Essays in Health and Labor Economics

by

James M. Flynn

B.A., James Madison University, 2007

M.S., Drexel University, 2016

M.A., University of Colorado Boulder, 2020

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Economics

2023

Committee Members:

Francisca Antman, Chair

Brian Cadena

Richard Mansfield

Jonathan Hughes

Amanda Stevenson

Flynn, James M. (Ph.D., Economics)

Essays in Health and Labor Economics

Thesis directed by Prof. Francisca Antman

(1) Chapter 1 - this paper demonstrates that expanding access to Long-Acting Reversible Contraceptives (LARCs) to lower-income women leads to positive selection in the health of the cohorts of children being born. I exploit the staggered timing of three privately funded programs which distributed LARCs at no cost to lower income-women in Colorado, Iowa and St. Louis. I implement an event-study design which compares trends in treated counties with other U.S. counties with similar family planning clinics, but which did not receive additional funding to distribute free LARCs. I find that expanded LARC access led to reductions of approximately 1.0 'extremely preterm' births and 1.1 infant deaths per 1,000 live births. I find significant reductions in deaths due to birth defects, maternal pregnancy complications, Sudden Infant Death Syndrome, and homicide. My estimates imply a reduction of approximately 90 infant deaths and 85 extremely preterm births per year in Colorado. Using a back of the envelope calculation, I find that the Colorado program, which was funded by a $27 million donation, led to reductions in ventilation costs for extremely preterm births of approximately $5.6 million per year. These results suggest that giving lower-income women the means to control their fertility has the potential to reduce adverse infant health outcomes and could decrease the infant mortality gap which exists between the US and other leading economies.

(2) Chapter 2 - In late January, 1990, the salary of every National Hockey League (NHL) player was suddenly disclosed, ending a decades-long culture of pay secrecy. I find that underpaid players respond to this new information by reallocating effort from defense to offense, which is more highly compensated within the league. Underpaid players begin scoring more, but allow their teams to get scored on by even more than the additional goals

they provide. Asymmetrically, overpaid players do not become more defensive-minded. Consistent with reference-dependent utility theory, I find suggestive evidence that this shift is more pronounced for underpaid players who play for teams with higher overall payrolls, as these players likely have a larger discrepancy between their actual salary and their reference point.

(3) Chapter 3 - Sugar-sweetened beverage (SSB) taxes have become an increasingly popular policy to combat the worldwide obesity epidemic, but relatively little is known about their impact on health outcomes, particularly among high school aged students. In this paper, I use public-use data from the Youth Risk Behavioral Surveillance System (YRBSS) to determine whether high school students living in three of the American cities which have implemented SSB taxes have experienced public health improvements. Using an event-study design that compares outcomes in treated districts to a group of similar control districts, I find significant reductions in both SSB consumption and average BMI, with suggestive evidence that the improvements are concentrated among female and non-white respondents.

Dedication

To Katie, Lucy, and BB.

Acknowledgements

This would not have been possible without the incredible support of my mentors and colleagues at the University of Colorado Boulder, including Francisca Antman, Brian Cadena, Amanda Stevenson, Rick Mansfield, Jonathan Hughes, Jeronimo Carballo, Tania Barham, Daniel Kaffine, Kyle Butts, Alexander Bentz, Anja Gruber, Saigeetha Narasimham, Nick Ferguson, Garet Crispin. I would also like to thank my colleagues on the Colorado Fertility Project, Sara Yeatman, Katie Genadek, Jane Menken, Stefanie Möllborn, Leslie Root, Liyang Xie, and Judith Mccabe. I would like to thank my mentors at Drexel University, Christopher Laincz, Mark Stehr, and Sebastien Bradley. Finally, I would like to thank my parents, Joe and Sarah Flynn, for instilling a love of learning in me from the very beginning, my wife, Katie Flynn, for supporting me in chasing this crazy dream, and my daughter Lucy Flynn, for making every day more fun than the last.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Can Expanding Contraceptive Access Reduce Adverse Infant Health Outcomes?

1.1 Introduction

In 2019, 5.8 out of every 1,000 infants born in the United States died before their first birthday, a rate that is over three times higher than Japan, and worse than 32 of the other 37 OECD countries ([242]). Although this disparity is widely known, there is a relative dearth of evidence as to exactly why the US lags so far behind other leading economies. One plausible explanation is the high rate of unintended pregnancies in the US, which are associated with delayed initiation of prenatal care, ([203]) low birthweight ([288], [272], [144]), and neonatal mortality ([68]). Unintended pregnancies alone cannot explain this gap, however, as the US has a similar rate of unintended pregnancies as many countries with much lower rates of infant mortality[1] . The US is unique, however, in which groups are experiencing unintended pregnancies. According to [291], low-income women in the United States are more than five times as likely to have an unintended pregnancy than higher income women, with barriers to contraceptive access playing an important role.

Long-acting reversible contraceptives (LARCs), such as intrauterine devices (IUDs) and subdermal hormonal implants, are the most effective form of reversible contraception available today, and are virtually immune to user error ([110]). Although they are cost effective in the long run, they come with high upfront cost of around $800, which low-income women are typically unable

[1] According to [43], 46% of all pregnancies from 2015-2019 in the US were unintended, while the three countries with the lowest IMRs in the OECD, Japan, Finland, and Slovenia, had rates of unintended pregnancy of 41%, 51%, and 51%, respectively.

to afford and unable to finance ([85]). They are also several times more expensive in the US than in much of Europe ([61]), and multiple studies cite cost as a major barrier to LARC access among low-income American women ([172], [66]). This leads many low-income women in the US to opt for more expensive, less effective means of contraception, even when they would prefer a LARC. Tragically, low-income women are also disproportionately likely to experience an infant death ([208]). This raises the question, then, of whether expanding LARC access to lower-income women has the potential to lower the infant mortality rate and reduce other adverse infant health outcomes.

In this paper, I answer this question by using the rollout of three privately-funded family planning programs as a source of plausibly exogenous variation in LARC access. These programs each gave thousands of LARCs for free to mostly low-income women in Colorado, Iowa and St. Louis. Using restricted-access natality data from the National Vital Statistics System (NVSS), which links birth certificates to infant death records, I implement an event-study design which compares trends in counties where private funding was received to expand LARC access with trends in other US counties which have similar family planning clinics, but which did not receive this additional funding.

My most conservative estimates find that expanded LARC access led to a reduction of just over 1.0 extremely preterm births (EPBs), which are births before 28 weeks of gestation, and 1.1 infant deaths per 1,000 live births across the three treated regions. These represent reductions of between 16-18% off of the base rates of these two outcomes[2] . I demonstrate that these results are robust to all of the recent concerns raised in the difference-in-differences with staggered treatment adoption literature. I also supplement the main analysis by estimating synthetic control specifications on each treated area separately, showing that large reductions in both adverse outcomes appear in each treated region approximately one to two years after the LARC interventions were initiated. Looking into specific causes of death, I find significant reductions in infant mortalities due to birth defects, maternal pregnancy complications, Sudden Infant Death Syndrome (SIDS),

[2] A nationwide reduction of 1.1 infant deaths per 1,000 live births would be enough to close the gap between the US and Japan by 26%.

and homicide, all of which are correlated with socioeconomic status. Additionally, I show that in Colorado and Iowa, these reductions only appear in counties with a nearby Title X clinic, where the LARC interventions were implemented, ruling out the possibility that statewide policy changes could be behind the reductions I find. Using a back of the envelope calculation, I find that the program in Colorado[3] , which cost $27 million total, led to reductions in medical ventilation costs for extremely preterm births of approximately $5.6 million per year.

This paper builds on a large literature which demonstrates how family planning access can lead to selection which impacts the health outcomes of the cohorts of children being born. A number of studies[4] find evidence that access to abortion reduces infant mortality, with [159] going so far as to say 'the increase in the legal abortion rate is the single most important factor in reductions in both white and nonwhite neonatal mortality rates' (pg. 695). This paper builds on these findings by demonstrating that LARC access also has the ability to reduce infant mortality, even in a setting where abortion is already legal. This is particularly important in light of the recent Supreme Court decision in Dobbs vs. Jackson, which overturned Roe vs. Wade and opened the door for restrictions in abortion access across the country. Since abortion and LARCs have complementary impacts on infant mortality, LARC access will become even more important as abortion access is restricted.

This paper also builds on a growing literature which documents the effects of expanding access to LARCs specifically. The programs I study in this paper have been shown to reduce unintended pregnancies ([226]), abortion ([47], [267]) and the teen birth rate ([215], [197]), while increasing female educational attainment ([296], [324]). I build on these findings by demonstrating that increased LARC access also led to a reduction in adverse infant health outcomes. The rest of the paper is organized as follows. Section two provides background on how family planning programs have impacted maternal and infant outcomes, as well as on LARCs and the three programs I study. Section three describes my data and empirical strategy, while section four presents my results on the effect of LARC access on the rates of extremely preterm births and infant mortality. Section

[3] The program in Colorado is the only one for which cost data is available

[4] [249], [104], [188] and [163]

five performs some back of the envelope calculations and concludes.

1.2 Background

1.2.1 Effects of Family Planning on Maternal and Infant Outcomes

Access to family planning services has been shown to have far-reaching impacts on health and economic outcomes for both mothers and their children, with these effects varying considerably across different forms of contraception. In addition to the many studies which find evidence that access to abortion reduces infant mortality, [99] find that abortion legalization in Mexico caused a sharp decline in maternal morbidity, while [235] finds that the liberalized access to abortion in the US in the 1970s gave agency to many young women in deciding if and when to get married and have children. A substantial literature, starting with [118][5] , tracks birth cohorts subject to legal abortion into adulthood, finding that criminality was substantially less than that of cohorts not exposed to legal abortion, though critics have disputed these findings[6] . Other research has found that cohorts exposed to abortion in utero were less likely to get pregnant as a teenager ([117]), while also being less likely become a single parent and more likely to graduate college ([25]). In this paper, I document a similar compositional impact, whereby allowing lower-income women to opt out of unwanted or unplanned pregnancies reduces the likelihood of adverse outcomes like preterm births and infant deaths.

Another literature investigates the consequences of the emergence of the birth control pill, finding some similarities and many differences in the effects of access to the pill versus abortion. In their seminal "Power of the Pill" paper, [151] exploit timing variation in state laws granting access to the pill to young women to show that it empowered women to delay the age of first marriage and lowered the cost of human capital acccumulation. [33] finds that legal access to the pill before age 21 increased the number of women in the labor force, increased their total number of hours worked

[5] [119], [117], [120], and [148]

[6] [145] point to a coding error in the original paper which weakens the results, while [186] points to changes in crack-cocaine use as a potential confounder. Both [190] and [62] investigate the abortion-crime hypothesis in European countries and fail to find an effect.

and decreased the likelihood of a birth before age 22, while [34] demonstrates that access to the pill at a younger age conferred an eight percent wage premium to young women and substantially reduced the gender wage gap in both the 1980s and 1990s.

Focusing instead on the children born to women exposed to pill access, [26] find that, in contrast to the effect of abortion access, the pill actually increased the share of children born with low birthweight and the share born to poor households in the short run. This effect appears to be driven by upwardly-mobile women delaying child-bearing while poorer women were not able to do so. These effects balanced out in the long run as these women began having children later. This paper highlights the importance of which types of women a contraceptive technology is made available to. While the birth control pill was a revolutionary breakthrough, it was not cheap and was rarely covered by insurance, meaning it was not available to all women who wished to use it.

This raises the question, then, of how expanding access to LARCs will shift the composition of births and whether the children born to women with this access will be healthier than their counterparts. LARC methods have become increasingly popular in recent years, with the American College of Obstetricians and Gynecologists recommending them as a first-line option for most women seeking to avoid pregnancy ([8]). Because of the high upfront costs, LARCs are difficult for low-income women to afford, even though they are cost effective in the long run. This suggests that, absent some type of intervention, the effects of LARC access are likely to be similar to the effects of access to the birth control pill. The programs I study in this paper, however, focus on expanding LARC access to low-income women specifically, so there is potential for them to give these women the same economic freedom that more upwardly mobile women attained with the emergence of the pill.

This means these programs have the potential to improve infant health both in the short run by shifting the composition of pregnancies towards potentially healthier ones, and in the long run by giving young women the power to delay pregnancies until they are economically better off and more capable of investing in their children. These programs have been shown to have many important benefits for the women using them, but relatively little is known about their effect on the health of the infants born to women with expanded LARC access. Because births to young, low-income women are more likely to result in a preterm birth or an infant death ([208], [140]), it seems plausible that these programs could reduce these outcomes, but to my knowledge no current work has drawn a causal link between programs that expand access to LARCs to low-income women and infant health outcomes.

1.2.2 Long-Acting Reversible Contraceptives

Long-Acting Reversible Contraceptives (LARCs), namely intrauterine devices and subdermal hormonal implants, are the most effective reversible contraceptive methods available, approximately 20 times more effective than pills, patches, and rings ([110]). LARCs are greater than 99% effective and can prevent pregnancy for anywhere from three to 10 years ([85]). As [297] point out, "they are not dependent on compliance with a pill-taking regimen, remembering to change a patch or ring, or coming back to the clinician for an injection."

LARCs are just as effective as sterilization ([205]), with the added benefit of not being permanent, and because they require no further action from the user after insertion, they are almost immune to user error. Oral contraceptives, the patch, and condoms are less effective than LARCs even when used perfectly, and they have much higher rates of user error ([307]). This risk of contraceptive failure is particularly high among low-income women ([300]), suggesting that making LARCs more available to low-income women has the potential to prevent many unwanted pregnancies. LARC users are also generally satisfied with their choice of contraception, with [254] finding that LARC users have higher rates of satisfaction than oral contraceptive users (80% compared with 54%) and are more likely to continue using them beyond a year (86% compared with 55%). Similarly, [116] find that LARC users are more likely to continue use after six months than users of oral contraceptives, and are at no greater risk of sexually transmitted infections.

Despite the many benefits to LARC usage, only 8.5% of women who were using a contraceptive in 2009 were using a LARC ([194]). Multiple explanations account for this disconnect. One reason is information. [271] documents a series of pervasive myths about LARC use, including that they cause disease, infertility, menstrual irregularities, weight gain, acne and hair loss. [49] and [226] both demonstrate that when women receive information about LARCs from a doctor, they become more likely to request them. There are also supply side issues, where not all healthcare providers have the equipment or training necessary to insert LARCs. [53] document that 21% of health centers have no staff trained in LARC methods, and that almost half (48%) of health centers do not offer IUDs or implants onsite, with these clinics concentrated in rural areas. Arguably the biggest impediment to LARC use, however, is the high upfront cost of up to $800 ([85]). Even though LARCs are cost effective for most users in the long-run[7] , many women, particularly low-income women, cannot afford to pay the upfront costs and end up using more expensive, less effective methods.

Multiple studies document an unmet demand for LARCs, with cost being the most frequently cited barrier to adoption and use ([172], [66]). [256] find that among women interviewed six months postpartum, two thirds had experienced a barrier to accessing their preferred method of contraception, while [257] find that 34% of postpartum women using less effective methods would prefer to be using LARCs. Because of this unsatisfied demand, there is potential for programs which improve LARC access to generate substantial improvements in public health, both for the women using LARCs and the children born to women with this improved access. I now describe the three large-scale programs which were rolled out with the express intention of addressing this unmet demand for LARCs.

1.2.3 The Colorado Family Planning Initiative

In 2009, the Colorado Department of Public Health and Environment (CDPHE) implemented the Colorado Family Planning Initiative (CFPI) with the help of an anonymous donor, who funded the program with a $27 million dollar donation. The goal of the program was to reduce unintended pregnancies in Colorado by increasing the number of family planning clients served and by increasing

[7] Oral contraceptives can cost up to $50 a month, which means that LARC methods can be cheaper as long as they are used for more than 16 months.

access to LARC methods[8] . The CFPI was implemented through Title X family planning clinics, which were operated in 38 of Colorado's 64 counties. Although just over half of Colorado counties had a Title X clinic, they are mainly located in population centers, and 91% of Colorado's population lived in a county with a Title X clinic in 2009. The guidelines for Title X clinics require that all contraceptive methods and counseling be provided at no cost to clients below 200 percent of the federal poverty line, but this was not always feasible with LARCs because clinics could not afford to purchase LARCs, did not have enough doctors trained in inserting LARCs, and lacked the clinic capacity to accommodate LARC insertions. The money from the CFPI went directly to fixing these bottlenecks, and made it possible for clinics to dramatically increase LARC insertions in Colorado.

While very few LARC's had been inserted in Colorado Title X clinics prior to the CFPI[9] , by the end of 2009 there had been almost 2,000 new insertions, and this number grew in each subsequent year. In each year from 2010-2014 between 4,000-7,000 LARCs were inserted at Title X clinics in Colorado, so that just under 30,000 had been given out by 2014, which translates to approximately one LARC per 24 women aged 15-35 in Colorado. In 2013, over 24% of Colorado teens visiting Title X clinics were LARC users, the highest rate of any state in the U.S. At the time, over 40 states had less than 10% of their Title X clients using LARCs, according the the Title X Family Planning Annual Report of 2013 ([238]). In response to the CFPI, teen pregnancy rates declined in Colorado counties with Title X clinics, with the largest impacts occurring in counties with high poverty rates ([215]), indicating that the CFPI made a significant difference for young, low-income women in Colorado.

1.2.4 Iowa Initiative to Reduce Unintended Pregnancies

In 2007, The Iowa Initiative to Reduce Unintended Pregnancies (IIRUP) was launched. The IIRUP was also funded by an anonymous donor and was aimed at reducing unintended pregnan-

[8] My discussion of the implementation of the CFPI draws on the detailed descriptions provided by [85], [215], and [267]

[9] The CDPHE reports that approximately 2,000 Title X clients in Colorado were LARC users in 2008

cies among women aged 18 to 30[10] . Although both Colorado and Iowa were targeted with an intervention to reduce unintended pregnancies, there is little evidence to suggest that they had above average rates to begin with. According to a report by the Centers for Disease Control and Prevention, Colorado and Iowa had the 31st and 35th highest rates of abortion among US States in 2006, respectively ([253]). Similar to the CFPI, the IIRUP was implemented through Title X family planning agencies, which operated 81 clinical sites across 46 of Iowa's 99 counties, covering 83% of the state's population. The funding was used to expand hours and locations, train clinic staff on how to talk about LARCs with patients, and purchase IUDs and implants which the clinics had previously been unable to afford. After receiving funding, all of Iowa's Title X agencies began offering LARCs.

LARC takeup increased dramatically in response to the Initiative. While only 1,047 Title X clients were using a LARC method in 2006, that number ballooned up to 10,092 by 2009 as 15% of all Title X clients were LARC users in 2009. Estimating a causal impact of the IIRUP on health outcomes is complicated by the fact that abortion access also increased in Iowa at the same time, with medication abortion via telemedicine becoming available in 2008. [47] demonstrate, however, that abortion in Iowa actually declined from a rate of 8.7 per 1,000 reproductive-age women in 2005 to 6.7 in 2012, so it seems more likely that any effects we see in this period are due to the IIRUP as opposed to the increased access to abortion.

1.2.5 St. Louis Contraceptive CHOICE Project

Also in 2007, researchers based at Washington University in St. Louis launched the St. Louis Contraceptive CHOICE Project (SLCCP) in order to study the contraceptive choices women make when cost and access barriers are removed and they are educated about the benefits of different contraceptive methods. The privately-funded study enrolled over 9,000 women aged 14-45 in the St. Louis metropolitan area who had been sexually active in the past six months or planned to

[10] My discussion of the Iowa Initiative to Reduce Unintended Pregnancies and the St. Louis Contraceptive CHOICE Project draw heavily from [298], [226], [48] and [47]

be sexually active in the next six months, wanted to avoid pregnancy for at least a year and were interested in trying a new form of contraception. The women were all read a script describing LARC methods, were counseled on the full range of contraceptive methods available and were screened for STIs. Once the participant chose a contraceptive method and it was approved by a phyisican, she received it at no cost for up to three years, and was allowed to change methods at any point. 75% of the participants chose a LARC method, which means that approximately 7,000 LARCs were inserted between 2007-2011, and the rates of teen pregnancy and abortion for women in the study were both four times lower than the national average.

