all cases, the hetero-robust errors were the least conservative, often giving p-values substantially smaller than the cluster-robust and bootstrapped errors (typically one magnitude of confidence smaller, e.g. 0.05 to 0.01). On the other side were the bootstrapped errors, which were extremely conservative and generally increased p-values by one magnitude (e.g. 0.05 to 0.1).

A.3 The Endogeneity Problem

One problem is that the choice of this particular area is endogenous to the problems the area is having, such as its violent crime. An area that “selects into” the policy treatments of the Promise Zone may be systematically different from other area of Philadelphia (Kuehn, 2014). Neumark and Young (2019), in a re-analysis of a previous enterprise zone study, suggest identifying control areas that are similar to the zones but where the policies did not apply. Similar to O’Keefe (2004) and Elvery (2009), I use propensity score matching to find tracts in Philadelphia with similar levels of violent crime and other covariates that may be linked to the decision to establish the Promise Zone.
However, those studies only use pre-treatment variables to perform their matching. As Neumark et al. (2014) points out, some studies perform their propensity score matching on variables both before and after the treatment was established; this is allowable only if the treatment has no potential causal link to these variables in the post-period (Greifer, 2022).

I used propensity score matching\textsuperscript{6} to determine the average treatment effect on treated by attempting to account for selection into the Promise Zone by city planners and government officials. To do this, I concatenated the data to a cross-section of pre-2014 tract-level averages. I used nearest neighbor matching without replacement to predict the “Zone” variable using all of the controls and violent crime as predictor variables. Unfortunately, this yielded a poor balance, where several of the control variables and pre-2014 violent crime fell outside the threshold of 0.1 standardized mean difference. I then attempted to find a better balance by removing some of the covariates. Promise Zone designations primarily target areas with high concentrations of people of color, low levels of educational attainment, low levels of income, high levels of joblessness, and high levels of crime. Therefore, I attempted matching on the variables Black, Hispanic, Less than high school, income, and No work 12 months. Unfortunately, this yielded a poor balance as well. While Promise Zones often target communities of color, (the Los Angeles Promise Zone is primarily Hispanic, for example), the Philadelphia Promise Zone is primarily African-American, and not Hispanic. Upon removing Hispanic from my matching algorithm, I am able to achieve balance as displayed by Figure A.1. The outcome of this procedure is a set of paired tracts: eight treated and eight control, each assigned to a “subclass”, which is the suggested level of clustering for subsequent regression models. I then employ the same difference-in-differences with two-way fixed effects models as before, but with only the eight treated tracts and their partners. For one model, I only use the controls that were matched on; for the other, I use the full set of controls. Generally speaking, the model using only the controls that were matched on is the one that should be most reliable. Table A.1 displays the results of this model, and suggests that, compared to tracts that have similar qualifications to tracts that ultimately receive the Promise Zone designation, violent crime fell substantially within the actual Zone.

\textsuperscript{6}See this vignette from the MatchIt package for more.
Table A.1: Difference-in-Differences for Violent Crime

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Violent Crimes per 1k Residents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Post x Zone</td>
<td>$-15.059^*$</td>
</tr>
<tr>
<td></td>
<td>(6.927)</td>
</tr>
<tr>
<td></td>
<td>[0.011]^{**}</td>
</tr>
<tr>
<td>TWFE?</td>
<td>X</td>
</tr>
<tr>
<td>Controls?</td>
<td>All</td>
</tr>
<tr>
<td>Observations</td>
<td>160</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.865</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.825</td>
</tr>
<tr>
<td>Residual Std. Error</td>
<td>7.677</td>
</tr>
</tbody>
</table>

*Note:* $^* p < 0.1; ^{**} p < 0.05; ^{***} p < 0.01$

Std. Errors Clustered at subclass level.
Wild cluster bootstrapped p-val in square brackets.
A.4 Blocks on or around the border compared to blocks within