In recent years, the 'tiered-effectiveness' counseling method used in the CHOICE project has come under criticism. In tiered-effectiveness counseling, the most effective methods are explained first so that patients can use information about the relative efficacy of contraceptive methods to make an informed choice. This approach was motivated by research in the early 2000's which found that misinformation among both patients and providers led to a low prevalence of LARC use ([28], [231]). This led to both the American College of Obstetricians and Gynecologists (ACOG) and the American Academy of Pediatrics (AAP) recommending that LARCs be offered as the 'first-line' of contraceptive methods to all patients.

Many have since argued that this conflicts with patient-centered care and actually coerces patients to use LARCs even if they would prefer another method ([56], [154]). The main concern is that if physicians believe that their job is to convince patients to use LARCs, they may believe that patients who wish to use other methods are only doing so because they have not been educated enough. This can lead to coercive behavior, particularly towards low-income patients, patients of color, and other vulnerable populations. This can be especially problematic with LARCs because they are more inherently capable of being used coercively, since patients need a doctor to remove them and cannot stop using them on their own.

While these three different initiatives had many differences in the populations they were serving and the scale and scope of their operations, they had several important characteristics in common that are useful for the purposes of this study. First, they all reduced the cost barrier of

LARC methods to low-income women by providing LARC insertions free of charge. In response to each program there was a dramatic uptick in the number of LARCs being used. The CFPI was the largest and most successful of the three initiatives, so we may expect to find larger effect sizes in Colorado, but if LARC access has a causal impact on infant health we should expect to see improvements in all three areas.

1.3 Empirical Approach

This section details the data used in my analysis as well as my strategy for estimating the causal effects of expanded LARC access on infant health outcomes.

1.3.1 Data

This paper uses data from several sources. Data on both 'extremely preterm' births (EPBs) and infant mortality come from restricted-access linked birth and infant death data from the National Vital Statistics System (NVSS). This data includes information from birth records for all live births which took place in the United States from 2002-2015. This includes the number of weeks of gestation, from which I calculate whether the birth was deemed 'extremely preterm', and also the county of residence of the mother, which I use to infer whether or not she lived in a treated county when the child was born. It also includes an indicator for whether that birth resulted in an infant death, and if so, it includes information from the death record including how old the infant was when they died and what the primary cause of death was.

This data allows me to calculate county-wide rates for both infant mortality and EPBs for each year. EPBs are important to measure independently of infant mortality, because although roughly 75% of EPBs will survive ([252]), these children are much more likely to suffer from serious cognitive and developmental disabilities ([282], [255]). In one sense, we can consider the infant mortality rate to measure the extensive margin of whether a child survives, while the rate of EPB measures the potential quality of life a child faces on the intensive margin.

Because not all counties in Colorado and Iowa have Title X clinics through which the LARC interventions were implemented, I define a county in these two states as treated if it had a Title X clinic in 2008. This clinic assignment was gathered by [215], with Colorado counties identified based on clinic addresses in the Colorado Department of Public Health and Environment's Directory of Family Planning Services. Clinics in other states were identified by geocoding the addresses of Title X clinics listed in the US Department of Health and Human Service's 340B Database. My event-study specifications will thus compare trends in infant health in treated counties (counties in Colorado and Iowa with Title X clinics as well as St. Louis county) with other counties in the U.S. which have similar Title X family planning clinics but which were not given additional funding specifically for a LARC program. Additionally, because infant mortality declined in states which expanded Medicaid after the passage of the Affordable Care Act of 2010 ([45]) relative to states which did not, I only include counties in the 39 states which expanded Medicaid as possible control counties since Colorado, Iowa and Missouri all expanded Medicaid.

To control for time-varying county characteristics, I use population data from the National Cancer Institute's Surveillance, Epidemiology and End Results Program (SEER) to construct demographic measures for the percent of the population that are teenagers (15-19 years old), the percent of the population which is Black, and the percent which is Hispanic. To control for time-varying economic conditions, I use county-level unemployment and poverty rates from the Bureau of Labor Statistics. Finally, I include two additional indicator variables which control for state-level policies. The first is whether emergency contraceptives are available over-the-counter, while the second controls for whether private insurance plans covering prescription drugs are required to cover any FDA-approved contraceptive. These variables were initially constructed by [215] using data collected from the National Conference of [244], the National Women's Law [86], and [329].

1.3.2 Methodology

I estimate the effect of expanding LARC access on infant health outcomes through two primary methodologies. First, I use event-study specifications of the form:

$$Y_{ct} = \sum_{k=-4}^{4} \theta_k LARC_{c,t+k} + \beta X_{ct} + \alpha_c + \gamma_t + \psi_i * t + \epsilon_{ct} \tag{1.1}$$

in order to estimate the joint effect of all three programs. Here, Y_{ct} measures the rate of a specific health outcome, either the rate of EPBs or infant mortality, for county c in year t. $LARC_c$ is an indicator for a county being treated with a LARC intervention at some point during the sample period, while k measures the years before and after the intervention took place. Therefore, θ_{-4} through θ_{-1} estimate differences in trends between treated and control counties before the LARC interventions went into effect and θ_1 through θ_4 measure the impact of the policies. If the LARC interventions had a causal impact on infant health outcomes, we should expect θ_{-4} through θ_{-1} to be close to zero and statistically insignificant, while θ_1 through θ_4 should be negative and significant. X_{ct} includes a vector of control variables that could impact infant health outcomes. α_c are county fixed effects, which control for time-invariant characteristics of each county which impact infant health, while γ_t are year fixed effects which control for ntaionwide trends in infant health across time. ψ_{c*t} is a county-specific linear time trend, which I include to prevent preexisting differences in trends between treated and control counties from being picked up as a treatment effect. I estimate this specification using weighted-least-squares, where the weights are determined by the total number of births in a county-year cell.

There are two reasons why I expect the effect of LARC access to increase over time. First, the policies continued over a period of several years, and we would expect their effects to be cumulative. Taking Colorado as an example, there were only about 2,000 LARCs inserted via the CFPI in 2009, but an additional 4,200 were inserted in 2010 and then between 5,000-7,000 were inserted in each subsequent year until 2015. Since women who received a LARC in 2010 were able to keep it for up to 10 years, the total stock of women protected by a LARC was increasing over time. Additionally, because of the unpredictable timing of sexual activity, even after LARCs are inserted we would not expect to see an immediate change in the number of unwanted births. In the counterfactual world without expanded LARC access, many unprotected women would still not get pregnant and even those that do would be unlikely to get pregnant right away. Therefore, the number of births that were avoided in each year would be increasing over time[11] , as would any effects this has on infant health outcomes.

I include all counties which had a clinic where free LARCs were distributed as treated. This includes 38 counties in Colorado and 46 counties in Iowa which had Title X clinics through which the CFPI and IIRUP were implemented, as well as St. Louis county. In choosing control counties, I begin with all counties which also have a Title X clinic but which did not receive additional funding for LARCs. I then exclude all counties in the 12 states[12] which did not expand medicaid following the passage of the Affordable Care Act of 2010, as [45] has demonstrated that infant mortality went up in these states relative to expanding states, which could bias my estimates.

I also drop all counties in Iowa and Colorado without a Title X clinic as well as all counties in neighboring states which border a treated county because of concerns over potential spillover effects. Since women could travel from neighboring counties to ones with a Title X clinic, these counties can be considered partially treated. Including them in the treated group could bias my estimates downward as the effects are almost certainly smaller for counties where it is more difficult to obtain LARCs. Including them as control counties could also bias my estimates downward by including counties which received a partial treatment in my control group, violating the Stable Unit Treatment Value Assumption (SUTVA). The easiest way to avoid these issues is by dropping these counties entirely ([69]). Overall, this approach results in 85 treated counties and 1,325 control counties. Figure 1.1 displays treated counties in red and control counties in light blue.

I supplement the event-study specifications by separately estimating the synthetic control method (SCM) of [6] and [5] separately on each treated region in the Appendix. The SCM constructs a control group which is a weighted average of all the possible controls, where the non-negative weights are determined by minimizing the sum of squared pretreatment differences between the

[11] This is consistent with the findings of [215] and [197] which find virtually no impact of the CFPI in 2009, and then a small decrease in 2010 which increases in 2011 and 2012

[12] Florida, Georgia, South Carolina, North Carolina, Tennessee, Alabama, Mississippi, Texas, Kansas, South Dakota and Wyoming

treated group and the synthetic control. This approach has both benefits and drawbacks which, overall, complement those of the event-study specifications. I estimate state-level synthetic controls on both Colorado and Iowa and a county-level synthetic control on St. Louis. One important benefit of this approach is that it selects a specific control group for each of the treated regions. Colorado, Iowa, and St. Louis are different places, each with its own idiosyncratic populations, so there is no perfect control group for the three of them combined. By allowing the SCM to select a control group for each of the treated regions separately, I am able to demonstrate that the effects I find in the combined regressions also show up individually for each treated area when compared with a control group chosen specifically to satisfy the equal counterfactual trends assumption for that area on its own.

One drawback of the state-level models is that since many counties in both Iowa and Colorado do not have Title X clinics, this approach essentially includes many untreated or only partially treated areas as treated. This will result in an understatement of the overall treatment effect and will bias my estimates toward zero. For both Colorado and Iowa, I correct for this issue by both running the SCM on the entire state and then again by first dropping all non-Title X counties from my sample before estimating. This will remove this bias and also serve as a falsification test, as when the untreated counties are removed, the treatment effects should at least stay as large, if not increase. If they were to decline after this procedure it would raise concerns that the effects I am picking up are from some other factor not related to the LARC interventions. Additionally, because the outcomes I am tracking are rare and somewhat noisy and because it is important for the SCM to match treated and control groups based on underlying trends and not on idiosyncratic noise, I also estimate the SCM on each group using a three-year moving average of the outcome of interest, which removes a substantial amount of noise without compromising the trends occurring in the data. In all cases, the economic inference is similar.

1.4 Results

This section details my estimates of the effect of expanded LARC access on both EPBs and overall infant mortality. Figure 1.2 displays overall trends in each of these outcomes in Colorado counties with a Title X clinic, compared with the annual number of LARCs inserted through the CFPI. For both outcomes, the rates hover between 5.7 and 6.5 occurrences per 1,000 live births from 2003 to 2009 with some noise but no apparent trend. As LARCs begin to be given away via the CFPI in 2009, both rates are at local maxima near 6.5, but begin to decline shortly thereafter. Both fall slightly in 2010 but then more aggressively in 2011 and 2012 as more and more LARCs are inserted.

These staggered declines make sense as it would take time after each insertion for a birth that would have happenned in the counterfactual world to be avoided. Both rates settle after 2012 to values mostly between 4.5 and 5.5 occurrences per 1,000 births, with reductions of greater than one occurrence per 1,000 each. In the remainder of this section I will argue that this relationship is a causal impact of the CFPI. First, I will focus on the EPB outcome and show that it occurred not just in Colorado but also in St. Louis and Iowa after similar LARC interventions and that it cannot be explained by changing demographics, economic indicators, policy changes or pre-existing trends.

1.4.1 Extremely Preterm Births

Table 1.1 displays estimates of the event-study specification outlined in equation (1), with coefficients detailing the changing rates of EPB across the three treated regions for three years before and four years after the LARC interventions were initiated. This means that for St. Louis and Iowa, pretreated estimates are displayed for 2004-2006 while postreatment estimates are displayed for 2008-2011. Likewise, for Colorado the pretreated estimates are for 2006-2008 while the postreatment estimates are for 2010-2013. The top panel of Table 1.1 displays the estimates on the pretreatment leads while the bottom panel displays estimates for the postreatment lags. The first thing to notice is that while all of the estimates in the top panel are insignificant at the 10% level and take on both positive and negative values, all of the posttreamtment lag coefficients are negative and all but two of the estimates for years two through four are significant at 5%.

Column one includes all three treated regions but does not include any controls beyond county and year fixed effects. The pretreatment leads are all negative and insignificant, which suggests EPBs may actually have been rising very slightly in the LARC-treated areas prior to the interventions. Still, the average difference is only 0.24 EPBs per 1,000 births and the p-value from a test that the average effect is zero is .3612. After the intervention, there is a small and insignificant decline in the first year, followed by declines of between 0.8 and 1.4 EPBs per 1,000 live births for years two through four, with each of these estimates significant at 5%. In column two I add county-specific linear trends to control for pre-existing patterns in the treated counties. If anything, these trends were biasing the estimates in column one towards zero. Each of the pretreated leads is now smaller in magnitude and less significant, with an average effect of just 0.15. Each of the posttreatment lags, on the other hand, is now larger and more significant with an average effect in years two through four of -1.72 EPBs per 1.000 live births. In columns three I add the demographic and economic controls, while in column four I include the two policy controls, and the story is roughly the same.

Of course, this standard two-way fixed effects (TWFE) model has come under scrutiny recently when treatments are staggered ([155], [71], [299], [54]), especially when there is potential for heterogeneous treatment effects over time. One of the main concerns is that later treated observations will be compared with earlier treated units whose treatment effects have been growing over time. This can even cause the parameter of interest to flip signs in certain situations leading to flawed inference. Although the treatment effects in this specification are staggered, these concerns do not present a serious threat to identification here because there are 1,325 untreated counties and only 85 treated counties, meaning the vast majority of 2x2 comparisons are between treated units and never-treated controls. The [155] decomposition of the TWFE specification is presented in Appendix Table A.1. 99.3% of the weight in the specification is from comparisons of treated units versus never-treated units, while only 0.4% of the weight is from the problematic later treated vs. earlier treated comparison group. Even though this comparison does bias the difference-in-difference estimate towards zero, it bears so little weight in the regression that it effectively makes no impact.

Columns five and six further address these concerns by removing the staggered component of the treatment and estimating the effects separately based on treatment timing. Column five estimates equation (1) on Colorado alone, and there is no concern over pretreatment trends in this specification. The leads are all small in magnitude and insignificant, and they bounce back and forth around zero. The lag on the first year after treatment is again negative but insignificant, while the lags on years two through four are all larger than in the previous columns and individually significant at 1%. In column six it becomes clear that the potentially concerning pretrends occurred in the St. Louis and Iowa sample, which was treated in 2007. Pretreatment leads decline from .65 to .41 and then .01 before remaining steady in 2007 and then declining much further. Still, the difference in the posttreated years is much larger than the changes taking place beforehand.

What is clear from Table 1.1 is that EPBs dropped substantially in the second through fourth year after treatment across all three interventions. To illustrate this, Figure 1.3 displays the coefficient estimates for columns 4-6 of Table 1.1 with 95% confidence intervals. In both the full sample and Iowa and St. Louis graphs, there is a slight pretrend leading up to treatment, but all pretreatment estimates are small and insignificant and there is much larger decline that occurs after the interventions. The Colorado graph demonstrates that pretreated outcomes track very closely with the control group before the CFPI and then a large reduction occurs in the second year after treatment which stays around 2 EPBs per 1,000 live births for each of years two, three and four.

Figure 1.4 displays estimates from columns 4-6 once again, only reestimated using [149]'s two-stage difference-in-difference estimator, which is robust to heterogeneous treatment effects with staggered timing. Here, because year and county fixed effects could be contaminated by the treatment effect, these effects are all estimated in a first stage using untreated observations to get year and county fixed effects[13] . The first stage is then residualized and regressed on the leads and lags, resulting in fixed effects that are uncontaminated and parameter estimates which are robust to heterogeneous treatment timing. In Figure 1.4, the parallel trends assumption looks even more plausible in both the full sample and Iowa and St. Louis graphs, suggesting that perhaps the fixed effects had been contaminated by the treatment effect in these groups. In each version there is little to no movement in the pretreatment period, followed by a small decline in the first year after treatment and then a larger, statistically significant decline in the second period after treatment.

1.4.2 Infant Mortality

Table 1.2 presents estimates of equation (1) with the infant mortality rate (IMR) replacing 'extremely preterm' births on the left-hand side. As with EPB's, there does not appear to be much movement in the three years before the interventions, and then there are large declines concentrated in years two through four following treatment. In the baseline two-way fixed effects model, IMR actually appears to be increasing in the treated areas relative to the control counties in the years before treatment, with that trend reversing after the LARC interventions began. Including county linear trends and demographic, economic and policy controls both reduce the pre-treatment differences and increases the post-treatment effect. In Colorado, the pretrends are somewhat concerning, though as in the basic TWFE model they appear to actually be increasing prior to treatment, and there is still a large decrease of 1.6 infant deaths per 1,000 live births by year three. The results are even larger when looking at only St. Louis and Iowa, with almost no movement prior to treatment and a large decline of between 1.5 and 2.5 infant deaths per 1,000 live births in year two through four.

Figure 1.5 display estimates of the same model, only this time using [149]'s two-stage difference-in-difference estimator. As before, the overall effect sizes are now somewhat smaller, but the pre-treatment trends look considerably more stable. For the full sample, there is an average treatment

[13] In other words, year fixed effects are estimated from the full group of never treated observations, while county fixed effects for treated observations are estimated using only the pretreated observations from these groups.

effect of 1.1 infant death per 1,000 live births for years two through four after treatment, with similar effects showing up when the samples are run separately to avoid concerns over the staggered treatment timing.

1.4.3 Infant Mortality by Cause of Death

Figure 1.6 displays coefficient estimates from the full-specified TWFE model separately for each of the six most commonly listed causes of death in the NVSS data. Large decreases appear for deaths due to birth defects, SIDS, maternal pregnancy complications and homicide. Deaths due to prematurity and low birth weight were actually declining throughout the period. Deaths from injuries do not respond to LARC treatment, which is comforting as it seems unlikely that they would be impacted. Low socioeconomic status is associated with higher rates of SIDS ([30]), birth defects ([323]) and maternal pregnancy complications ([200]), so it makes sense that expanding LARC access to low-income women might improve these outcomes.

The decrease in deaths due to homicide is perhaps the most surprising result, though unwanted births have been shown to increase the risk of violence to the mother ([270]) and can prolong the mother's relationship with the father in the short term ([225]). If these unwanted pregnancies are causing women to stay longer in potentially abusive relationships, that mechanism could also explain the reduction in infant deaths due to homicide.

1.4.3.1 Where are These Improvements Happening?

So far, it has been established that large, statistically significant declines in EPBs and infant mortality occurred in the regions treated with a LARC intervention. In order to establish a causal impact of LARC access on this outcome, however, it is important that the treatment effects are concentrated near the Title X clinics through which the programs were implemented. In this section, I compare counts of EPBs and infant mortalities in treated versus untreated counties in Colorado and Iowa, in order to rule out any statewide policies which could have impacted infant health across the entire state. Since non Title X counties were not used as controls, it is not important that they satisfy the equal counterfactual trends assumption, but it should be the case that any treatment effect which shows up should predominantly occur in counties with Title X clinics.

To that end, Figure 1.7 displays the raw number of EPB cases for both Colorado and Iowa, broken out by whether or not they were born to a resident of a county with a Title X clinic. The top left graph displays the EPB count over time for counties with a Title X clinic in Colorado. As over 90% of births occur to such women, it is perhaps not surprising that the overall shape of the graph in the top right panel looks similar to the trends for Colorado overall. From 2003 to 2009, the count hovers between 400 and 450. The rate drops slightly in 2010 before declining sharply in 2011 and then remaining between 300 and 350 for the remainder of the sample. The count in counties without a Title X clinic tell a very different story, fluctuating apparently at random throughout much of the sample period and actually rising from 2009-2012 when EPBs were falling throughout the treated counties.

The story is similar for treated counties in Iowa, with rising EPB counts from 2003 through 2008, before a dramatic decline in which the count dropped from 210 to just under 170 before rebounding somewhat. Non Title X counties also show a decline around this time, but this looks similar to the noise which occurred throughout the sample and does not necessarily look like a treatment effect. The rebound in non Title X counties is also much larger, and brings the total in 2010 to a point even higher than it was before the IIRUP. Additionally, it is worth pointing out that counties in Iowa are much smaller on average than counties in Colorado, so a resident of an untreated county in Iowa would not have to travel nearly as far to get to a treated county as a resident of an untreated county in Colorado might. This could explain why the contrast between treated and untreated counties is not as clear in Iowa as it is in Colorado. In order to determine whether spillovers onto neighboring counties could explain the reduction in non-Title X Iowa counties from 2007-2009, Appendix Figure A.1 displays a map of Iowa, with all Title X counties in blue and all Title X counties which saw a reduction in the average number of EPBs from 2004-2006 to 2007-2009 in red. All of the untreated Iowa counties which experienced EPB reductions border treated counties, so it seems plausible that this reduction could have been caused by spillovers. Regardless, whatever caused the dramatic decline in EPBs in Colorado and Iowa was mainly happening in the counties which had Title X clinics, which is consistent with the hypothesis that expanded LARC access was the driving force behind the reduction.

Figure 1.8 repeats this process for infant mortality counts, and the results tell roughly the same story. Counts for Colorado Title X counties hover around 400 from 2003 to 2009, before declining each subsequent until 2012, where the count settles around 300 per year. In non Title X counties in Colorado, counts actually reach a minimum in 2009, before rebounding back up to pre-CFPI levels. There appears to be a clear treatment effect in Colorado Title X counties, but none in non Title X counties.

For Iowa, infant mortalities also rise from 2004 to 2008 before declining from around 180 in 2008 to 130 in 2010. Again there is a slight rebound, but infant mortality cases are still far less common in Title X counties in the years following the IIRUP than in the years preceding it. For non Title X counties, there is again no clear treatment effect. Counts fluctuate apparently at random from 2003 to 2010 and do not appear to be meaningfully effect by the IIRUP.

Figure 1.9 displays the change in the average number of EPBs and infant deaths per year from the four years before to the four years after a LARC intervention for Colorado and Iowa counties compared with the number of Title X clinics in that county. Each observation is weighted by the average number of births per year. As both the change in the outcomes and the number of clinics are highly correlated with population, it is not surprising that the counties with the most clinics saw the largest declines, but it is comforting that there appears to be a dose response, where more clinics typically translates to a larger decline, even among relatively similarly sized circles. The negative relationship is particularly clear for EPB's, where large declines occur in the population centers of Denver and Des Moines. For infant deaths, the relationship is less obvious, but still shows that more populated areas with multiple clinics had the largest improvements.

Finally, Appendix Figures A.2 and A.3 display placebo event-study specifications, where the untreated counties in Colorado and Iowa are considered treated, and are compared with the other untreated counties across the US. In both cases, there is clearly no treatment effect showing up, which suggests that it is unlikely that some other statewide intervention is behind the results I documented in the previous section. Overall, it appears that the declining rates of both EPB and infant mortality occurred mainly in areas which had the most access to LARCs through the CFPI and IIRUP.

1.5 Back of the Envelope Calculations/Conclusion

1.5.1 Back of the Envelope Calculations

1.5.1.1 *

Avoided Ventilation Costs for EPBs

EPBs are a tragic and traumatic event, but they are also incredibly costly as the procedures used to treat EPBs are quite expensive. In order to understand the cost savings in care for EPBs in Colorado, I use a series of estimates from [169]. This paper details the likelihood of receiving medical ventilation as a result of births at various gestational ages, conditioned on whether or not the infant ultimately survives, along with the associated average costs of ventilation for each gestational age/survival cell. The authors use 2009 data from the Agency for Healthcare Research and Quality (AHRQ). My estimates from section 4 imply a reduction of around 85 EPBs per year in Colorado following the CFPI. Of these EPBs, many would have received ventilation. The likelihood of receiving this care varies both by the gestational age of the birth, and by whether or not the infant ultimately survived.

Because of this, I first demonstrate that the reductions in EPBs associated with the CFPI did not cause large changes in the proportions of EPBs occurring at each gestational age. For example, if the reductions in EPBs were all occurring in births at less than 24 weeks of gestation, this would imply very different cost savings than if they were all occurring at 27 weeks of gestation. Table 1.3 displays the proportion of EPBs in Colorado which occurred at each gestational length, both before and after the Colorado Family Planning Initiative. Column 3 of this table displays a p-value for whether the proportion of EPBs at each gestational age are the same before and after the CFPI. In each row, equality cannot be rejected at .05, suggesting that the distribution of gestational ages did not change in response to the CFPI. There is suggestive evidence of a relative reduction in births at 24 weeks (p-value = .092) and a relative increase in births at 25 weeks (p-value = .061), but I will demonstrate that the overall costs associated with EPBs at these gestational ages are similar. This means that it is reasonable to treat the 85 EPB reduction as if it had the same proportions in each gestational age bracket as all of Colorado.

This assumption allows me to calculate the proportion of the 85 EPBs which would have come from each gestational age group. I then use the estimates from [169] to calculate the likelihood of survival, the probability of being ventilated given survival or non-survival, and then the average cost avoided for each of the 85 EPBs. Tables 1.4 and 1.5 display the estimates of this calculation for the proportion of the 85 EPBs which would be predicted to survive and not survive, respectively. For each gestational age, the total cost avoided, conditional on survival outcome, is equal to:

$$Cost_{age,survival} = N_{age,survival} * P(Vent|age, survival) * VentCost_{age,survival} \tag{1.2}$$

Where $N_{age,survived}$ is the predicted number of EPBs at that gestational age to survive (or not survive), $P(Vent|age, survival)$ is the probability of being ventilated conditional on being in that gestational age group and survival outcome, and $VentCost_{age,survival}$ is the average cost of ventilation care, also conditional on being in that gestational age group and survival outcome. The total costs avoided due to the CFPI can then be calculated by summing the individual avoided costs across each gestational age/survival cell.

Column 1 of Table 1.4 distributes the 85 EPB reduction across the gestational age categories based on their proportional occurrence in Colorado from 2006-2013. Column 2 displays the likelihood of survival for an EPB of that gestational age, based on [169][14] . Column 3 then calculates the predicted number of counterfactual EPBs which would have survived, given the prevailing survival odds. Column 4 gives the probability of being ventilated, conditional on gestational age and ultimate survival, while column 5 displays the average cost of ventilation for that gestational age,

[14] Note that the odds are identical for gestational ages 25-27 because [169] group these together in their estimates

also conditional on survival. Finally, column 6 presents the estimated costs avoided, which is the product of columns 3-5, as displayed in equation 2. For surviving EBPs, the predicted cost savings across gestational age categories totals $5.4 million annually.

Table 1.5 repeats this exercise for EPBs which would be predicted not to survive. It is notable that the average cost of ventilation is much smaller for an infant who does not survive, as infants who do survive can require ventilation for weeks and even months. The predicted cost savings from EPBs who would be predicted not to survive is only around $192,000, which means the total costs avoided are roughly $5.6 million. Considering the fact that the CFPI only cost $27 million, this means the program could pay for itself in avoided ventilation costs for extremely preterm births in 4.8 years. Of course this is only one of many potential avoided costs associated with EPBs and infant mortalities. The EPBs who ultimately survived would have likely experienced higher than average medical costs throughout their entire lives, to say nothing of the effect of these traumatic events on the health and wellbeing of the parents and their friends and family members.

1.5.1.2 Cost per Infant Death Avoided

Using the event-study estimates for Colorado, the CFPI appears to have reduced infant mortality by an average of 1.4 infant deaths per 1,000 live births for 2011-2013. With a back-of-the envelope calculation, this translates to 74 avoided infant deaths in 2011, 104 in 2012, and 93 in 2013, for a total of 271 avoided infant deaths in the first four years after the CFPI was implemented. If avoiding infant deaths were the entire goal of the CFPI, it would have cost approximately $99,600 per infant death avoided.

1.5.2 Conclusion

This paper uses the staggered implementation of three privately-funded family planning programs to investigate whether expanding access to long-acting reversible contraceptives to low-income women can reduce adverse infant health outcomes. Because these women are the most likely to experience an extremely preterm birth or an infant death, improving their ability to avoid unwanted pregnancies has the potential to create positive selection in the health of the cohorts of children being born. By comparing trends in treated counties with trends in other counties across the United States with similar family planning clinics which did not receive additional funding specifically to improve LARC access, I demonstrate that expanded LARC access led to large reductions in both the rates of 'extremely preterm' births and overall infant mortality.

The programs I study in this paper have been shown to have many other important benefits, including reducing the teen birth rate and increasing female human capital accumulation. I demonstrate an important unintended consequence of expanding LARC access to low-income women, in that it creates positive selection in the health of the cohorts of children being born. These results are particularly important in light of the recent Supreme Court decision in Dobbs vs. Jackson, as legalizing abortion and expanding LARC access to low-income women both appear to reduce adverse infant health outcomes. As abortion becomes more restrictive in many states, effective contraceptive access will become even more important.

Table 1.1: Event-Study Specifications Measuring the Effect of LARC Access on the rate of Extremely Preterm Births - 2003-2015

	(1) EPB	(2) EPB	(3) EPB	(4) EPB	(5) EPB	(6) EPB
3 Years Before	-0.305	0.285	0.504	0.652	-0.101	0.748
	(0.344)	(0.335)	(0.385)	(0.393)	(0.504)	(0.681)
2 Years Before	-0.203	0.193	0.374	0.466	0.0377	0.463
	(0.307)	(0.348)	(0.370)	(0.376)	(0.416)	(0.780)
1 Year Before	-0.219	-0.0235	0.0618	0.101	-0.147	0.0364
	(0.345)	(0.355)	(0.361)	(0.362)	(0.491)	(0.571)
Avg pretreated effect	-.242	.151	.313	.406	-.070	.416
p-value (avg effect = 0)	.3620	.5820	.3004	.1875	.8580	.4684
1 Year After	-0.351	-0.551	-0.616	-0.652	-0.393	-0.754
	(0.470)	(0.475)	(0.488)	(0.485)	(0.580)	(0.905)
2 Years After	-1.375***	-1.775***	-1.892***	-1.776***	-2.047***	-1.709**
	(0.263)	(0.282)	(0.316)	(0.339)	(0.505)	(0.620)
3 Years After	-1.157**	-1.756***	-1.959***	-1.880***	-2.178**	-1.468
	(0.448)	(0.457)	(0.510)	(0.526)	(0.774)	(0.921)
4 Years After	-0.844**	-1.637***	-1.925***	-1.885***	-1.751**	-1.914*
	(0.280)	(0.264)	(0.371)	(0.393)	(0.640)	(0.752)
Avg effect years 2-4	-1.125	-1.723	-1.925	-1.847	-1.992	-1.697
p-value (avg effect = 0)	.0001	.0000	.0000	.0000	.0006	.0141
Ratio of pre-post effect	4.65	11.41	6.15	4.55	28.46	4.09
County and year FE's	Y	Y	Y	Y	Y	Y
County linear trends	N	Y	Y	Y	Y	Y
Main controls	N	N	Y	Y	Y	Y
Policy controls	N	N	N	Y	Y	Y
Only Colorado	N	N	N	N	Y	N
Only Iowa/St. Louis	N	N	N	N	N	Y
Observations	15510	15510	15510	15510	12267	12348

Note: Standard errors in parentheses, clustered at the county level. This table displays estimates of the effect of LARC interventions on the rate of extremely preterm births per 1,000 live births. Column one estimates the standard two-way fixed effects (TWFE) specification. Column two adds county-specific linear trends. Column three add demographic and economic controls. Column 4 adds policy controls for whether emergency contraceptives were available over the counter and whether private insurance plans were required to cover any FDA-approved contraceptive. Columns five and six address concerns about staggered treatment timing by estimating the model separately based on when the intervention took place. Column five includes only Colorado as treated, while column six includes only St. Louis and Iowa. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.2: Event-Study Specifications Measuring the Effect of LARC Access on the Rate of Infant Mortality - 2003-2015

	(1) IMR	(2) IMR	(3) IMR	(4) IMR	(5) IMR	(6) IMR
3 Years Before	-0.982*	-0.323	-0.119	-0.0171	-0.984	0.0161
	(0.416)	(0.451)	(0.448)	(0.452)	(0.631)	(0.645)
2 Years Before	-0.587*	-0.147	0.0177	0.0944	-0.526	0.0529
	(0.284)	(0.316)	(0.319)	(0.321)	(0.413)	(0.568)
1 Year Before	-0.504	-0.284	-0.224	-0.203	-0.423	-0.400
	(0.355)	(0.381)	(0.379)	(0.380)	(0.546)	(0.531)
Avg pretreated effect	-.691	-.251	-.108	-.042	-.644	.110
p-value (avg effect = 0)	.0179	.4362	.7349	.8966	.1342	.8259
1 Year After	-0.547	-0.766	-0.846*	-0.870*	-0.606	-0.786
	(0.415)	(0.417)	(0.427)	(0.426)	(0.509)	(0.826)
2 Years After	-0.970**	-1.410***	-1.548***	-1.568***	-1.136	-1.566**
	(0.353)	(0.370)	(0.385)	(0.390)	(0.664)	(0.517)
3 Years After	-1.438***	-2.096***	-2.324***	-2.370***	-1.592**	-2.429***
	(0.397)	(0.360)	(0.395)	(0.405)	(0.584)	(0.699)
4 Years After	-1.079**	-1.956***	-2.244***	-2.324***	-1.433*	-2.182***
	(0.370)	(0.353)	(0.403)	(0.409)	(0.691)	(0.635)
Avg effect years 2-4	-1.162	-1.821	-2.039	-2.087	-1.387	-2.059
p-value (avg effect = 0)	.0003	.0000	.0000	.0000	.0131	.0004
Ratio of pre-post effect	1.68	7.25	18.88	49.69	2.15	18.72
County and year FE's	Y	Y	Y	Y	Y	Y
County linear trends	N	Y	Y	Y	Y	Y
Main controls	N	N	Y	Y	Y	Y
Policy controls	N	N	N	Y	Y	Y
Only Colorado	N	N	N	N	Y	N
Only Iowa/St. Louis	N	N	N	N	N	Y
Observations	15554	15554	15544	15544	12293	12374

Note: Standard errors in parentheses, clustered at the county level. This table displays estimates of the effect of LARC interventions on the number of infant deaths per 1,000 live births. Column one estimates the standard two-way fixed effects (TWFE) specification. Column two adds county-specific linear trends. Column three add demographic and economic controls. Column 4 adds policy controls for whether emergency contraceptives were available over the counter and whether private insurance plans were required to cover any FDA-approved contraceptive. Columns five and six address concerns about staggered treatment timing by estimating the model separately based on when the intervention took place. Column five includes only Colorado as treated, while column six includes only St. Louis and Iowa. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3: Comparison of Gestational Ages for Extremely Preterm Births in Colorado: Before and After Colorado Family Planning Initiative

Gestational Age	(1) 2006-08	(2) 2010-14	(3) P-value
<24 Weeks	.326	.325	.589
24 Weeks	.135	.133	.092
25 Weeks	.151	.154	.061
26 Weeks	.179	.180	.610
27 Weeks	.208	.207	.916

Note: This table compares the proportion of extremely preterm births in Colorado which fall under each gestational age category, both before and after the Colorado Family Planning Initiative. Column 1 displays the proportion of EPBs in 2006-2008 (pre-intervention) which were in each gestational age category. Column 2 displays the proportion of EPBs in 2010-2014 (post-intervention) in each gestational age category. Column 3 displays a p-value on a test for equality of the proportions before and after the Colorado Family Planning Initiative.

Table 1.4: Annual Cost Calculation for Extremely Preterm Birth Reduction for Infants who Would be Predicted to Survive

	(1)	(2)	(3)	(4)	(5)	(6)
Gest. Age	Pred. N	Surv. Odds	N Survived	P(Vent|A,S)	(Cost|A,S)	Pred. Costs Avoided
<24 Wks	27.6	.091	2.5	.94	205,000	481,750
24 Wks	11.2	.505	5.6	.91	200,000	1,019,200
25 Wks	11.7	.883	10.4	.74	130,000	1,000,480
26 Wks	15.9	.883	14.0	.74	130,000	1,346,800
27 Wks	18.6	.883	16.4	.74	130,000	1,577,680
Total	85.0		48.9			5,425,910

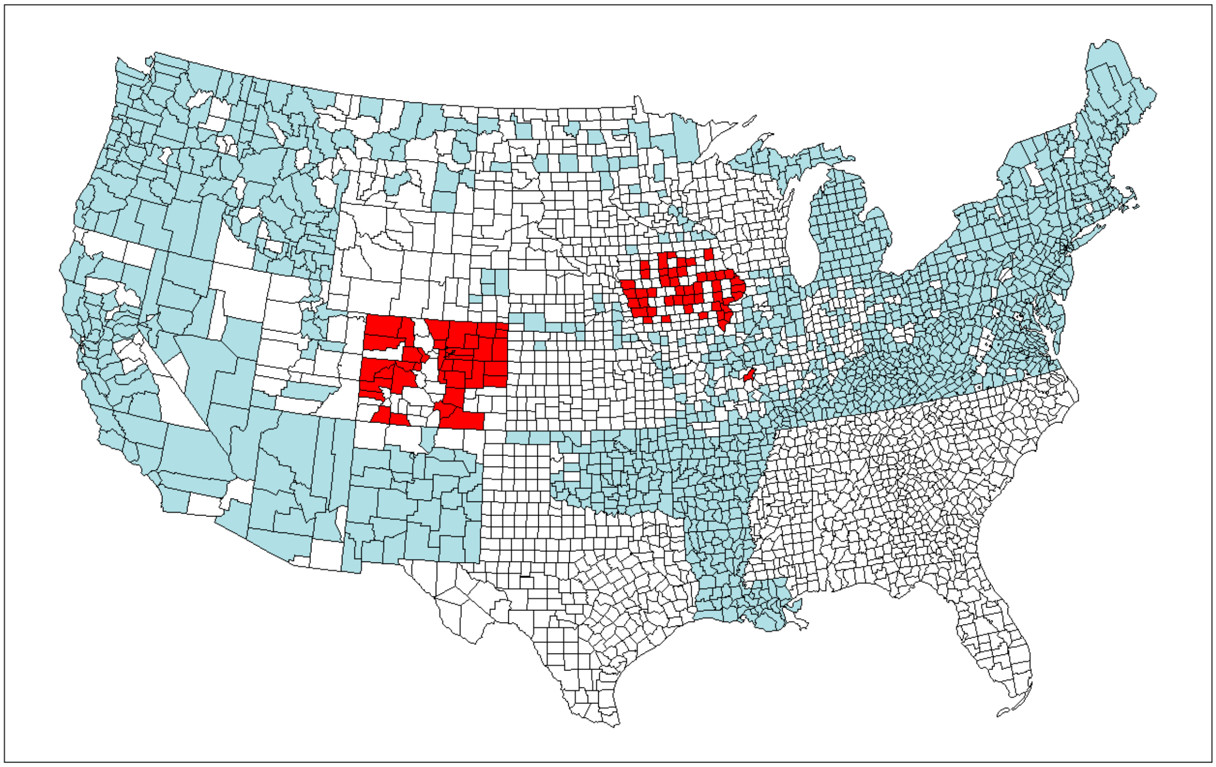
Note: This table calculates the predicted annual cost savings due to avoided ventilation care among infants who would have been predicted to survive. Column 1 distributes the estimate of a reduction of 85 extremely preterm births across gestational age categories, based on the proportional occurrence of each age in Colorado from 2006-2013. Column 2 displays the likelihood of survival for a birth of that gestational age, taken from [169]. Column 3 calculates the number of births at each gestational age which would be predicted to survive, based on the likelihood in column 2. Columns 4 and 5 display the probability of being ventilated and the average cost of ventilation, conditional on gestational age and survival, respectively. Finally, column 6 calculates the total predicted avoided ventilation costs from surviving infants for each gestational age group.