I choose not to perform a block-level analysis for much of this paper since the treatment is not at the block-level and, in choosing to analyze blocks, I lose the precision that the ACS 5-year tract population and demographic controls provide. The primary weakness of block-level analysis is that I can only measure crime incidents without controlling for changes in population. That being said, block-level analyses allow me to examine the border of the Promise Zone and compare the area immediately at the edge to the area within the Zone. Additionally, I can include crime observations back to 2006 and aggregate at a shorter time period than year if I am not relying on ACS data; this results in 56 quarters of block-level crime incidents. I compare blocks completely contained within 1500 meters of the border of the Promise Zone to blocks contained fully within the Promise Zone which have no borders lying outside the Zone. The exact choice of border down to the block level\(^7\) may be arbitrary, as the zone, to my knowledge, does not facilitate any kind of treatment cutoff explicitly at the border, but instead facilitates grant money to organizations based within the Zone. For that reason, I allow partially contained blocks to be considered untreated.

I employ both a zero-inflated\(^8\) and a non zero-inflated generalized linear mixed model using the package \textit{glmmTMB} in R. Block and date (quarter) are included as “crossed” random effects. For the model without zero-inflation, the treatment is associated with fewer violent crime incidents. For the zero-inflated model, treatment is significantly associated with observing zero violent crimes in a block-quarter despite the zone being less likely to observe zero violent crimes in general. This is in line with the MVCDC claim that they observed a street corner where crime fell to zero. Block-quarters with 0 incidents become more likely with the establishment of the Promise Zone.

\(^7\)The general choice of boundaries is not arbitrary in itself, but whether a specific street block at the edge is included or not is likely arbitrary.

\(^8\)Approximately 2/3rds of the block-quarter violent crime observations are 0s.
<table>
<thead>
<tr>
<th></th>
<th>GLMM</th>
<th>GLMM w/ ZIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.54***</td>
<td>1.26***</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>post</td>
<td>-0.05**</td>
<td>-0.13**</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>zone</td>
<td>0.23***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>post:zone</td>
<td>-0.06**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
</tr>
</tbody>
</table>

Count model: (Intercept) 1.26***
Count model: post -0.13**
Count model: zone 0.07
Count model: post:zone 0.03
Zero model: (Intercept) 0.85***
Zero model: post 0.04
Zero model: zone -0.99***
Zero model: post:zone 0.19**

<table>
<thead>
<tr>
<th></th>
<th>GLMM</th>
<th>GLMM w/ ZIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>97313.21</td>
<td>74045.73</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-48649.61</td>
<td>-37009.87</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>34608</td>
<td>34608</td>
</tr>
</tbody>
</table>

***p < 0.001; **p < 0.01; *p < 0.05

Table A.2: GLMM Models
Figure A.2: Blocks contained within 1500m of the border
A.5 Robbins and Davenport Block-Level “MicroSynth” Methodology for Synthetic Control

As an alternative method of disaggregating the potential crime-reducing effect of the Zone, I employed the Synthetic Control Method for Microdata to offer an alternative counter-factual to the Zone had the treatment not occurred. It is important to note that this methodology operates under a separate set of assumptions from the Synthetic Difference-in-Differences applied elsewhere in this research. It also operates at a finer level of aggregation and studies crime incident frequency as opposed to population-adjusted rates.

I apply the exact methodology of Robbins and Davenport (2021) in their showcase of the package Microsynth to apply a block-level analysis of crime incidents as opposed to crime incidence. One benefit of this package and its methodology is that it can result in a perfect pre-treatment match between treated and synthetic, which is unlikely in most other synthetic control applications. Block-level micro-data is only available in the decennial census, which I acquired using the Tidycensus package in R. The Robbins and Davenport (2021) methodology requires covariates to be time-invariant, and the authors use cross-sectional decennial census data available in the “SeattleDMI” dataset. There are 18,872 census blocks in Philadelphia, 1898 of which had no crime occurrences (of any type) during the period studied. These blocks were removed leaving 16974, 379 of which are in the Promise Zone. Applying my code in the exact manner that the authors demonstrate, I use the same covariates, which are block-level total population, Black population, Hispanic population, number of households, number of owner-occupied family households, number of female-headed households with no husband present, number of renter-occupied households, number of vacant houses, and number of males 15 to 21. I then merged this data with all crime occurrences from 2010 to 2019, which I aggregated at the quarterly level, once again following the methodology of Robbins and Davenport (2021). Similar to the authors, I use every available outcome variable for

---

9 Note that the dataset I build is essentially the same as the SeattleDMI dataset, except for Philadelphia. It is the combination of crime panel data and a cross-section of block demographic data.