Table 1.5: Annual Cost Calculation for Extremely Preterm Birth Reduction for Infants who Would be Predicted Not to Survive

	(1)	(2)	(3)	(4)	(5)	(6)
Gest. Age	Pred. N	NS Odds	N NS	P(Vent|A,NS)	(Cost|A,NS)	Pred. Costs Avoided
<24 Wks	27.6	.919	25.1	.18	10,000	45,180
24 Wks	11.2	.495	5.6	.74	15,000	62,160
25 Wks	11.7	.117	1.4	.77	20,000	21,560
26 Wks	15.9	.117	1.9	.77	20,000	29,260
27 Wks	18.6	.117	2.2	.77	20,000	33,880
Sum	85.0		36.2			192,040

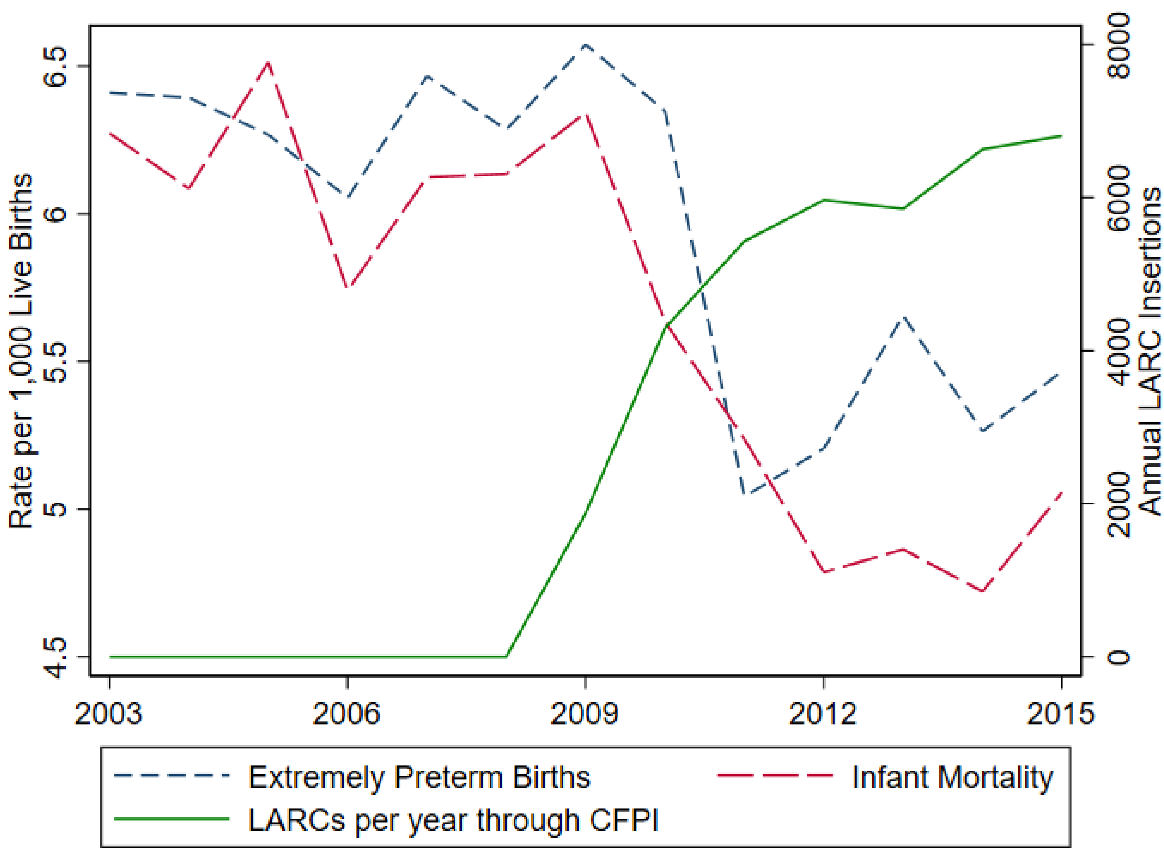
Note: This table calculates the predicted annual cost savings due to avoided ventilation care among infants who would have been predicted not to survive. Column 1 distributes the estimate of a reduction of 85 extremely preterm births across gestational age categories, based on the proportional occurrence of each age in Colorado from 2006-2013. Column 2 displays the likelihood of death for a birth of that gestational age, taken from [169]. Column 3 calculates the number of births at each gestational age which would be predicted not to survive, based on the likelihood in column 2. Columns 4 and 5 display the probability of being ventilated and the average cost of ventilation, conditional on gestational age and non-survival, respectively. Finally, column 6 calculates the total predicted avoided ventilation costs from non-surviving infants for each gestational age group.

Figure 1.1: Treatment and Control Assignment



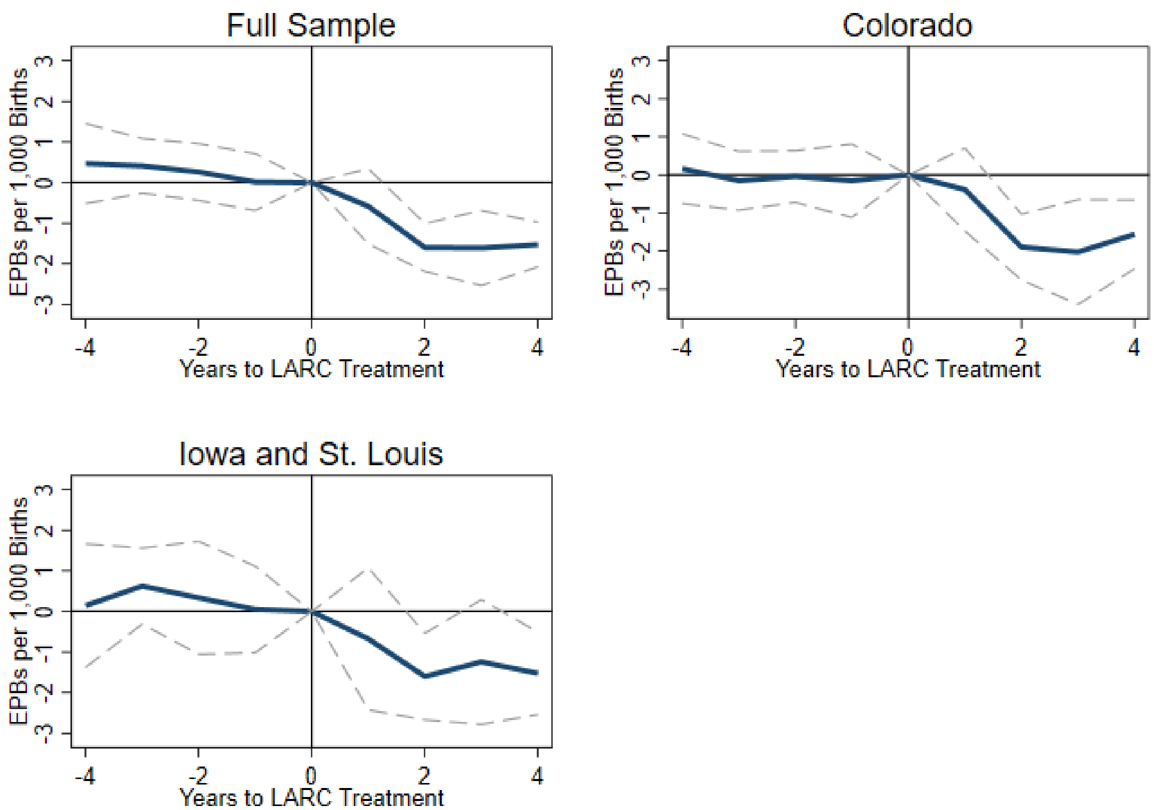
Note: This figure displays (treated) counties which have a Title X family planning clinic and which received specific funding to expand LARC access to low-income women in red. Control counties, which are other US counties which have a Title X clinic but which did not received specific funding to expand LARC access are denoted in blue. Counties which do not have a Title X clinic, or which are located in a state which did not expand Medicaid with the Affordable Care Act of 2010 are omitted from all subsequent regressions, as are all counties which border a treated county for concerns about potential spillovers which would violate the Stable Unit Treatment Value Assumption (SUTVA).

Figure 1.2: CFPI and Infant Health - 2003-2015



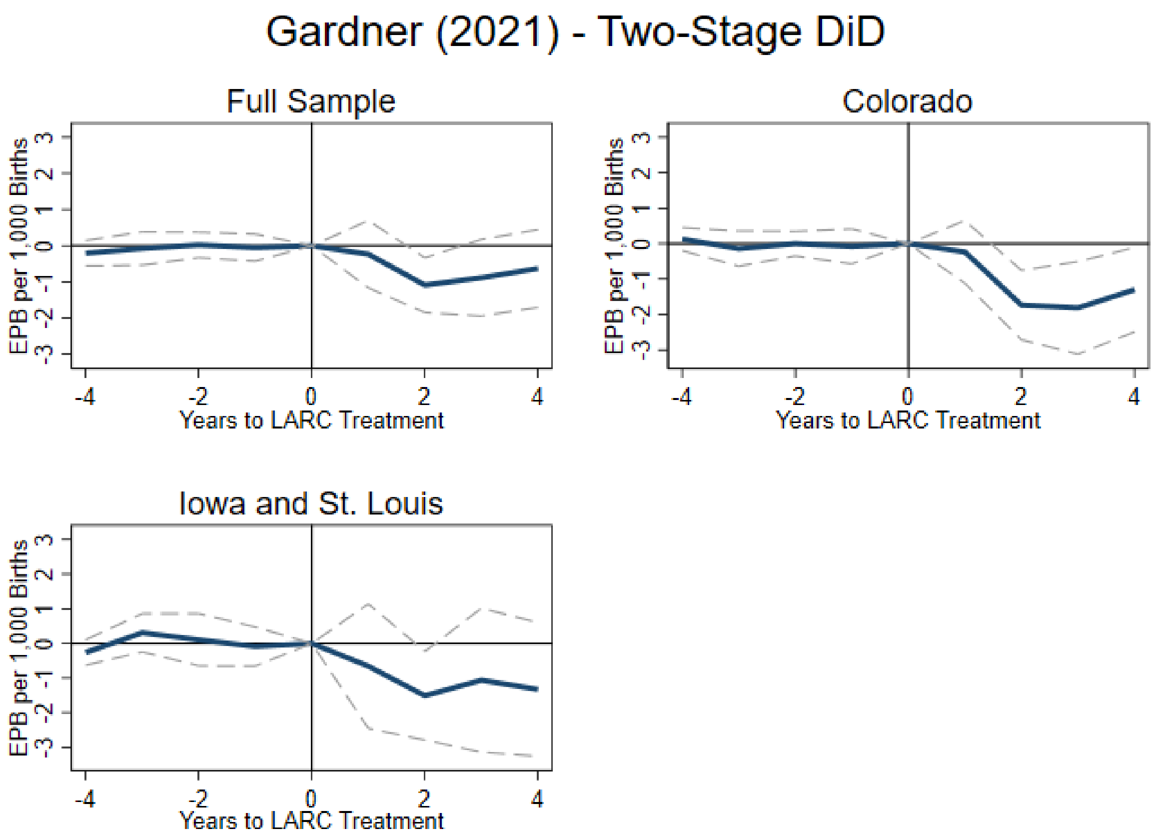
Note: This figure displays the annual number of LARCs inserted through the Colorado Family Planning Initiative compared with the rates of extremely preterm births (births before 28 weeks gestation) as well as the infant mortality rate in Title X counties in Colorado, both calculated using restricted-access data from the National Vital Statistics System.

Figure 1.3: Event-Study Graphs - the Effect of LARC Access on the rate of Extremely Preterm Births - 2003-2015



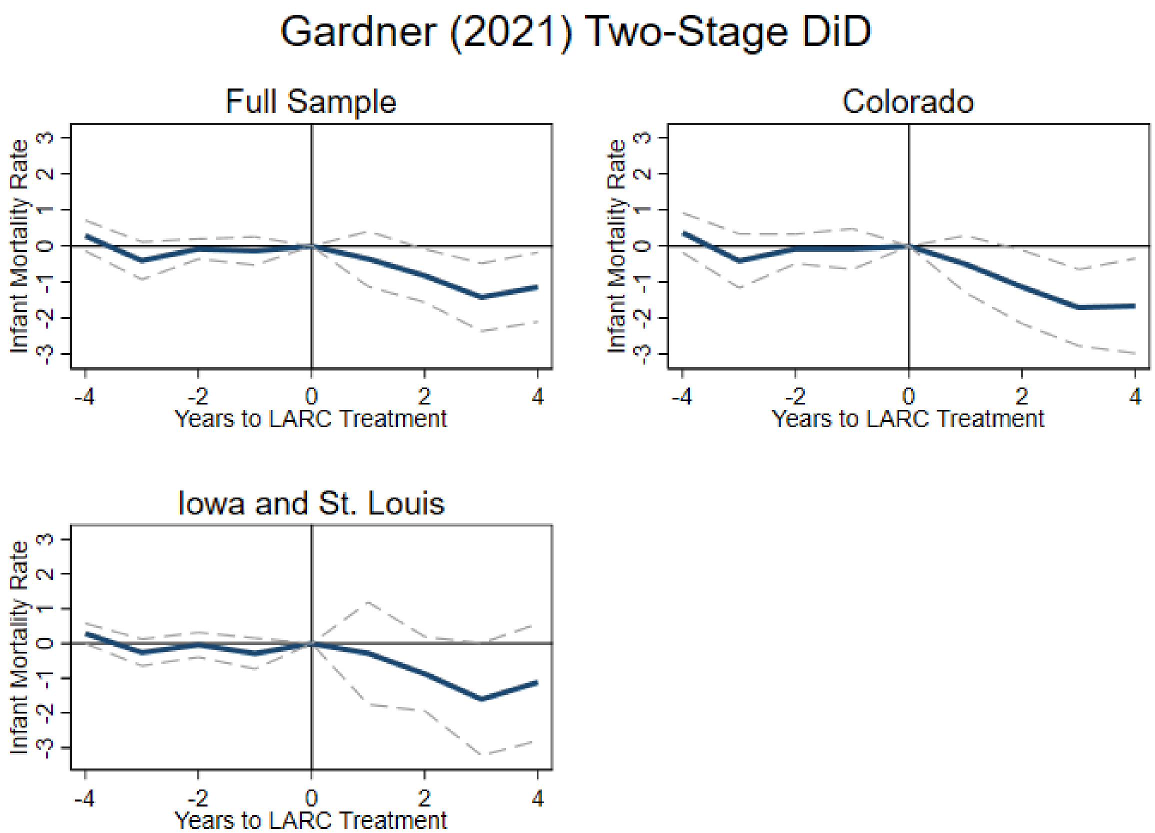
Note: This figure uses natality data from the National Vital Statistics System to plot coefficients from the event-study specification comparing preterm birth rates in the three treated areas compared with other counties with Title X clinics which are in states that expanded medicaid and do not border treated counties. The top left graph displays the fully specified version including all three treated areas, the top right graph displays the results for only Colorado, while the bottom graph displays the results for only Iowa and St. Louis.

Figure 1.4: Event-Study Graphs Using Two Stage DiD (Gardner (2021)) to Estimate the Effect of LARC Access on the Rate of Extremely Preterm Births



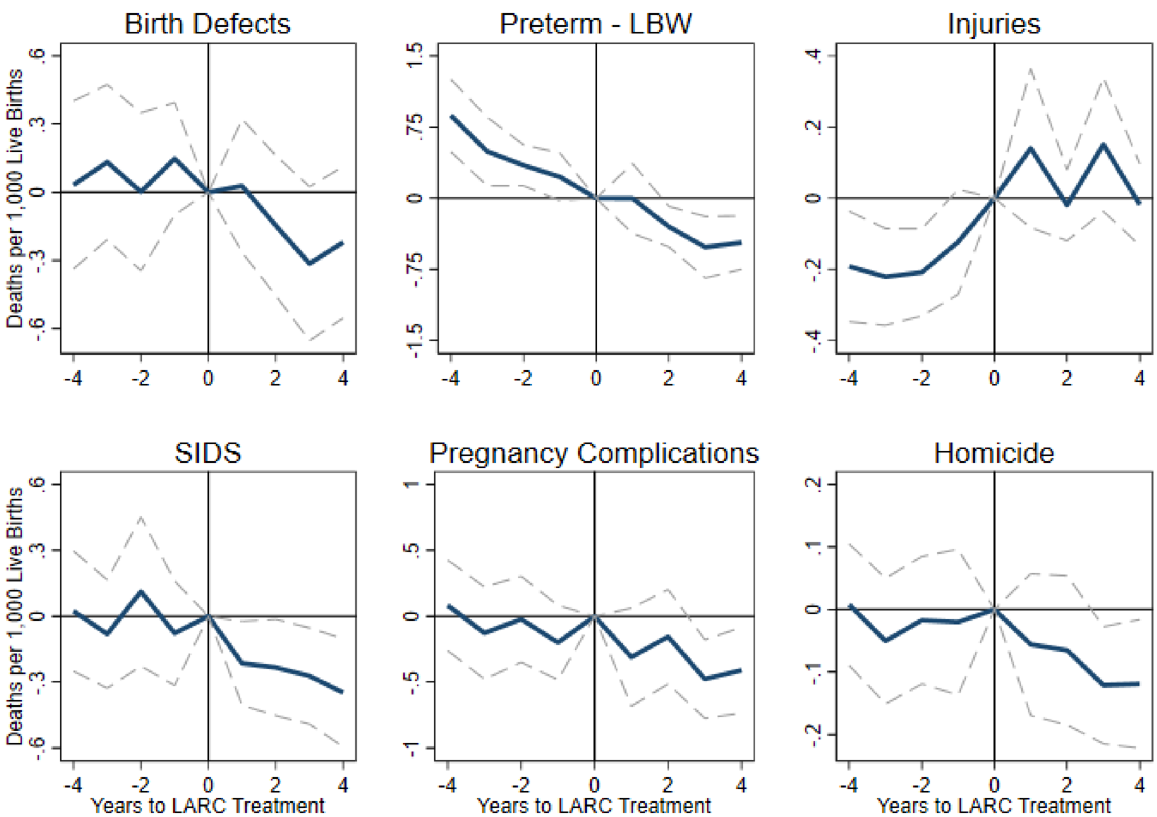
Note: This figure uses natality data from the National Vital Statistics System to plot coefficients from the event-study specification utilizing the two-stage difference-in-difference method of [149], comparing preterm birth rates in the three treated areas compared with other counties with Title X clinics which are in states that expanded medicaid and do not border treated counties. The top left graph displays the fully specified version including all three treated areas, the top right graph displays the results for only Colorado, while the bottom graph displays the results for only Iowa and St. Louis.

Figure 1.5: Event-Study Graphs Using Two Stage DiD (Gardner (2021)) to Estimate the Effect of LARC Access on the Rate of Infant Mortality



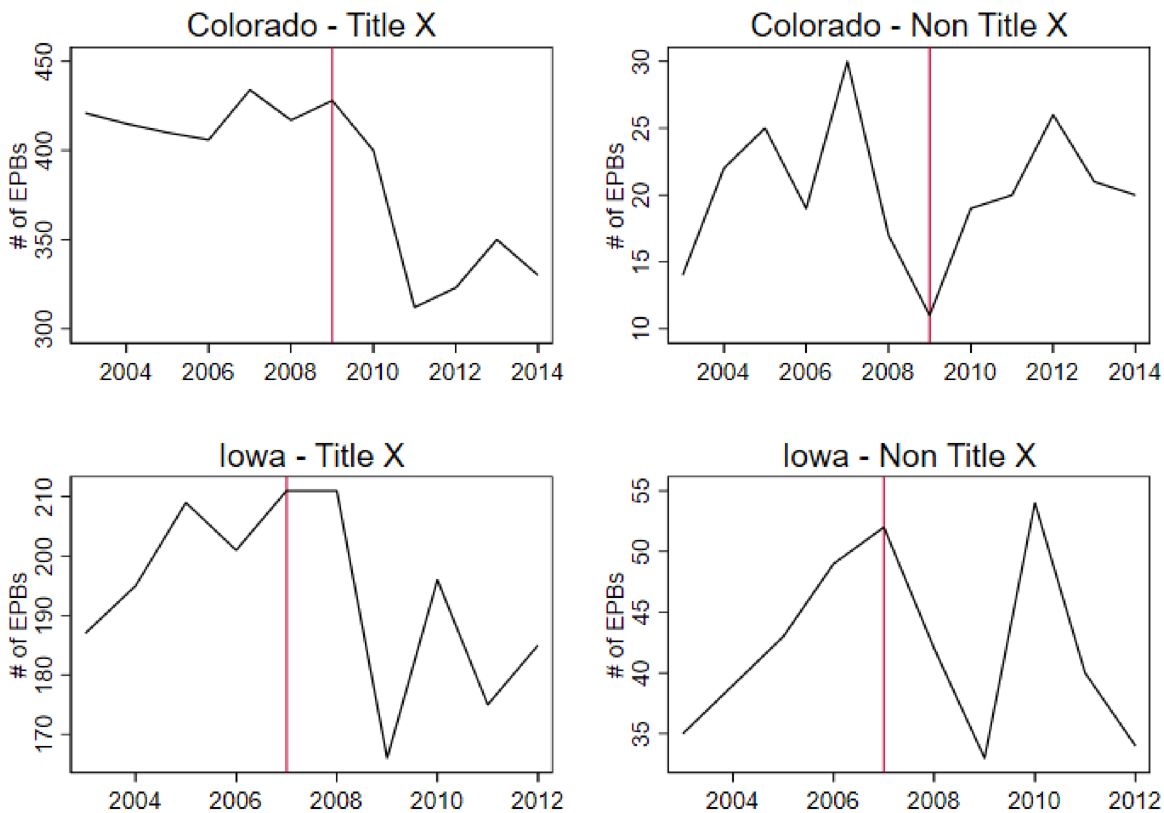
Note: This figure uses natality data from the National Vital Statistics System to plot coefficients from the event-study specification comparing infant mortality rates in the three treated areas compared with other counties with Title X clinics which are in states that expanded medicaid and do not border treated counties using [149]'s two-stage difference-in-difference estimator. The top left graph displays the fully specified version including all three treated areas, the top right graph displays the results for only Colorado, while the bottom graph displays the results for only Iowa and St. Louis.

Figure 1.6: Event-Study Graphs on the Effect of LARC Access on the Rates of Individual Causes of Infant Mortality



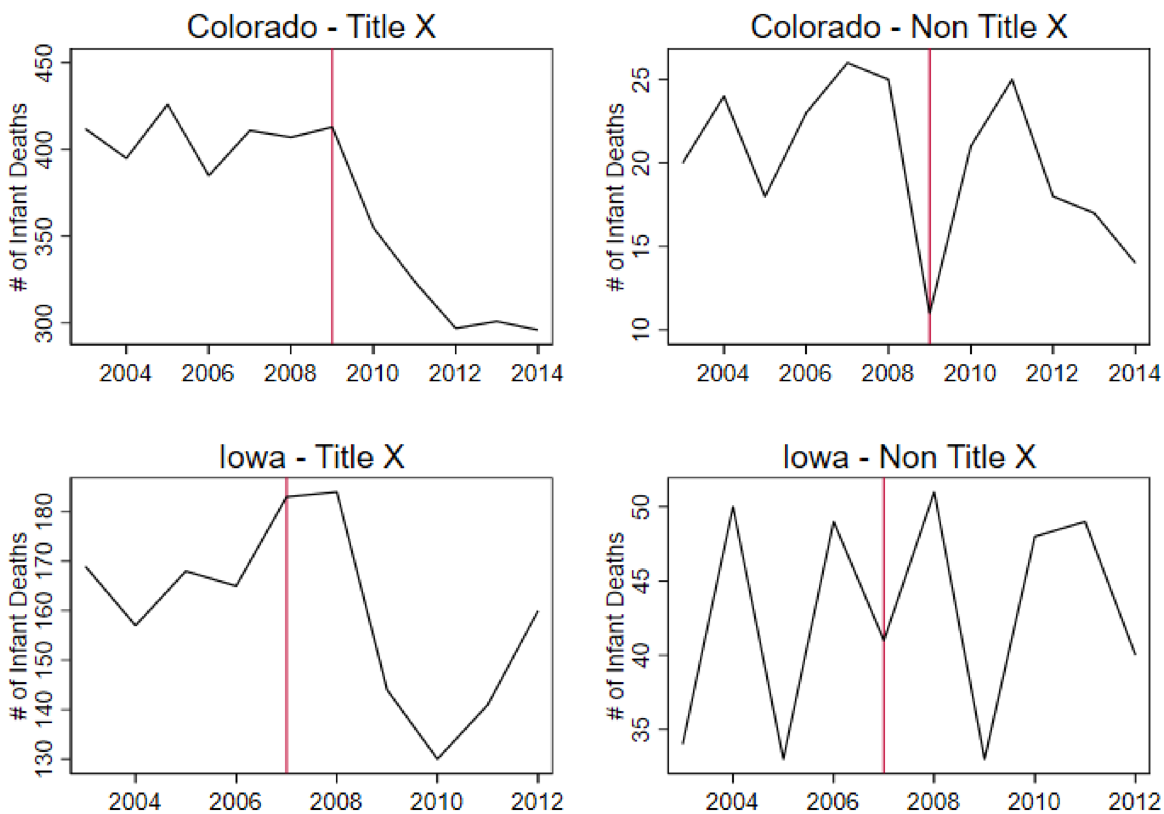
Note: This figure uses natality data from the National Vital Statistics System to plot coefficients from event-study specification for each of the six most common causes of infant death

Figure 1.7: Time Series of Extremely Preterm Birth Outcomes in Colorado and Iowa by Title X Status



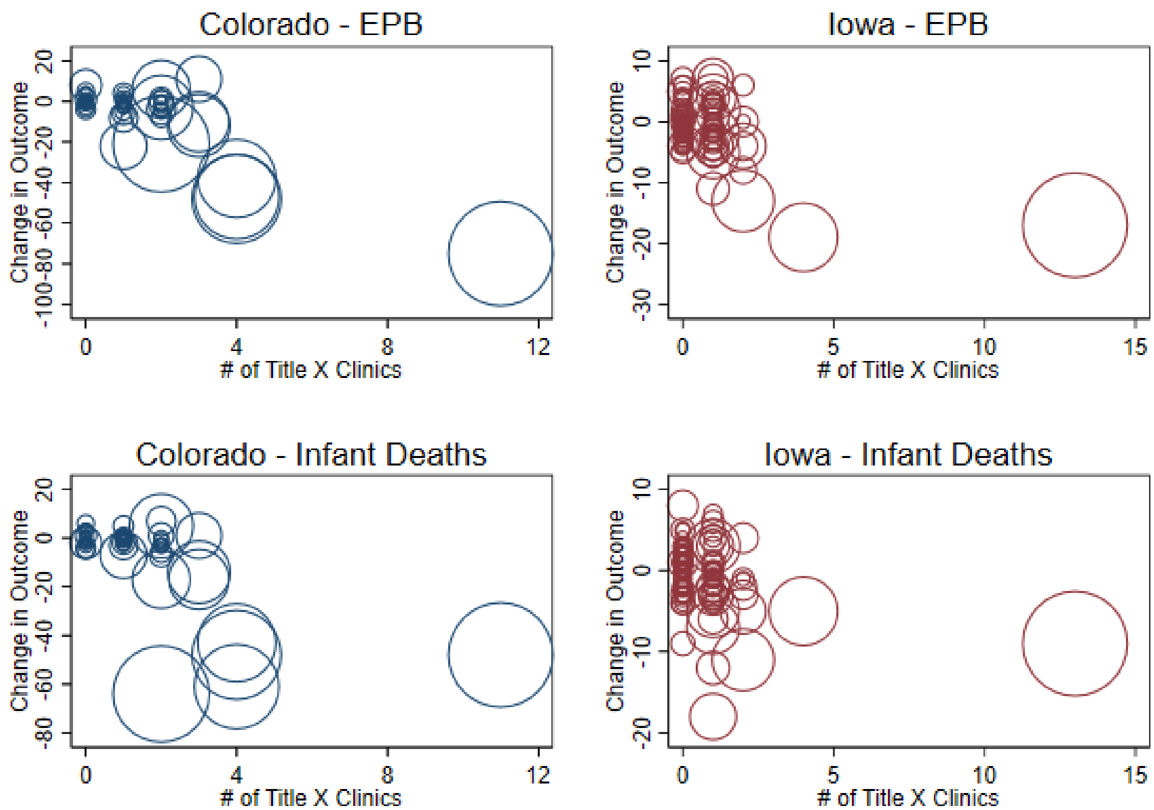
Note: This figure displays the raw number of extremely preterm birth cases in Colorado and Iowa counties with and without a Title X clinic. Graphs on the left display the outcome for Title X counties, while graphs on the right display the non-Title X counties. The top row displays outcomes for Colorado, while the bottom row displays outcomes for Iowa, using data from the National Vital Statistics System.

Figure 1.8: Time Series of Infant Mortality Outcomes in Colorado and Iowa by Title X Status



Note: This figure displays the raw number of infant mortality cases in Colorado and Iowa counties with and without a Title X clinic. Graphs on the left display the outcome for Title X counties, while graphs on the right display the non-Title X counties. The top row displays outcomes for Colorado, while the bottom row displays outcomes for Iowa, using data from the National Vital Statistics System.

Figure 1.9: Change in Infant Health Outcomes from Pre to Post LARC Initiative by of TitleX Clinics - Iowa and Colorado



Note: This figure displays the change in the average number of EPB's and infant deaths from the four years leading up to a LARC intervention to the four years after the intervention for counties in Colorado and Iowa compared with the number of Title X clinics in that county. The circles for each county are weighted by the average number of births in the county across all sample years. Blue circles represent Colorado counties, while red circles represent Iowa counties.

Chapter 2

Salary Disclosure and Invidual Effort: Evidence from the National Hockey League

2.1 Introduction

Meritocratic compensation is an ideal for any market economy, with high performers reaping the rewards of their oversized relative contributions. An individual's salary therefore says a great deal about how their firm values their work and it can become an important part of one's identity. Research has shown that as long as pay discrepancies can be traced to differences in performance, workers will accept them, but when employees of similar calibre are paid differently, problems can arise ([57], [55]). Because of this, Economists have long studied the ways in which individuals care not just about their own salary, but how their salary compares to their relevant peer group. Several papers have shown that job satisfaction is inversely related to an individual's comparison wage ([98], [74], [150], [281]). [217] finds that, controlling for a person's income, their happiness is decreasing in their neighbor's income, while [290] find that people would be willing to accept 50% less real income in order to have higher relative income. [220] finds that low relative status leads to an increased risk for disease and early death, while [143] find neurophysiological evidence that social comparison has a direct effect on the reward processing centers of the brain. Clearly, information about relative wages has important implications for individual health and happiness. The effect of comparison income on worker effort has been studied in laboratory settings[1] , but to my knowledge no current study uses real world data to answer the question of what happens to

[1] [89], [14], [10], [11], [76], [41], [101]

individual effort within a job and overall firm performance when workers suddenly learn their place in the salary hierarchy.

In this paper, I do precisely that by exploiting a natural experiment in the National Hockey League. On January 29, 1990, almost two thirds of the way through the grueling seven month long NHL regular season, the Montreal Gazette published the salaries of every player in the league, ending a decades-long culture of pay secrecy in the NHL. Overnight, players went from having little idea where they stood in the leaguewide salary distribution to knowing precisely where they fit in. Because both salary and statistical data is available for the mostly untreated 1990-91 season, I am able to classify players as underpaid or overpaid by comparing their salaries to those of similarly performing players. By then tracking the performance of players in response to this new information, I find that underpaid players began shifting their effort from defensive production to offense, which is more highly rewarded in the NHL labor market. Their goals went up significantly, assists rose slightly, but their defensive production declined, with these players allowing their teams to get scored on by even more than the additional goals they provided. Crucially, teams who gave significant raises to underpaid players after salary disclosure were able to avoid this outcome.

Asymmetrically, I find no evidence of a behavioral response for players who were overpaid prior to salary disclosure. While overpaid players appear to continue to allocate their effort to maximize firm performance, underpaid individuals appear to be playing in a manner designed to maximize the value of their next contract. Though underpowered, I find that this behavioral shift is larger for underpaid players who play for teams with higher than average payrolls. This is in line with reference-dependent utility theory, which would suggest that underpaid players on wealthier teams would be more surprised to learn that they are underpaid, and thus would have the largest discrepancies between their actual salary and their reference point.

At the firm level, I find that in the season leading up to salary disclosure, overall team payroll had no correlation with team performance. In the season following disclosure, however, team payroll and performance became immediately and permanently linked, with higher-paying teams consistently outperforming lower-paying ones. Because the restrictive rules of free agency in the early 90's effectively prevented elite players from switching teams by choice, this result can only be explained by the higher-paying teams getting relatively more production out of their players following disclosure, as opposed to a sorting explanation where the better players are reallocated to the higher-paying teams.

This paper contributes to several strands of the literature. There exists a large experimental literature which investigates the link between pay inequality and effort provision within a job, though experimental evidence has been mixed. While both [89] and [41] fail to find evidence that pay inequality has a material impact on worker effort, [76] find that it can affect the quality of output produced, suggesting that there are internal margins on which effort may be withheld in response to perceived unfairness in pay. My results build on these findings by demonstrating that workers can respond to perceived unfairness by changing the type of effort they exert, as opposed to withholding effort entirely.

A pair of important papers, [101] and [100], document an asymmetric response to both reductions and increases in relative wages. [101] finds that when a group member's wage is arbitrarily cut, that member reduces effort while the other member, whose relative wage has risen, displays no effort response. Conversely, [100] show that underpaid workers will respond to a raise with increased effort while overpaid workers will not, suggesting that the removal of perceived unfairness is the key mechanism. I document a similar asymmetric response whereby underpaid workers reallocate their effort from defense to offense, but underpaid workers who receive a raise which removes the perceived unfairness appear to reallocate their effort in the opposite direction. Also, it is important that my findings complement some of these experimental studies. Experiments have many important benefits in their ability to randomize participants and isolate causal impacts, but it is difficult for them to imitate the conditions of an actual job, where workers show up and interact with the same people over months and years. By using a natural experiment where all workers in a labor market learned whether they were over or underpaid and then tracking their performance, I am able to provide evidence that many of these important results may generalize outside of their experimental settings.

I also contribute to the literature which shows how differences in individual versus firm goals can create misincentives for agents. This includes [51], which find that optimal CEO compensation contracts may emphasize short-term stock performance at the expense of long-run value, and [228], who find that higher risk firms perform worse in the long run when CEO compensation is highly sensitive to stock returns. I build upon these findings by showing that underpaid NHL players appear to reallocate their effort to optimize their next contract to the detriment of team performance. This result should be of interest to any organization which employs individuals to perform several different tasks, where one of the tasks is more noticeable and easy to track, though all of the tasks are vital to the organization's goals. When a worker feels undervalued, they might rationally respond by focusing more of their effort on the more noticeable task, letting their performance on other dimensions suffer. Examples of this might include sales associates focusing on maximizing short-term sales numbers instead of building long-term relationships with clients, or public school teachers 'teaching to the test' instead of working to instill a love of learning in their students.

Third, I contribute to the literature which analyzes the potential impact of public policies which promote salary transparency by either requiring disclosure from certain public employees or by outlawing firms from preventing employees from sharing their salary information with their peers. Such laws have already been adopted by 10 states and are being considered by several others. [201] and [38] both demonstrate that laws which promote pay transparency can reduce the gender pay gap, while [241] argue they have the potential to reduce pay inequity and inequality and [224] finds they compressed the top salaries of public-sector managers in California. My results suggest that while such laws may have a detrimental impact on the bottom line of firms which underpay their workers, they can actually be beneficial to higher paying firms.

I also contribute to the literature which studies the importance of reference-dependence in worker decision-making. While experimental results on reference-dependent utility have been mixed[2] , applied work has demonstrated the importance of reference points in the labor supply de-

[2] While [7], [133], [127], and [293] find experimental evidence consistent with reference-dependent preferences, [170], [289], and [88] all fail to do so.

cisions of cab drivers ([106]) and stadium vendors ([243]), the performance of police officers ([223]), and the relationship between upset losses for NFL and NBA teams and domestic violence ([73], [75]). Though some of my results are underpowered, I contribute to this literature by showing that underpaid NHL players who play for higher-paying teams, who are thus more likely to be surprised to learn they are being underpaid, change their behavior more sharply than underpaid players on lower-paying teams.

The remainder of the paper proceeds as follows. Section two provides background on the game of hockey and how it is played, the abrupt end to the culture of pay secrecy that existed in the NHL prior to 1990 and the relevant performance statistics that I use to assess the behavioral response in the wake of salary disclosure, as well as providing a conceptual framework using the relevant research related to this topic. In section three, I describe my data and present descriptive findings of what occurred in the NHL labor market in the years surrounding salary disclosure. In section four, I track the performance of overpaid and underpaid players from the pre-disclosure era into the season following disclosure in order to understand how this information impacted player performance and team performance. Section five concludes.

2.2 Background

2.2.1 What is Hockey?

The game of hockey is played with two opposing teams, each consisting of five skaters and a goaltender who are on the ice at any given time. The game is played with a small rubber disc known as the ‘puck’, with the object of the game being to get the puck into the opposing team’s net more times than they get it into yours. The standard NHL rink is 200 feet long and 85 feet wide, is rounded in the corners and surrounded by a wooden or fiberglass fence of between 40-48 inches high (known as the ‘boards’), with safety glass extending a further eight feet above the boards. The game is played for 60 minutes (not including stoppages), broken up into three 20-minute ‘periods’, with ‘intermissions’ occurring between periods where the ice is cleaned.

Because of the fast-paced nature of the game, substitutions occur continuously throughout the game and are allowed 'on the fly', meaning that teams do not have to wait for a stoppage in play to change who is on the ice. A typical 'shift' for a player usually lasts between 30-90 seconds of gametime before they are replaced. Of the five skaters, there are two defensemen and three forwards. The two defensive positions are somewhat interchangeable, but the three forward positions are the left wing, the center and the right wing. That said, all players play both offense and defense, with defenseman regularly scoring goals and forwards often blocking shots from the opposing team. For the 1989-90 season, the average defenseman scored five goals and had 18 assists, while the average forward had 17 goals and 23 assists. In the late 1980s and early 1990s there were, on average, between 7-7.5 total goals scored per game, compared with about 2.5 goals per game across European football leagues (known as 'soccer' in the United States and Canada)[3] .