10 Even though I am examining violent crime, there are blocks where no offenses, including minor non-violent ones going back to 2006, never occurred a single time. These blocks are likely devoid of activity and their removal should not bias my synthetic control. On the other hand, I choose not to exclude blocks with zero population, as these blocks could still have corner stores, gas stations, and other points of interest that are subjected to crime.

11 The authors suggest that if the treatment is expected to not have immediate effects, the final “pre-treatment” period should be the period the treatment begins. Therefore, the 17th quarter (Q1 2014) is the final pre-treatment period in my code.
my analysis. This includes a measure of all violent crimes, all non-violent crimes, property crimes, homicide, (other) assaults, aggravated assault, aggravated assault with a firearm, rape, robbery, robbery with a firearm, and arson.
Agg Assault w/ Firearm

Quarterly Block-Level Crime Incidents

Treatment

Synthetic Control

All cases (scaled)

Robbery

Quarterly Block-Level Crime Incidents

Treatment

Synthetic Control

All cases (scaled)

Robbery w/ Firearm

Quarterly Block-Level Crime Incidents

Treatment

Synthetic Control

All cases (scaled)
These results, which are tabulated in Table A.3 suggest a statistically significant reduction in violent crime occurrences (6.6%), mostly attributable to an 11.4% reduction in occurrences of non-aggravated “other” and simple assaults. In the OpenDataPhilly API, “other assaults” is any assault that is not aggravated. The simple/other assaults variable is significant at the 1% level for all three methods of causal inference provided by Microsynth. Note that, graphically, the largest reductions in violent crime and assault occur around 2018, which is where the result was most significant in the event study analysis. There is limited evidence that there may also be a reduction in aggravated assault with a firearm, which is significant at the 10% level for all three methods. It should be noted that this methodology does not account for changes in population, and takes the demographic covariates as time-invariant (which is required for the package according to the authors). This may result in an understatement of effect if population is increasing, and an overstatement of the effect if population is decreasing. On the other hand, while yearly population estimates are unavailable for blocks, the tract-level ACS 5-year estimates show an increase in Zone population throughout the period studied. This is especially important when noting the significant increase in non-violent crime occurrences attributable to the zone as demonstrated by the synthetic control. More people and more economic activity in an area leads to more incidents of non-violent crime, which includes not only property crime but “all other offenses”, the most common crime type reported in the database. “All other offenses” includes minor crimes that are neither Part 1 nor Part 2, but excludes traffic offenses. Non-violent offenses that are not property-related are minor and not particularly costly to society, such as loitering.
Table A.3: Microsynth SCM Output Table

<table>
<thead>
<tr>
<th></th>
<th>Treat</th>
<th>Ctrl</th>
<th>% Chng.</th>
<th>linear.pVal</th>
<th>Jack.pVal</th>
<th>Perm.pVal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violent</td>
<td>6329</td>
<td>6776.22</td>
<td>-6.6%</td>
<td>0.0003</td>
<td>0.0141</td>
<td>0.0080</td>
</tr>
<tr>
<td>Simple/Other Assault</td>
<td>3445</td>
<td>3888.71</td>
<td>-11.4%</td>
<td>0.0000</td>
<td>0.0006</td>
<td>0.0000</td>
</tr>
<tr>
<td>Agg. Assault</td>
<td>970</td>
<td>936.21</td>
<td>3.6%</td>
<td>0.8056</td>
<td>0.7556</td>
<td>0.7280</td>
</tr>
<tr>
<td>Agg. Assault w/ Firearm</td>
<td>358</td>
<td>400.13</td>
<td>-10.5%</td>
<td>0.0625</td>
<td>0.0851</td>
<td>0.0880</td>
</tr>
<tr>
<td>Robbery</td>
<td>605</td>
<td>630.46</td>
<td>-4.0%</td>
<td>0.1916</td>
<td>0.2551</td>
<td>0.2480</td>
</tr>
<tr>
<td>Robbery w/ Firearm</td>
<td>448</td>
<td>431.71</td>
<td>3.8%</td>
<td>0.7528</td>
<td>0.7015</td>
<td>0.7240</td>
</tr>
<tr>
<td>Homicide</td>
<td>48</td>
<td>53.92</td>
<td>-11.0%</td>
<td>0.2208</td>
<td>0.2344</td>
<td>0.2120</td>
</tr>
<tr>
<td>Rape</td>
<td>211</td>
<td>210.09</td>
<td>0.4%</td>
<td>0.5214</td>
<td>0.5189</td>
<td>0.4680</td>
</tr>
<tr>
<td>Other Sex Offenses</td>
<td>169</td>
<td>156.62</td>
<td>7.9%</td>
<td>0.7776</td>
<td>0.7472</td>
<td>0.7040</td>
</tr>
<tr>
<td>Arson</td>
<td>75</td>
<td>68.39</td>
<td>9.7%</td>
<td>0.7590</td>
<td>0.7303</td>
<td>0.7720</td>
</tr>
<tr>
<td>Non-Violent</td>
<td>23018</td>
<td>20962</td>
<td>9.8%</td>
<td>0.9992</td>
<td>0.9749</td>
<td>0.9640</td>
</tr>
<tr>
<td>Property</td>
<td>8068</td>
<td>8010.85</td>
<td>0.7%</td>
<td>0.6261</td>
<td>0.5741</td>
<td>0.5800</td>
</tr>
</tbody>
</table>

To understand the Average Treatment Effect on Treated (ATT), Microsynth calculates the cumulative number of cases between the treated and synthetic control groups as shown in Table A.3. Since the match is perfect in the pre-treatment period, the difference between the Trt and Con columns indicates the number of incidents of crime potentially prevented by the zone. From 2014 to the end of 2019, 447 violent crimes are estimated to have been prevented. Additionally, 443 simple/other assaults are potentially prevented, making up most of the violent crime prevention estimated by the model. If we accept that aggravated assaults with firearms are reduced, then the Zone’s programs may have prevented up to 42 of them from occurring during the period studied. While these are just estimates, they can help inform a cost-benefit analysis of crime-prevention in the zone.

A.6 Cost-Benefit Analysis of Crime Reducing Policies

An Executive Summary from Rebecca Rhynhart, City Controller of Philadelphia, states that a single year of a 10 percent homicide reduction in the city would net the city approximately $13 million per year. Of course, the focus of this study (and the $3.38 million spent) is on this much smaller area of 30,000 or so residents. I do not find any evidence that the Zone reduced homicides. I do find evidence of a significant reduction in simple assaults and very limited evidence of a modest reduction in aggravated assaults with firearms (which include non-fatal shootings, something of great interest to policymakers).
Simple assault is defined by the FBI as an assault “where no weapon was used or no serious or aggravated injury resulted to the victim. Stalking, intimidation, coercion, and hazing are included.” Aggravated assault is “an unlawful attack by one person upon another for the purpose of inflicting severe or aggravated bodily injury. This type of assault is usually accompanied by the use of a weapon or by means likely to produce death or great bodily harm.” McCollister et al. (2010) provides an extensive analysis of both tangible and intangible costs of violent crime victimization. They estimate the tangible cost of an aggravated assault, which includes direct costs to the victim and the legal system, at around $20,000 per aggravated assault in 2008 dollars. They further estimate the “pain and suffering” intangible cost at an additional $13,000. They also include a “risk of homicide” cost relative to the risk that an aggravated assault escalates into a homicide; that being said, I ignore this amount (often much larger than the other costs mentioned) since I know that the aggravated assaults in my data did not result in a homicide. They estimate $20,000 in tangible costs and $13,000 in intangible costs for an aggravated assault. If 42 fewer aggravated assaults with firearms occur because of the Promise Zone, then $840,000 in tangible costs are prevented, along with $546,000 in intangible costs. In 2022 dollars, this is approximately $1.9 million. McCollister et al. (2010) does not cover simple assaults since, according to the authors, they are too common and often go unrecorded. However, the Office of Justice Programs estimates that an “other assault” with no injury as having a total cost of $2000. Since some minor injury and/or domestic abuse instances may be counted in the “other assaults” variable provided by the opendataphilly API, this $2000 estimate is a lower bound. Adjusted for inflation, given that the original estimate was in 1993 dollars, results in a cost of $4100. 443 fewer simple assaults translates to $1,816,300 in costs prevented. In other words, 5 years of the Promise Zone may have prevented approximately $3.7 million in social costs related to violent crime victimization. When adjusted for inflation, this is remarkably close to the amount of money spent on public safety grants in the Zone, but pales in comparison to the tens of millions spent on education in the area. That being said, if the effect on aggravated assaults is assumed away, then the city only recoups about half the value of the investment in public safety from the grants.