A typical 'lineup' for a team includes four 'lines' of forwards, or four pairings of a left wing, center and right wing who will most often play together, as well as three pairings of defensemen. Though they typically play together, through the course of a game there are many situations in which the lines are 'shuffled' or changed, so that in any given game a player will play with many more players than just their usual line. The lines are also changed frequently throughout the course of a season because of injuries, form or other strategic considerations. The National Hockey League is considered to be the premier professional hockey league in the world ([221]). Despite the name, the league includes teams from both the United States and Canada. The league began in 1917 with four teams, all of which were Canadian, before expanding to include American teams in 1924. Because of the league's success, it expanded rapidly in the 1960's and 70's to the point where it had 21 teams in 1979. This remained the case until the 1991-92 season, where the San Jose Sharks were added to the league.

[3] For 1989-90, there were 2.6 goals per game in the German Bundesliga and the English First Division (now known as the Premier League), 2.2 goals per game in the Italian Serie A, and 2.4 goals per game in Spain's La Liga

2.2.2 Salary Disclosure in the NHL

While there had been talk of salary disclosure among some circles within the NHL, the idea had been resisted by the NHLPA's long-serving executive director, Alan Eagleson. In the late 80's, however, unrest within the NHLPA over Eagleson's leadership came to a head, as players felt the NHL was being left behind as the other three major American sports leagues (the National Basketball Association, the National Football League and Major League Baseball) had all experienced rapidly rising salaries in the 1980s. In the summer of 1989 Eagleson was forced to hire an assistant who would take over as executive director in 1992 ([125]). Eventually, beginning in September of 1991, a journalist named Russ Conway would publish a series of articles documenting Eagleson's extensive corruption as the leader of the NHLPA, including accusations that he embezzled money from the players' pension funds and from disability claims made by former players. Conway's investigative journalism opened the door for a series of criminal investigations and earned him a Pulitzer prize nomination, while Eagleson would eventually plead guilty to six counts of fraud and embezzlement ([103]).

The assistant who would apprentice under Eagleson was an agent named Bob Goodenow, who was in favor of salary disclosure as a way to make the arbitration process in hockey more fair. While Goodenow would have to wait three years before taking power, he was able to get the NHLPA to take a vote of its members on the issue of salary disclosure, which was passed by an overwhelming 469-49 majority on November 25, 1989 ([9]). Salary numbers were scheduled to be made public the following February, but on January 29, 1990, the Montreal Gazette acquired and published the list prematurely, to the chagrin of several owners who complained that the list did not include renegotiations they had made since the offseason ([132]). The following day, several other papers, including the Washington Post, the Los Angeles Times and the Chicago Tribune, confirmed and reported on the figures ([131], [198], [294]).

The biggest way in which salary disclosure impacted the NHL labor market was by making the arbitration process more fair. Arbitration relies heavily on bench-marking player salaries to those of ‘comparables’ who have similar statistical output ([40]). Prior to salary disclosure, the league was able to select favorable comparisons which would make the player in question appear to be worth as little as possible. The player and their agent, without access to the full set of NHL salaries, had little in the way of recourse until salaries were disclosed ([37], [265]). Once this happened, they were able to present a counterargument to the one presented by the league in order to argue for a more realistic ‘fair market value’ of the player. This also made negotiations more favorable for players who would not eventually go through arbitration because the threat of doing so became more credible.

Before this, a decades-long culture of pay secrecy had persisted within the league, and many famous players have described the difficulty in getting other players to share information on what they were making. Future Hall of Famer Mike Gartner said “Believe it or not, we still had a lot of players that didn’t want to disclose their salaries. They didn’t think it was a good idea, they didn’t see the bigger picture that if everybody knows what everybody is making, we’re all going to make more.” According to Goodenow, “There was one player, he was very vociferous as being against salary disclosure. He said he did not want to back out of his driveway and have his neighbors look at him and have them know what his salary was. About a year later, he sure liked salary disclosure as he was preparing for his salary arbitration case.” Though anecdotal, this captures the general feeling many players had at the time.

After salaries were disclosed, future Hall of Fame defenseman Ray Bourque said “you always felt uncomfortable going up to a guy and asking ‘Hey, how much are you making?’ This way all you have to do is peek at the list.” At the time he said this, Bourque was already an 11-time NHL All-Star and three-time Norris Trophy winner as the league’s best defenseman. If a player of his stature felt uncomfortable asking other players what they were making, less influential players likely did as well. In Appendix B, I give a more complete description of labor relations in the NHL around this time and outline the restrictive rules of free agency which prevented star players from switching teams via free agency.

2.2.3 Key Statistics

The most straightforward statistic to understand for the purposes of this paper is goals scored. Any time a goal is scored in hockey, credit goes to the last player who touched the puck for the scoring team, regardless of whether or not they shot it directly in. For any goal scored, zero, one or two assists will be credited to members of the scoring team. A ‘primary’ assist is awarded to the player who passed the puck to the ultimate goal scorer, while a ‘secondary’ assist goes to the player who passed the puck to the player receiving the primary assist. If the goalscorer won the puck directly from the opposition before scoring, then no assists are awarded. Both primary and secondary assists count as one assist, so no distinction is made between the two for statistical purposes. Shots on goals are any attempt on goal by an opposing player that requires a save from the goaltender. This means that shots blocked by a defender and shots that miss the net are not counted. The term ‘points’ can be slightly confusing in hockey because it has a different meaning when it is associated with an individual player or with a team. For a player, points are simply the sum of their goals and assists. For a team, points determine their place in the standings. A team gets two points from each victory and one point for a tie[4] . Team points are summed over the course of a season to determine who makes the playoffs and where teams are seeded once the playoffs start.

Another important statistic is plus-minus. A player is awarded a ‘plus’ each time she is on the ice when her team scores an even-strength or shorthanded goal[5] . She receives a ‘minus’ if she is on the ice for an even-strength or shorthanded goal scored by the opposing team. Over the course of a season, these pluses and minuses are summed up to give the player an overall plus-minus rating. More than just being an offensive statistic like the previous three, plus-minus serves as a proxy

[4] During the period this paper examines, if a game was tied at the end of regulation, the teams would be a five minute ‘sudden death’ overtime. If a goal was scored during overtime, the game would end and the scoring team was declared the winner. If no goals were scored in overtime, the game ended in a tie and both teams received a point in the standings

[5] Because the NHL has ‘power-plays’, where one team has to play with one fewer player as punishment for a rule infraction, even-strength situations are ones where both teams have the same number of players on the ice, while shorthanded situations are where a team is playing with fewer players. This means that goals scored on the power-play do not affect plus-minus, unless they are scored by the shorthanded team.

for a player's overall contribution to the team's performance. Since players are constantly subbing in and out of the game and being paired up with different teammates, comparing the plus-minus of players within a team reveals which players have a more positive impact on team performance throughout the season. For the purposes of this paper, there are three important things to note about plus-minus as a statistic. The first is that plus-minus contains more noise than the offensive statistics mentioned above, as players will often be on the ice when the team gets scored on through no fault of their own, and likewise their team will score goals they played no part in.

The second important detail is that despite the noise, plus-minus still measures something real and valuable. Both the noisiness and the value of plus-minus as a proxy for overall team contribution are displayed in Appendix Figures B.1 and B.2, which show the plus-minus of the winner of the Selke and Norris Trophies (given to the league's best defensive forward and best defenseman, respectively), compared to the means for their position. Though the plus-minus of the award winners bounces around quite a bit, it is virtually always positive and often among the highest in the league. That these players consistently have much higher plus-minuses than other players suggests that the statistic is tracking something important. In recent years, more advanced statistics that incorporate data on shots on goal as well as goals scored have been introduced to replace plus-minus, allowing for a bigger sample size over the course of a season. While these new statistics undoubtedly improve upon plus-minus, they are also highly correlated with plus-minus. Appendix Figure B.3 displays scatterplots for plus-minus versus three of these advanced statistics for 2019, Corsi For %, Fenwick For % and Expected Plus-Minus, showing a strong positive associaiton with each[6] .

While plus-minus can be misleading when looking at individual players, looking at trends in plus-minus across larger groups still tells a meaningful story. Finally, despite this, improvements

[6] Corsi For % tracks the percentage of total shot attempts which were made by a player's team while that player was on the ice, regardless of whether they were saved, blocked, missed the net, or scored. Because of the much larger sample size of shots attempts compared with goals, this greatly reduces the noise that occurs in plus-minus. Fenwick For % is similar to Corsi For % but excludes blocked shots as its creator, Matthew Fenwick, argues that blocked shots are not as clear a scoring opportunity as unblocked shots ([67]). Finally, Expected Plus-Minus uses the leaguewide shooting percentage from each location on the ice to assess how likely each shot attempt was to result in a goal. Expected Plus-Minus is the sum of the expected value of all shot attempts, regardless of whether or not they resulted in a goal.

in plus-minus go virtually unrewarded in salary negotiations. Appendix Figure B.4 shows the coefficients from a hedonic regression that predicts salary based on statistical performance[7] for each year from 1990-2003. While points are always a positive and significant predictor of salary, plus-minus is only positive and significant in 2000 and 2002, and is even negative in five of the 14 years displayed. Since offensive production is controlled for in these regressions, this means that holding a players offensive production constant, improvements in defense are not compensated accordingly.

Taken together, the fact that plus-minus is noisy and that it is not well compensated suggest that an underpaid player wishing to maximize their next contract could shade their effort on defense and focus more on offense. This would lead to more goals and assists, which are highly valued in the labor market, and a lower plus-minus, which a player could attribute to noise in salary negotiations and which is not generally punished in the labor market regardless.

2.2.4 Conceptual Framework

Multiple theories contribute to how we might expect individual players to respond to salary disclosure. [17]'s fair-wage effort hypothesis would predict that upon learning they were being paid less than other similar players, underpaid players may respond by withholding effort in the future. This may not be a rational response in this particular case, however, as even the most underpaid NHL players are likely making more than they could in their next best option, and they all have a strong incentive to continue playing in the NHL. Each player knows, for instance, that there are hundreds of minor league players who are eager for the chance to take their place in the NHL. A more rational response from such a player, then, after learning that the NHL labor market rewards offense more than defense, would be to still exert maximum effort, but to shift how they allocate that effort from defense to offense. Players would not be able to get away with drastic shifts in effort allocation without being noticed, but subtle changes over the course of a season could potentially

[7] These regressions include points (goals + assists), shots, games played, plus-minus, age and penalty minuts (PIMs) and also include age and position fixed-effects

lead to statistical improvements in offense that could lead a player to earn a higher salary in the long run. It is particularly plausible that players could make such a change in the early 90s when virtually no defensive statistics were being tracked aside from plus-minus.

The idea that individuals will respond to inequity by altering their behavior in such a way as to correct it dates back to an important social psychology literature from the 1950s and 60s. In his 'Theory of Cognitive Dissonance', [136] writes that the existence of inequity creates tension in an individual which is proportional to the magnitude of the inequity, and that the tension will motivate the individual to reduce or eliminate the inequity. Similar to [17], [181] and [175] both argue that the threshold of inequity which must be attained before a response is triggered is larger for overpaid than for the underpaid, though a series of experiments by John Stacy Adams ([14], [10], [11]) all find that overpaid workers do respond to inequity in ways designed to reduce or eliminate it, so it is unclear whether we should expect to see an asymmetrical response to salary disclosure in the NHL.

The other important strand of research which contributes to how we might expect players to respond to salary disclosure investigates reference-dependent utility theories, which are based on [310] and [191]'s 'prospect theory.' These theories posit that utility depends not just on a realized outcome, but how that outcome contrasts with the individuals reference point or expectation for the outcome. These theories have been used to explain why cab-drivers quit working early when they are making more than they expected[8] , declining police performance after losing a salary arbitration case ([223]), and the link between surprise upset losses for NFL and NBA teams and domestic violence in the cities of the losing teams ([73], [75]).

Reference-dependent utility theories would predict that whether we observe a behavioral response from players depends on how their revealed status compares with their prior beliefs about where they stood in the salary hierarchy. If, for example, underpaid players knew they were being underpaid, perhaps because they play in smaller media markets where teams are not earning as much revenue, then learning that they were being paid less than their equally productive counter-

[8] See [72], [130], and [106]

parts would not be surprising and would thus not lead to any particular response. In the popular model developed by [206], players' reference points would be based on their rational expectations held in the recent past, in this case the period just before salaries were disclosed.

In this setting, we would expect players who are surprised to learn they are underpaid to exhibit the strongest behavioral response to salary disclosure. Players likely have some idea of how relatively rich or poor their team is, based on the quality of amenities in their stadium and training facilities, the ticket prices the team charges, and how consistently they are able to fill their stadium. Players on a poor team would likely not be surprised to find out they are being underpaid compared with the rest of the league. Players on a wealthier team, on the other hand, would have a higher reference point, and would be more likely to be surprised to find out they are underpaid. Because of this, reference-dependent utility theory would predict that underpaid players on higher-paying teams would exhibit the largest response to salary disclosure.

2.3 Data/Descriptive Findings

In order to understand the effects that salary disclosure had on the NHL's labor market, I gathered data from four sources. I obtained official salary information for every player in the league from 1990-2018 from markerzone.com, as well statistical performance data for both players and teams from 1970-2018 using hockey-reference.com. These data are also available elsewhere, and I performed a series of checks to make sure these sources lined up with what was available at other sites[9] . Appendix Table B.1 displays the linkage rates across the salary and statistical data. In each year, about 70% of players who show up in either dataset are matched. There are typically around 200 players, however, who appear in the statistical data but not the salary data. These are typically minor league players who are called up at some point in the season, but who do not have an NHL contract. This is why their average number of games played is far fewer than for players with both salary and statistical data.

[9] For example, salary data is also available at capfriendly.com and statistical performance data can be found at hockeydb.com. To make sure my data were accurate, I randomly selected dozens of players and compared the salary and statistical data in my dataset with these alternative options and found no discrepancies.

In order to test for whether players changed their behavior during the 1989-90 season, I gathered data from NHL.com, which allows users to view statistics over a given date range. This enabled me to gather performance statistics for the 1989-90 season, both before and after January 29. Finally, I gathered data on every player transaction including trades, drafts and free agent signings from 1980-2018 using prosportstransactions.com.

Because the statistics for goaltenders are not comparable to the rest of the skaters, and because the sample size for goaltenders is too small to obtain precise estimates when analyzed on their own, I omit goaltenders from all of the analysis in this paper. Figure 2.1 displays histograms of the salary distribution for each year from 1990-1996, with 1994-1995 omitted because of the shortened season due to that year's lockout. In all graphs in this section, a year on the graph will represent the season ending in that year. So 1990 on the graph represents the 1989-90 season, and so forth.

It is worth noting that 1990 captures the state of salaries in the NHL in the pre-salary disclosure era, as salaries were released towards the end of the season and contract negotiations typically take place before the season begins. Looking at salaries in 1990, they appear to be tightly distributed and somewhat truncated on the left. The two outliers in the right tail represent Mario Lemieux and Wayne Gretzky who were making $2 million and $1.72 million, respectively. It seems that a premium was paid to the super-elite players, but most everyone else was clustered relatively closely around the mean salary of $202,000, with a league-wide standard deviation of $141,000. In 1991, a right-tail begins to develop, and this continues into 1992 and beyond, with a much larger right tail and more dispersion overall. In the first two years after disclosure, the inflation-adjusted league-wide mean increased by 41% to $284,000, with the standard deviation increasing 66% to $236,000. By 1996, the distribution has changed drastically, with an inflation-adjusted mean of $689,000 and standard deviation of $691,000.

Another interesting question is whether more players started to leave their team in free agency in the wake of salary disclosure. The number of players switching teams in free agency in each calendar year from 1980-2000 is reported in Figure 2.2. From 1980-1990 there was an average of 23 free agent moves per season, with a maximum of 32 in 1985. In 1990, the year disclosure took place, 28 free agents switched teams, a slight uptick from the previous year but still within the normal range for that period. Shortly after this, however, the number of players leaving their team in free agency began rising rapidly as the league's rules regarding free agency were liberalized. There was a minor dip in 1994, but overall free agent moves skyrocketed throughout the decade, settling between 100-120 from 1998 onward. Clearly, the labor market in the NHL changed dramatically in the years directly following salary disclosure, but it is unclear how much of this is causal versus being due to changes that were already in motion. In order to understand the causal impact of salary disclosure, I now focus specifically on the behavioral response of underpaid and overpaid players just after disclosure took place.

2.4 The Effect of Salary Disclosure on Player Performance

The conceptual framework described in section 2 suggests that when players realize they are being underpaid, they may shift their efforts from defensive production to offense, and we may expect this change to be larger for underpaid players on typically high-paying teams. Before salary disclosure, absent knowledge of other players' salaries, NHL players may assume they are being paid fairly, as the average salary of $202,000 was over 16 times the United States GDP per capita of $11,944[10] . Without detailed information about what their peers are making, players may have set their reference points flexibly, believing that a range of pay outcomes could be considered fair. This uncertainty would make underpaid players more likely to believe their compensation was reasonable. When salary disclosure occurs, however, having complete information would cause players to update their reference point and respond accordingly.

That being said, the timeline for when we might expect to see a behavioral change show up is not immediately clear. As of January 29, 1990, teams had played between 49-52 of their 80 regular season games, or 61%-65%. It is possible that we might see changes take place immediately, with

[10] Per https://data.worldbank.org, US GDP per capita in 1990 was $23,888 in current $US. Adjusting for inflation means 1990 US GDP per capita was $11,944

players digesting the new information and then responding accordingly. Hockey players are not econometricians, however, and it likely would have taken some time and effort to form an honest opinion of whether they were being fairly compensated, by comparing their salary and statistics to many other with whom they believe they have similar abilities. Doing this would have been difficult while maintaining their grueling schedule of practices, travel, and playing three to four games per week, on average. At that point in the season, players would also be focused on making the playoffs to give themselves a chance at winning the Stanley Cup. At that time, 16 of the 21 NHL teams made the playoffs, so almost every team still had a chance to win. Frustrated underpaid players could also have decided to give their team a chance to rectify the situation in the coming offseason. Ultimately, when we might expect to see a behavior change is an empirical question.

I begin by looking for a midseason response taking place immediately after the January 29 disclosure of salaries. In order to identify which players were overpaid and which were underpaid, I run an OLS specification to predict log salary in 1989-90 based on performance statistics from the 1988-89 season. I do this using a lasso model which selects the most predictive variables from a list including a third degree polynomial in goals, assists, an interaction of goals and assists, plus-minus, games played, penalty minutes, shots on goal, game-winning goals, a third degree polynomial in age as well as age and position fixed-effects. I include the third degree polynomial in goals because I expect that the marginal benefit of scoring an additional goal is likely different for a player with five goals compared with a player with 40. Similarly, I expect there to be non-linearities in the effect of age over time as more experienced players can command higher salaries, but as players age their abilities eventually decline. The goals*assists interaction sheds light on whether it is more valuable to be a player who gets a decent amount of both goals and assists or whether it is better to get a high number of one or the other. I intentionally leave out team fixed effects because I believe that players are concerned with how their salary fits into the overall league distribution and not just within their team. The lasso model keeps everything except for goals cubed. I use the residuals from this regression to create two new variables:

$$Under_i = |min(R_i, 0)| \tag{2.1}$$

and

$$Over_i = max(0, R_i) \tag{2.2}$$

where $Under_i$ is the absolute value of the residual for all players for whom the residual is negative. In other words, $Under_i$ measures, in log terms, the extent to which player i is underpaid, and takes a value of zero for all overpaid players. Likewise, $Over_i$ measures the extent to which player i is overpaid and takes a value of zero for all underpaid players. For each player, I then calculate the change in each statistic (goals, assists, shots, plus-minus) on a per-game basis from the portion of the 1989-90 season played under pay secrecy to the games played after disclosure. For example, a player who scored 20 goals in 40 games before January 29, but increased their per-game production after disclosure to 15 goals in 20 games would have a change in goals per game of .25 ($\frac{15}{20} - \frac{20}{40} = .75 - .5 = .25$). I then use this data to test whether there was a midseason response by estimating the following specification on the 389 players who played in the 1988-89 season as well as in both the pre and post salary disclosure portion of the 1989-90 season:

$$\Delta Y_{iap} = \beta_1 Under_i + \beta_2 Over_i + \gamma_a + \delta_p + \epsilon_i \tag{2.3}$$

where ΔY_{iap} measures the change in a performance statistic per game from before to after salary disclosure for player i of age a who plays position p. γ_a and δ_p are age and position fixed effects. Results on a number of statistical outcomes are displayed in Table 2.1.