On the other hand, if we trust the synthetic difference-in-differences point estimate of -4, and

---

12 It should be noted that these estimates are conservative compared to the Rand Corporation’s estimate of $87,000 per aggravated assault as noted here.
assume the entire effect is from the least costly violent crime of simple assault, we arrive at a slightly different number. Considering a reduction of four violent crimes per thousand residents per year, and assuming an average Promise Zone population of 30,000 residents, this is 720 fewer violent crimes over a six year span. Given the $4100 potential cost of a simple assault, this works out to $2,952,000. Even given this more conservative estimator, it is still reasonable to believe that the Zone recouped millions of dollars in reduced crime victimization due to the investment in the Zone.

A.7 What about partially treated tracts?

In all of the analyses done throughout this paper, the only tracts considered “treated” are ones that lie fully within the Zone. However, there are several tracts that lie partially within the Zone: 86.02, 87.01, 87.02, 88.01, 88.02, and 369. I apply the Synthetic DiD estimator, which provides an estimate of 2.09 with 95% CI [0.23, 4.09]. Figure A.3 helps demonstrate that if the partially treated tracts experienced reductions in violent crime in the post-period similar to the weighted control group, the average level of violent crime would be two crimes per 1k residents lower per year. This positive ATT for the partially treated units seems largely driven by the spike in 2016 in these six tracts. Considering that the event study applied in Section 1.6.2 demonstrated the strongest reductions in crime attributable to the Promise Zone occurring in 2018 and 2019, it is not clear that the Promise Zone is pushing crime “around the corner”.\textsuperscript{13} It is also important to note that if the Promise Zone is somehow increasing violent crime in these six tracts, the effect size is half that of the reduction that the eight fully treated tracts experience.

\textsuperscript{13}Weisburd et al. (2006) examined whether the practice of hotspot policing simply pushes crime “around the corner”. The Philadelphia Promise Zone helps facilitate hotspot policing strategies, and it is plausible that criminals would simply go elsewhere to commit crimes if there is a heightened police presence. However, the author uses a spatial econometric analysis to argue that crime not only does not move around the corner, but there are positive spillover effects in crime deterrence to areas outside the hotspot.
Figure A.3: Synthetic DiD for Partially Treated Tracts
A.8 What about Drexel University in Tract 90?

One possible source of bias in my point estimates is that Census Tract 90 contains Drexel University and its many on-campus resident students. While the other tracts which lie fully within the Promise Zone have relatively high rates of violent crime, low educational attainment, and large African-American communities, Tract 90 has a very low crime rate, a median age of 21 (expected of a college campus), high rates of high-school completion (once again, expected of a college campus) and a majority white community, with Asian-Americans being the largest minority group. While the level of measured poverty is similar to the rest of the zone, this is likely due to the low incomes and ages of full-time college students. While the Promise Zone happens to cover Tract 90, it is not likely that the BCJI or the Promise Zone were placed there for the benefit of the college students of Drexel University, but the residents of the other seven tracts.

Additionally, one could argue that the Drexel police’s patrol zone (which covers most of Tract 90) could influence the results. However, when I apply the Synthetic DiD estimator to the data with Tract 90 removed (and therefore only 7 treated units), I obtain an estimate of -4.40 with 95% CI [-8.95, -0.33]. This estimate is larger in absolute value than the estimate that included Tract 90.
B Additional Explanations, Tables and Figures for Chapter 2

B.1 What if respondent WTA changes once they are offered money.

In the WTA analyses from the body of the paper, I assume that individuals who answered in support of a facility on their block before any money was offered have an implicit WTA of 0 and their responses to follow-up WTA questions were ignored. If I relax that assumption and allow them to be inconsistent, developing a WTA above zero once money starts being offered, I get a median WTA estimate of $2430.
Table B.1: Allowing WTA to change once respondents are offered money

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Answer = “Yes”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offer</td>
<td>0.0009141***</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
</tr>
<tr>
<td>&lt;10,000</td>
<td>−0.946</td>
</tr>
<tr>
<td></td>
<td>(1.213)</td>
</tr>
<tr>
<td>20k-34,999</td>
<td>−0.585</td>
</tr>
<tr>
<td></td>
<td>(0.955)</td>
</tr>
<tr>
<td>35k-49,999</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>(0.934)</td>
</tr>
<tr>
<td>50k-74,999</td>
<td>−0.924</td>
</tr>
<tr>
<td></td>
<td>(0.891)</td>
</tr>
<tr>
<td>75k-99,999</td>
<td>0.772</td>
</tr>
<tr>
<td></td>
<td>(0.935)</td>
</tr>
<tr>
<td>100k-149,999</td>
<td>1.505</td>
</tr>
<tr>
<td></td>
<td>(0.937)</td>
</tr>
<tr>
<td>150k-199,999</td>
<td>0.551</td>
</tr>
<tr>
<td></td>
<td>(1.183)</td>
</tr>
<tr>
<td>200,000 or more</td>
<td>−0.527</td>
</tr>
<tr>
<td></td>
<td>(1.182)</td>
</tr>
<tr>
<td>Prefer not to answer.</td>
<td>−2.257*</td>
</tr>
<tr>
<td></td>
<td>(1.257)</td>
</tr>
<tr>
<td>Treated</td>
<td>0.559</td>
</tr>
<tr>
<td></td>
<td>(0.419)</td>
</tr>
<tr>
<td>Constant</td>
<td>−2.221***</td>
</tr>
<tr>
<td></td>
<td>(0.808)</td>
</tr>
</tbody>
</table>

| Observations        | 2,196          |
| Log Likelihood      | −903.357       |
| Akaike Inf. Crit.   | 1,832.714      |
| Bayesian Inf. Crit. | 1,906.741      |

Note: *p<0.1; **p<0.05; ***p<0.01
B.2 Including starting point fixed effects.

I also include a model where I add dummy variables for the starting point a respondent was shown, $200, $500, or $3000 ($500 is excluded as the reference point). The WTA estimate is still around $2800 in this case. While the coefficients on the extra fixed effects are not significant, they follow the expected sign in relation to $500: $200
Table B.2: Including Starting Point Fixed Effects

<table>
<thead>
<tr>
<th>Offer</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0005733***</td>
<td>(0.0002)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>10,000</td>
<td>-0.986</td>
<td>(0.908)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>34,999</td>
<td>-0.408</td>
<td>(0.719)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>49,999</td>
<td>-0.047</td>
<td>(0.708)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>74,999</td>
<td>-0.846</td>
<td>(0.672)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>99,999</td>
<td>0.382</td>
<td>(0.708)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>149,999</td>
<td>1.209*</td>
<td>(0.734)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>199,999</td>
<td>0.269</td>
<td>(0.906)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>200,000 or more</td>
<td>-0.617</td>
<td>(0.910)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>Prefer not to answer</td>
<td>-1.697*</td>
<td>(0.939)</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Treated</td>
<td>0.342</td>
<td>(0.322)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>$X=200</td>
<td>-0.558</td>
<td>(0.387)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>$X=3000</td>
<td>0.136</td>
<td>(0.431)</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.587**</td>
<td>(0.642)</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Observations: 1,793
Log Likelihood: -778.172
Akaike Inf. Crit.: 1,586.344
Bayesian Inf. Crit.: 1,668.719