There is very little evidence of a midseason response to disclosure, with nine of the ten estimates having p-values greater than .25. There appears to be a slight reduction in goals per game for overpaid players (p-value = .09), while the change for underpaid players is a relatively precisely estimated zero. Since there are ten coefficients estimated, we should expect that one of them would be significant at 10% merely by chance. There also does not appear to be any trend across the statistics for overpaid vs. underpaid players. If salary disclosure led to any changes in player performance, it appears that it did not take place until at least the following season.

In order to test whether there was a behavioral response in the following season, I run a similar lasso specification which predicts log salary for the 1989-90 season based on performance statistics for the same season. The lasso specification chooses the same variables as in the previous specification, except that it also omits age squared and age cubed. A regression table of these coefficients can be found in Appendix Table B.2. The goals-assists interaction and shots on goal are the most significant positive predictors in salary, while plus-minus is close to zero and insignificant (p-value = .62). Since offensive production is controlled for, plus-minus mostly becomes a measure of defensive performance in this regression, and the fact the it is not significant reflects the fact the defensive production is not highly compensated in the NHL labor market[11] . A residual plot of this regression is displayed in Appendix Figure B.5 and it appears to be completely noise.

As mentioned above, one reason why there may not have been a midseason response to salary disclosure might be that players wanted to give their team a chance to rectify the situation in the offseason. Indeed, teams may have informally suggested that they would address some of the more egregrious discrepancies that players were facing. Of particular interest, then, is what happens to the players who are deemed to be underpaid based on their performance in the 1989-90 season, who then do not receive a raise in the summer of 1990.

To investigate this, I break the group of 362 players for whom both salary and statistical performance data is available for both the 1990 and 1991 seasons into four groups: underpaid players who don't receive a raise in the summer of 1990 (henceforth UN), underpaid players who do receive a raise (UR), overpaid players who do not receive a raise (ON) and overpaid players

[11] The goals-assists interaction has a p-value of just .004, while shots has a p-value of .000. Games played is actually negative and significant, meaning that if, for example, a player is going to score 20 goals and get 20 assists, they are rewarded for doing that in fewer games played. While the goals-assists interaction is positive and significant, both goals and assists are individually weakly negative, suggesting that what is important is being a player who gets both goals and assists, not just one or the other. Penalty minutes, which could be used a proxy for aggression, is weakly positive, while plus-minus is weakly negative. Age is the only other significant predictor, indicating that player salaries rise as they get older. When controlling for all of these measures, they are no significant differences in salary across positions. This does not mean forwards and defensemen are paid the same, however, as defensemen score fewer points and take fewer shots.

who receive a raise (OR). Many player's contracts at this time were designed to give them nominal raises (often of around $10,000) each year, but I am only interested in players who negotiated a new contract (or were not able to), so I count any player whose salary increased by more than 20% between 1990 and 1991 as having received a raise. Importantly, as I will show in the robustness section of the appendix, the results that follow are robust to setting this threshold at either 15% or 25% and the economic interpretation of the parameters remains unchanged. This calculation leaves 86 players in the UN group (underpaid/no raise), 91 players in UR (underpaid/raise), 129 players in ON (overpaid/no raise) and 56 players in OR (overpaid/raise). This breakdown makes intuitive sense, as among underpaid players, 51% (91/177) receive a raise whereas only 30% (56/185) of overpaid players receive a raise. Also, just under two thirds of players (215/362) do not receive a raise, suggesting an average contract length of around three years.

Table 3.1 displays the means of various performance measures for the 1989-90 season across the four groups along with a p-value from a test of whether the means of all four groups are equal. The overpaid group (ON & OR) clearly makes more than the underpaid group, which is unsurprising, but aside from that there is relative balance across the four groups. They have very similar numbers for goals, assists, points, plus-minus, penalty minutes (PIMs), shots on goal and shooting percentage. Age is the only metric for which equality across the four groups can be rejected, with the no-raise group (UN & ON) being slightly older than the big-raise group, but even the oldest group (UN) has an average age of 26 which is still young in hockey terms and in the early prime of their careers. Overall, the balance across Table 3.1 suggests that whether a player got a raise in the summer of 1990 or not has less to do with how good a player they are, and is likely more dependent on whether or not that player's contract was up to be renegotiated[12] .

Using these groupings, I then estimate the following model:

[12] To test whether the UN group in 1990 differed in ability from UN players in subsequent seasons, Appendix Table B.3 displays the same performance measures for the UN group in 1990 compared with UN players in 1991 and 1992. Again, aside from slight differences in age, the groups are similar in nearly all performance measures.

$$\Delta Y_{iap} = \beta_1 Under_i * NoRaise_i + \beta_2 Under_i * BigRaise_i + \beta_3 Over_i * NoRaise_i + \beta_4 Over_i * BigRaise_i + \gamma_a + \delta_p + \epsilon_i \tag{2.4}$$

where $Under_i$ and $Over_i$ are calculated as before and can therefore be interpreted as the amount in percentage terms which player i was under or overpaid in the 1989-90 season. ΔY_{iap} measures the change in a performance statistic from 1989-90 to 1990-91 for player i of age a who plays in position p. $NoRaise$ is an indicator for receiving less than a 20% pay raise going into the 1990-91 season, while $BigRaise$ indicates at least a 20% raise. β_1 through β_4 measure the differential performance change due to the amount which a player is over or underpaid for each group of players. γ_a and δ_p are age and position fixed-effects. Results from estimating Equation 4 on a range of statistics are reported in Table 2.3.

Column 1 shows that UN players saw significant upticks in goalscoring. β_1 can be interpreted as meaning that for an underpaid player who did not get a raise going into the 1990-91 season, a 10% increase in the amount which they are underpaid is associated with an approximate one goal increase in output (p-value = .001). There is virtually no change for formerly underpaid players who received a raise going into the new season, or for either of the groups of overpaid players. Column 2 shows that UN players also increased their shots on goal (p-value = .033), while column 3 shows that their assists went up by about half as much as their goals, though this result is statistically insignificant (p-value = .19).

It could be that UN players are shifting their effort towards offense, or it could be that they are just playing better overall, perhaps looking to prove themselves after discovering they were being undervalued. It is worth pointing out that goals increase by substantially more than assists, which is surprising considering the average player recorded approximately 15 goals and 24 assists in 1990. It makes sense that players get more assists than goals in general because any given goal only has one goalscorer, but can have up to two assists credited for it. If UN players were improving their overall play, we could reasonably expect assists to increase by more than goals. The fact that goals increase by more is suggestive that these players are focused more on scoring goals themselves than on increasing the team's overall offensive production.

Plus-minus is a good statistic to flush out whether this increase in scoring is due to better play overall or more offensively-focused play. Column 4 estimates the change in plus-minus from year-to-year for different players. It is immediately telling that UN players see a drop in plus-minus when they themselves are scoring more, even though the estimate is not significantly different than zero (p-value = .20). There are two reasons why this result is surprising. First, since plus-minus is mechanically correlated with a player's goals and assists (they get a +1 for every goal they score or assist they provide), they should move in the same direction 'ceteris parabis'. The fact that the estimate is negative indicates that these players' teams are getting scored on *even more* than the additional goals they are creating. Second, these coefficients are measuring these players relative change compared to the rest of the league. Remember that a plus for one player is a minus for the opposing player. When UN players score additional goals, not only is their own plus-minus rising but their opponents' plus-minuses are falling, which makes it all the more surprising that plus-minus appears to be falling for the group which is scoring more.

Although the UN estimate in Column 4 is not significantly different than zero, it is significantly different than the estimate for the underpaid/raise group (p-value = .021) and the overpaid/no raise group (p-value = .001). Column 5 shows the results for an adjusted version of plus-minus, which holds a player's offensive production constant from the previous year. Adjusted plus-minus asks the question 'what would have happened to this player's plus-minus if their offensive production remained the same as the prior year'. Since plus-minus tracks both offensive and defensive contribution, holding the offensive side constant means adjusted plus-minus mainly tracks changes on the defensive side.

The estimate of β_1 in Column 5 shows that a 10% increase in the amount in which a player is underpaid is associated with a 2.6 point decrease in their adjusted plus minus (p-value = .005). It looks convincingly like these players are sacrificing defense in order to pursue personal offensive statistics. Since offense is more highly rewarded in salary negotiations, this is likely an attempt to improve their bargaining position for their next contract. Interestingly, the coefficient for UR players is positive and significant. This would suggest that an underpaid player who receives the raise they deserve responds by 'buying in' and playing a more team-oriented game. The fact that they are also taking fewer shots on goal suggest that perhaps they are filling in the defensive gaps left by their UN counterparts.

In Appendix D, I run a series of robustness checks on the main results from above. In short, they are robust to including team fixed effects, controlling for changes in games-played and penalty minutes as proxies for playing time, using a more fully specified predictive model instead of the lasso model, using a much simpler predictive model, changing the threshold for whether a player received a raise to 15% or 25%, dropping the outliers Wayne Gretzky and Mario Lemieux from the predictive model, and dropping any combination of up to three observations from the UN group.

While this outcome seems correlated with salary disclosure in the NHL, it is also possible that this is just a constant feature of the NHL, where underpaid players become disgruntled and play more selfishly. In order for this to be related to salary disclosure, these impacts should be most pronounced in the first year after disclosure took place, before the market has had time to adjust to the new information. To address this, Figure 2.3 plots the same coefficients for goals, assists, plus-minus and adjusted plus minus for every year from 1991-2004. Going into each new season, I once again sort players into the same four groups, based on whether they were overpaid or underpaid based on their performance in the prior year, and by whether they received a raise in the offseason. For goals, the 1991 coefficient is the only one of the 14 estimates that is positive and significant, and is significantly higher than the coefficient in every other year except 1996, which followed the shortened 1994-95 season. Similarly, the adjusted plus-minus coefficient in 1991 is the most negative of the coefficients, confirming that the results in Table 2.3 are unique to the period immediately following salary disclosure. In Appendix E, I track what happens to each group of players over time and demonstrate that the behavior change I document above both led underpaid/no raise players to be more likely to receive a raise and increased the size of the average raise received.

Next, I test whether the behavioral change in the UN players is greater for underpaid players on relatively richer teams. As noted above, reference-dependent utility theory would predict the behavioral response to salary disclosure to be larger for underpaid players on relatively richer teams. I test for this by further splitting each group into players on a team whose overall payroll was above the league average when salaries were disclosed and players on a team whose overall payroll was below the leaguewide average. Henceforth, I refer to UN players on a team with an above average payroll as UNA and UN players on a team with a below average payroll as UNB. Doing this puts 39 players into UNA and 47 players in to UNB[13] . With the increasingly small sample size, we might expect statistical power to become an issue, but there are enough players in each group to potentially see some effects. Table 2.4 reestimates Equation 1 using these new groupings.

The first thing to note is that the main results from Table 2 apply to both UNA and UNB, with large increases in goals and decreases in adjusted plus-minus. Regardless of where a players' team falls in overall payroll, underpaid players who do not get a raise appear to shift their effort from defense to offense. On the other hand, this shift appears to be substantially larger for underpaid players on teams with above average payrolls. The increase in goals is nearly three times as large, from about five and a half (p-value = .046) for UNB to about 14 (p-value = .000) for UNA. A test for equality of the coefficients on UNB and UNA when measuring their effect on goals yields a p-value of .016. The increase in shots for UNA is about 44 compared with just 14 for UNB, though because of the large standard errors the p-value for equality of the two coefficients is now .2. The change in assists is now weakly negative for UNB, compared with an increase of 12 for UNA, with a p-value comparing the two coefficients of .10. The change in (unadjusted) plus-minus actually appears to be larger for UNB (p-value = .50), but this could be a result of the mechanical correlation that exists between plus-minus and goals and assists.

Column 5 indicates that this is likely the case, as when this mechanical correlation is removed the decrease in adjusted plus-minus appears to be much larger for UNA. This result is also somewhat underpowered, however, likely because of the smaller sample sizes, and the p-value for the difference

[13] Similary, this puts 49 players in URA, 42 players in URB, 78 players in ONA, 51 players in ONB, 24 players in ORA and 32 players in ORB

between the coefficients on UNA and UNB is .18. Overall, the results in Table 3 are consistent with the predictions of reference-dependent utility, with underpaid players on high-paying teams appearing to adjust their behavior more drastically than underpaid players on lower-paying teams. Appendix Table D.2 reestimates this specification with team fixed effects included in order to control for the possibility that it is easier to get more points on a higher-paying team. If anything, however, the results are more convincing when team fixed effects are included.

I now consider whether these changes that happen on an individual level aggregate up to affect overall team performance. Because underpaid players appear to be shifting their effort from defense to offense, their teams are getting scored on by even more than the additional goals they provide. On aggregate, this is obviously problematic for these players' teams, and it could lead to a correlation between team payroll and team performance. Even though these effects appear to be larger for underpaid players on higher-paying teams, lower-paying teams have more underpaid players and thus might still perform worse overall.

To this end, Figure's 2.4 and 2.5 plot the relationship between team salaries and regular season league performance for 1989-90 and 1990-91, respectively. Most teams had between 22-28 salaried players in these two seasons, likely due to a combination of injuries, minor-league callups, performance struggles and other on-ice reasons. To prevent teams with more players from looking like they pay more, I use the log of the sum of each team's top 20 earners, though the results that follow are robust to using either the top 15 or even 10 players.

There appears to be no correlation in 1989-90. The Quebec Nordiques were by far the worst team in the league with only 31 points, less than half of second-worst Vancouver with 64, and yet they were the 12th highest-paying team out of the 21 teams in the league. There even appears to be a slight downward relationship, with the two highest paying teams, the Los Angeles Kings and the Pittsburgh Penguins, both struggling. In 1990, 16 of the league's 21 teams made the playoffs, but the Penguins missed out entirely and the Kings barely made the playoffs, finishing with the league's 15th best record. They would eventually get swept four games to zero in the second round of the playoffs by the eventual champion Edmonton Oilers. Everything else on Figure 2.4 looks to be little more than noise, with no evidence that higher pay translates into better performance.

In Figure 2.5, on the other hand, the relationship is obvious, with a clear upward trajectory suggesting that teams which paid higher salaries performed much better. In 1991, both the Kings and the Penguins (still by far the highest paying teams) comfortably made the playoffs with the Penguins eventually winning the Stanley Cup (the Kings lost in the second round once again to the Oilers). Perhaps the most interesting thing about this transformation is that it happened so quickly. Over time, we might expect the best players to begin to relocate to the higher paying teams, but when salary disclosure occurred, most players were under contract for a period of multiple years, making them unable to switch teams or renegotiate their salaries right away. As mentioned above, 28 free agent moves took place in 1990, which was similar to the number of players switching in previous seasons. Also, because of the restrictive nature of free agency in the NHL at the time, the majority of these players were fringe-level players who switched teams more because their previous teams were not interested in resigning them, rather than in search of bigger contracts. This suggests that this change is due more to changes in how players performed as opposed to changes in which teams had which players.

Figure 2.6 plots the coefficients from a regression of regular season points on team salary with 95% confidence intervals for 1990-2004. There appears to be a clear shift from 1990 to the era of salary transparency. In 1989-90, the coefficient is slightly negative with a p-value of .16. For each of the next 14 years the coefficient is positive each year and is significant at 5% every year except 1995 and 1998. A test for equality of the coefficients in 1990 and 1991 produces an F-statistic of 25.44 (p-value = .0000), whereas equality cannot be rejected at even 10% for any other consecutive years in the sample.

Interestingly, the positive correlation between payroll and performance is the strongest in the first year after disclosure, which suggests an immediate behavioral adjustment which leveled off after the initial change. The relationship between payroll and performance declines in magnitude over time, perhaps as discrepancies are addressed when underpaid players are able to negotiate a new contract, but the correlation remains positive and mostly significant with estimates becoming more precise over time. There are two potential identification threats for this to be a causal effect of salary disclosure. The first is if there was another change in the NHL labor market which made salary more important for team performance. Since there were no major changes to the CBA or any other significant policy changes in 1990, this seems unlikely. The second threat is that perhaps the lack of a correlation in 1989-90 was simply an anomaly, and that in the years before disclosure it was normal for there to be a positive association between payroll and performance. Unfortunately, this is untestable as salary information is unavailable for seasons before 1989-90, but the fact that there is such a sharp break in the relationship which evolves relatively smoothly in the subsequent years suggests that this explanation is not likely.

2.5 Conclusion

In this paper I exploit a natural experiment in the National Hockey League in order to understand how employees respond to learning their place in the salary hierarchy. I find that underpaid players respond by shifting their effort from defensive production to offense, which is more highly rewarded in the NHL labor market. This pattern is consistent with agents working to maximize their own future pay instead of overall firm performance. Asymmetrically, overpaid players do not appear to become more defensive-minded after learning they are overpaid. Consistent with reference-dependent utility theory, these behavioral responses are most pronounced for underpaid players who play for higher-paying teams. Overall, these results are consistent with other experimental, observational and theoretical studies which find that having low relative pay can lead to decreased job satisfaction. At a firm level, I find that while total payroll was unrelated to firm performance before salary disclosure, the two became immediately and permanently linked after disclosure took place.

In terms of policy, these findings suggest that low-paying firms have a strong incentive to keep salary information secret, while higher-paying firms could benefit from policies designed to increase salary transparency. If salaries are disclosed, low-paying firms will have an incentive to increase wages, as underpaid agents who received a substantial raise appeared to continue to allocate their effort in order to maximize team performance.

Table 2.1: Effect of Salary Disclosure on Performance - Midseason 1989-90

	(1) Goals PG	(2) Assists PG	(3) Points PG	(4) Shots PG	(5) +/- PG
Underpaid	-0.00126	0.0372	0.0360	0.0648	0.130
	(0.0380)	(0.0624)	(0.0723)	(0.145)	(0.147)
Overpaid	-0.0873	0.159	0.0715	0.161	0.231
	(0.0491)	(0.147)	(0.130)	(0.225)	(0.215)
Observations	389	389	389	389	389

Note: This table displays regression coefficients comparing changes in statistical output in response to salary disclosure during the 1989-90 season for players deemed overpaid and underpaid based on their 1988-89 performance statistics using salary data from markerzone.com and statistical data from NHL.com. Age and position fixed effects are included in each specification.

Table 2.2: Balance on Observables

	UN	UR	ON	OR	p-value
Log Salary	11.96	11.85	12.33	12.28	0.00
Goals	13.44	16.54	13.29	16.12	0.19
Assists	24.32	25.38	21.93	25.33	0.44
Points	37.76	41.92	35.22	41.45	0.29
+/-	0.26	-0.99	-0.03	3.48	0.22
PIMs	71.44	87.37	86.38	76.63	0.34
Shots	104.47	118.17	106.18	122.78	0.29
Shooting %	11.30	11.66	10.28	10.45	0.34
Age	26.43	25.31	26.09	25.03	0.04
Observations	86	91	129	56	

Note: This table compares statistical means for the four groups of players for the 1989-90 season using data from hockey-reference.com. The four groups are UN (Underpaid players who did not receive a raise), UR (Underpaid players who did receive a raise), ON (Overpaid players who did not receive a raise) and OR (Overpaid players who received a raise).