Note: *p<0.1; **p<0.05; ***p<0.01
B.3 Linear Probability Models

As an alternative way of conceptualizing the model, where each question is a separate dependent variable, a linear probability model to measure the effect of the information treatment is specified as follows:

\[
(Support)_i = \beta_0 + \beta_1(Treated)_i + \beta_2(Gender) + \beta_3(Race) + \beta_4(HouseholdIncome) + \beta_5(AgeBracket) \\
+ \beta_6(EducationLevel) + \beta_7(Housing) + \beta_8(Kens) + \beta_9(AddictedFamily) + \beta_{10}(SeenTrash) \\
+ \beta_{11}(FamiliarNEP) + \beta_{12}(FamiliarSIF) + \beta_{13}(No.Children) + \beta_{14}(HouseholdSize) 
\]  

(2)

where \(Treated\) indicates whether the respondent received the information treatment. All controls are in the form of factor variables, where each control is a set of binary regressors, one of which equals 1 while the rest equal 0. In this case, \(Support\) is specific to the question being asked, with \(n = 382\) and a separate model being specified for each attribute/level combination.

The difference between Far/Frequent and Near/Never were difficult to hypothesize about, since it is not clear whether respondents would be more averse to a site near them that they never interact with (something that may be difficult to imagine) or a site far away from their living space in a place they often go. Interestingly enough, the untreated group seems to prefer Near/Never, whereas the treated group prefers Far/Frequent. That being said, the two are not significantly different when comparing within treatment status. In fact, support does not decline in a significant way for the untreated until Far/Frequent is introduced after Far/ Never, and again when ”On Block” is introduced after Near/Frequent. The latter is where the largest drop occurs, suggesting that respondents overall are much more averse to a site right next to them than a site nearby that they have to interact with often. The decline from Near/Frequent to “On Block” is so significant that even the confidence intervals do not overlap.

Analysis of the bar chart aside, Table B.3 shows the main results, demonstrating effects significant at the 5% level for Far/Frequent and “On Block”. These results suggest that the treatment increases likelihood of support by around 10% for these situations, and possibly similarly for “Far/Never”, “Near/Never”, and “Near/Frequent”. Note that while many of the demographic variables are significant and have the expected signs, these characteristics are endogenous to selection.
into Dynata surveys. Since treatment is randomly assigned and not endogenous, a causal claim can be made that this information treatment does significantly increase people’s openness to the facility. By providing a respondent with generalized information about the relative harmlessness of SIFs, we can reduce opposition to the site.

When the analysis is repeated with Zipcode Fixed Effects, the treatment effect is weaker. In this case, every level from Far/Frequent to “On Block” is significant at the 10% level. These results are shown in Table B.4.

Table B.3: Baseline Linear Probability Model

<table>
<thead>
<tr>
<th></th>
<th>Legal</th>
<th>Far/Never</th>
<th>Far/Freq</th>
<th>Near/Never</th>
<th>Near/Freq</th>
<th>Block/Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated</td>
<td>0.086*</td>
<td>0.096*</td>
<td>0.124**</td>
<td>0.079</td>
<td>0.085*</td>
<td>0.088**</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.052)</td>
<td>(0.048)</td>
<td>(0.049)</td>
<td>(0.048)</td>
<td>(0.042)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>All Controls?</th>
<th>Observations</th>
<th>R²</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>382</td>
<td>0.257</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>382</td>
<td>0.311</td>
<td>0.241</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>382</td>
<td>0.309</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>382</td>
<td>0.277</td>
<td>0.185</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01
HC1 Robust Standard Errors

Table B.4: Fixed Effects Linear Probability Model

<table>
<thead>
<tr>
<th></th>
<th>Legal</th>
<th>Far/Never</th>
<th>Far/Freq</th>
<th>Near/Never</th>
<th>Near/Freq</th>
<th>Block/Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treated</td>
<td>0.084</td>
<td>0.111**</td>
<td>0.116**</td>
<td>0.131***</td>
<td>0.134***</td>
<td>0.105**</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.052)</td>
<td>(0.048)</td>
<td>(0.049)</td>
<td>(0.048)</td>
<td>(0.042)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>All Controls?</th>
<th>Observations</th>
<th>R²</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>372</td>
<td>0.385</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>372</td>
<td>0.406</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>372</td>
<td>0.402</td>
<td>0.184</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>372</td>
<td>0.374</td>
<td>0.178</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01
Zipcode FE. HC1 Robust Standard Errors
B.4 Effect of Information

As demonstrated by Figure B.1, there might be a statistically significant difference in the proportion of respondents supporting the site at any Attribute/Level between the treated and untreated group when a raw comparison is made.\textsuperscript{14} Since we have already established random treatment assignment, this graph conveys reliable information. As expected, support declines substantially in both groups as the distance decreases and the interaction rate increases.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure_b_1.png}
\caption{Support Grouped by Treatment Status}
\end{figure}

\textsuperscript{14}Overlapping confidence intervals do not necessarily rule out significant differences: https://cscu.cornell.edu/wp-content/uploads/73_ci.pdf.
B.5 But what about NIMBYism?

This fact alone is interesting and useful, but still does not address whether the information treatment affects NIMBYism specifically. After all, NIMBYism is conditional on someone supporting the site far away but opposing it on their block. For that purpose, I perform a Welch’s two-sided t-test, comparing the mean difference between “Far/Never” and “On Block” between the treated and untreated groups. This test demonstrates that there is no significant difference in the difference between the treatments effect on bridging the gap between “Not in my backyard” support and “Yes, in my backyard” support.

<table>
<thead>
<tr>
<th>Treated</th>
<th>Untreated</th>
<th>t-Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.28</td>
<td>0.27</td>
<td>0.37</td>
<td>0.71</td>
</tr>
</tbody>
</table>

The implication of this result for the main results is that there is no heterogeneity in how treatment affects respondents at various levels of interaction/distance. The treatment provides a fixed “lump sum” boost in support that does not seem to vary across the spectrum from “Far/Never” to “On Block”.

As for whether NIMBYism exists in the first place, we can examine Figure 4 or simply perform another t-test of means. It is very clear that a substantial number of respondents who believe SIFs should be legal would not like one on their block. Unlike Boyle et al. (2019), which finds that individuals who view wind power positively are consistent in supporting the establishment of windmills in their viewshed. Here, more than half of the individuals who view SIFs positively in a general sense still do not want one in their area. The mismatch between my results and Boyle et al. (2019) can be easily understood, though, since a windmill in one’s viewshed is likely perceived as much less costly than a facility where individuals are using hard drugs.

<table>
<thead>
<tr>
<th>Legal</th>
<th>Block/Always</th>
<th>t-Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.51</td>
<td>0.23</td>
<td>8.63</td>
<td>2.2e-16</td>
</tr>
</tbody>
</table>

B.6 A note about Kensington

Since there is so much selection bias in opting into a Dynata survey, it is difficult to make causal claims about anything other than the treatment, which is the only truly random assignment in this experiment. That being said, it is important to note that respondents who self-identified as
working or residing in Kensington expressed significantly more support for the site at all levels than other respondents in the sample. As mentioned previously, much of the buildup to Safehouse’s opening was based around Kensington, widely considered to be the epicenter of Philadelphia’s heroin problem. Safehouse had even secured a location in Kensington; however, the press conference announcement of “the facility is opening next week” pointed to South Philadelphia as the first site location, nowhere near Kensington. Kensington residents were about 25% more likely to support the site at most levels, suggesting that Safehouse may have much better luck continuing to push for a site in Kensington, rather than one in South Philadelphia.