Table 2.3: Effect of Salary Disclosure on Performance - 1990-91

	(1) Goals	(2) Shots	(3) Assists	(4) +/-	(5) Adj. +/-
Underpaid/No Raise	10.01	29.83	5.475	-10.26	-25.74
	(2.611)	(13.02)	(4.043)	(7.772)	(8.051)
Underpaid/Big Raise	-0.465	-19.61	-5.260	9.553	15.28
	(3.093)	(9.033)	(3.240)	(5.016)	(5.354)
Overpaid/No Raise	-0.198	-0.114	-2.378	16.20	18.77
	(3.949)	(23.11)	(7.567)	(4.635)	(14.71)
Overpaid/Big Raise	3.208	44.00	11.32	5.692	-8.841
	(3.273)	(29.69)	(7.941)	(6.579)	(15.23)
Observations	362	362	362	362	362

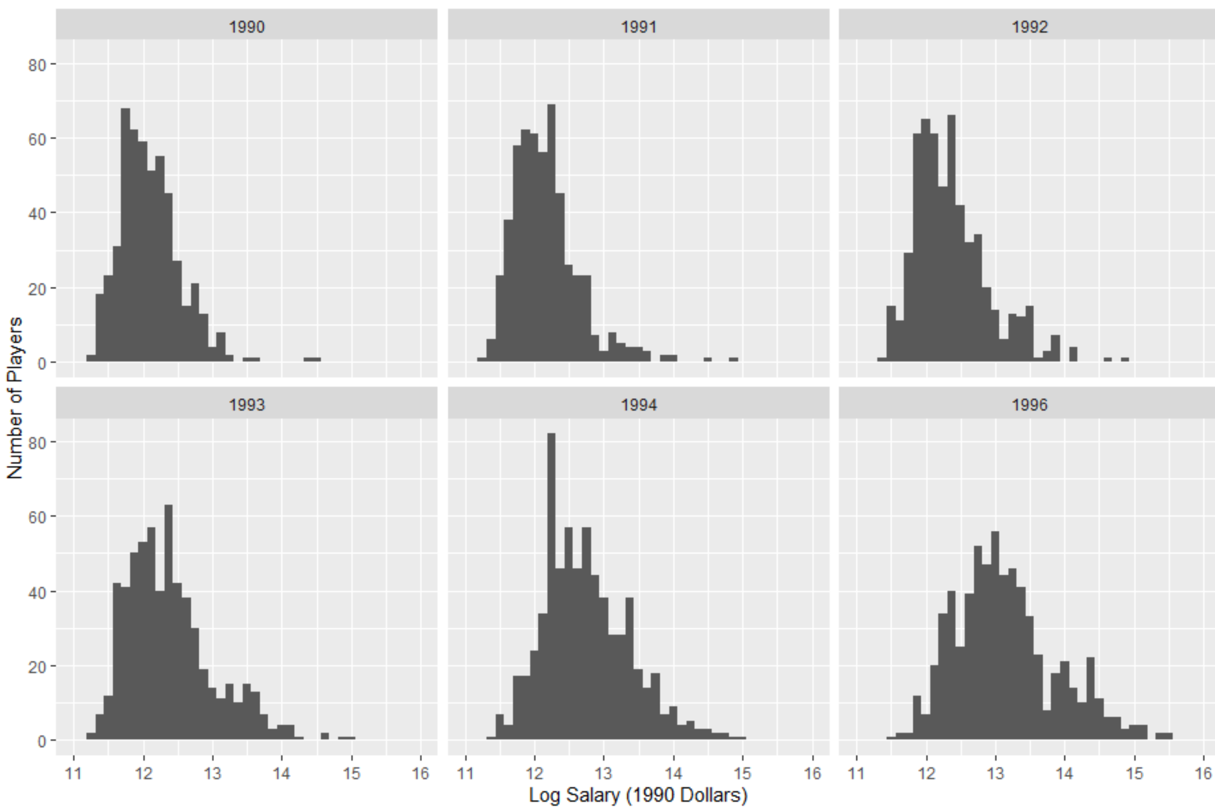
Note: This table displays regression coefficients comparing changes in statistical output for the eight different groups from 1989-90 to 1990-91 using salary data from www.markerzone.com and statistical data from www.hockey-reference.com. Age and position fixed effects are included in each specification.

Table 2.4: Reference-Dependent Utility - 1990-91

	(1) Goals	(2) Shots	(3) Assists	(4) +/-	(5) Adj. +/-
Underpaid/No Raise - Below	5.422	14.17	-1.754	-15.05	-18.72
	(2.547)	(20.85)	(4.823)	(7.676)	(9.786)
Underpaid/No Raise - Above	14.22	44.20	12.11	-5.856	-32.18
	(2.480)	(11.78)	(5.889)	(11.55)	(9.264)
Underpaid/Big Raise	-0.520	-19.80	-5.347	9.495	15.36
	(3.107)	(9.246)	(3.406)	(5.025)	(5.445)
Overpaid/No Raise	-0.258	-0.319	-2.473	16.13	18.87
	(3.997)	(23.29)	(7.675)	(4.542)	(14.79)
Overpaid/Big Raise	3.160	43.84	11.25	5.641	-8.767
	(3.242)	(29.62)	(7.958)	(6.663)	(15.18)
Observations	362	362	362	362	362

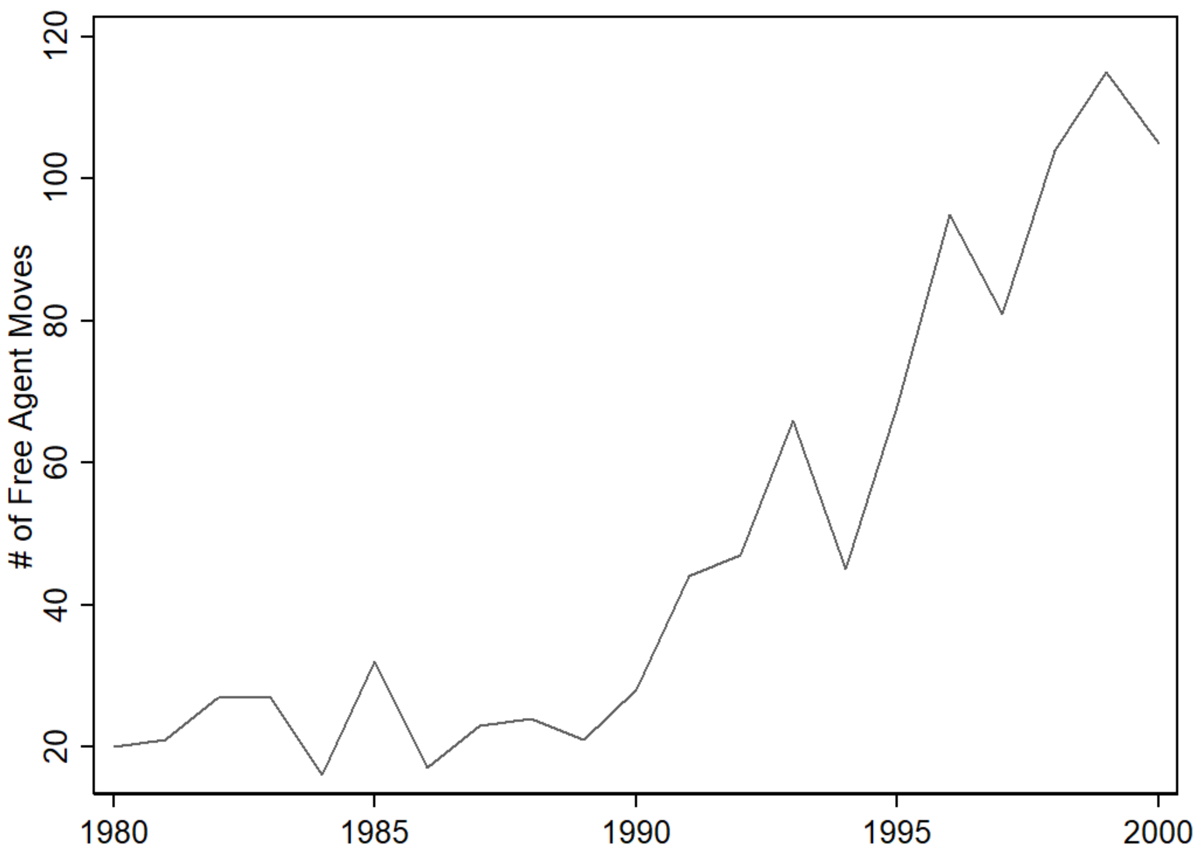
Note: This table displays regression coefficients comparing changes in statistical output for five different groups (underpaid/no raise players on below average payroll teams, underpaid/no raise players on above average payroll teams, underpaid/big raise players, overpaid/no raise players, and overpaid/big raise players) from 1989-90 to 1990-91 using salary data from www.markerzone.com and statistical data from www.hockey-reference.com

Figure 2.1: NHL Salary Distributions Following Salary Disclosure



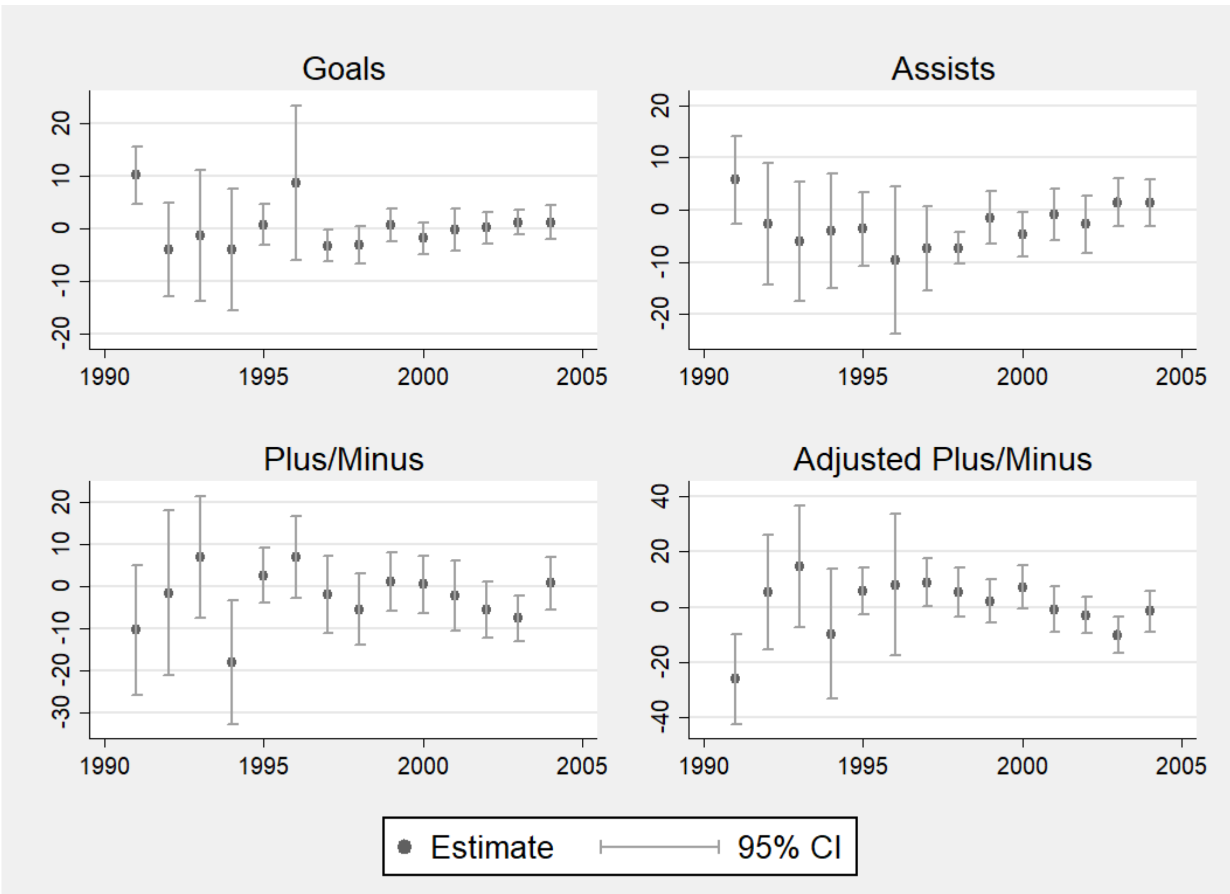
Note: This figure displays salary distributions using data from markerzone.com for each year from 1989-90 to 1995-95, omitting 1994-95 due to the shortened season. In each graph, the year listed is the later year in the season, so 1990 represents the 1989-90 season.

Figure 2.2: Yearly Free Agent Moves - 1980-2000



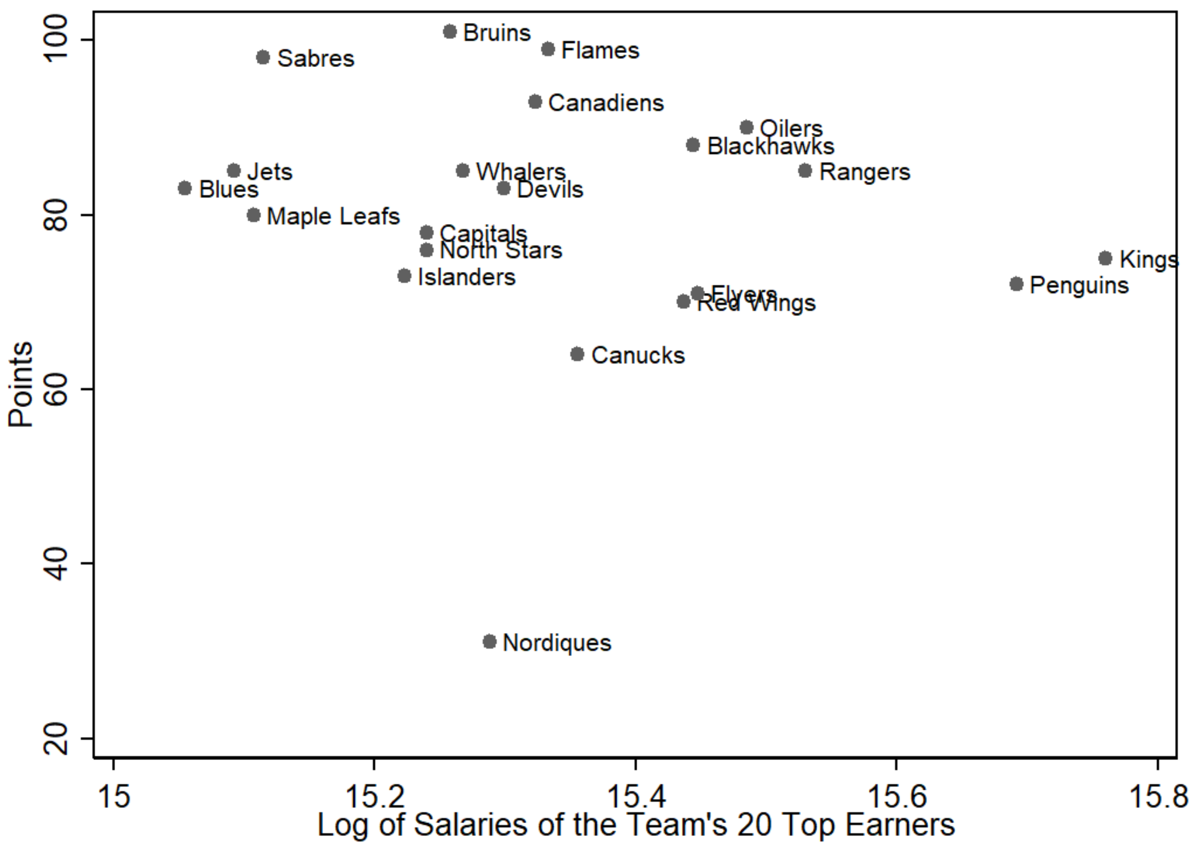
Note: This figure displays the number of players switching teams in free agency in each calendar year from 1980-2000 using data from prosportstransactions.com.

Figure 2.3: Coefficient on Underpaid/No Raise by Season - 1991-2004



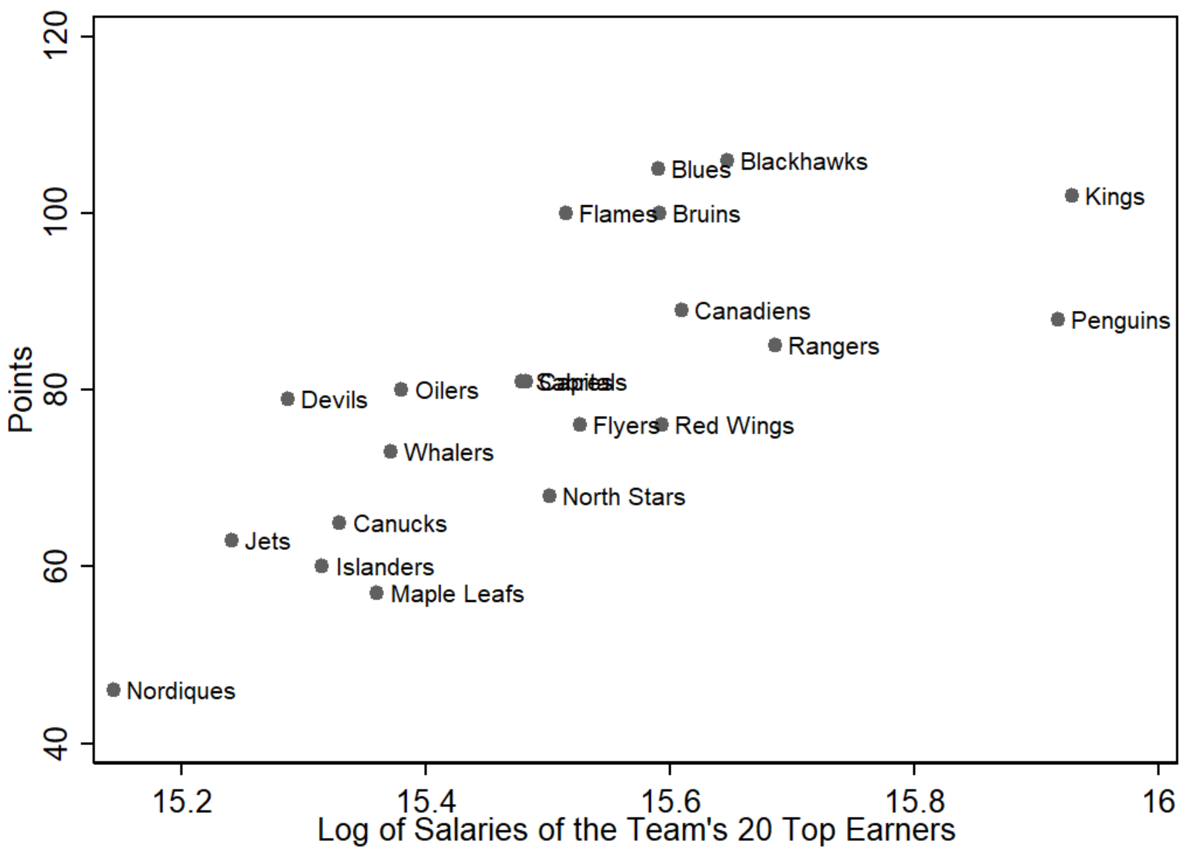
Note: This figure displays coefficients estimating the effect of being underpaid on player performance each year from 1990-91, the first season after salary disclosure to 2003-04. These coefficients are estimated using salary data from markerzone.com and statistical data from hockey-reference.com

Figure 2.4: Salaries and Team Performance: 1989-90



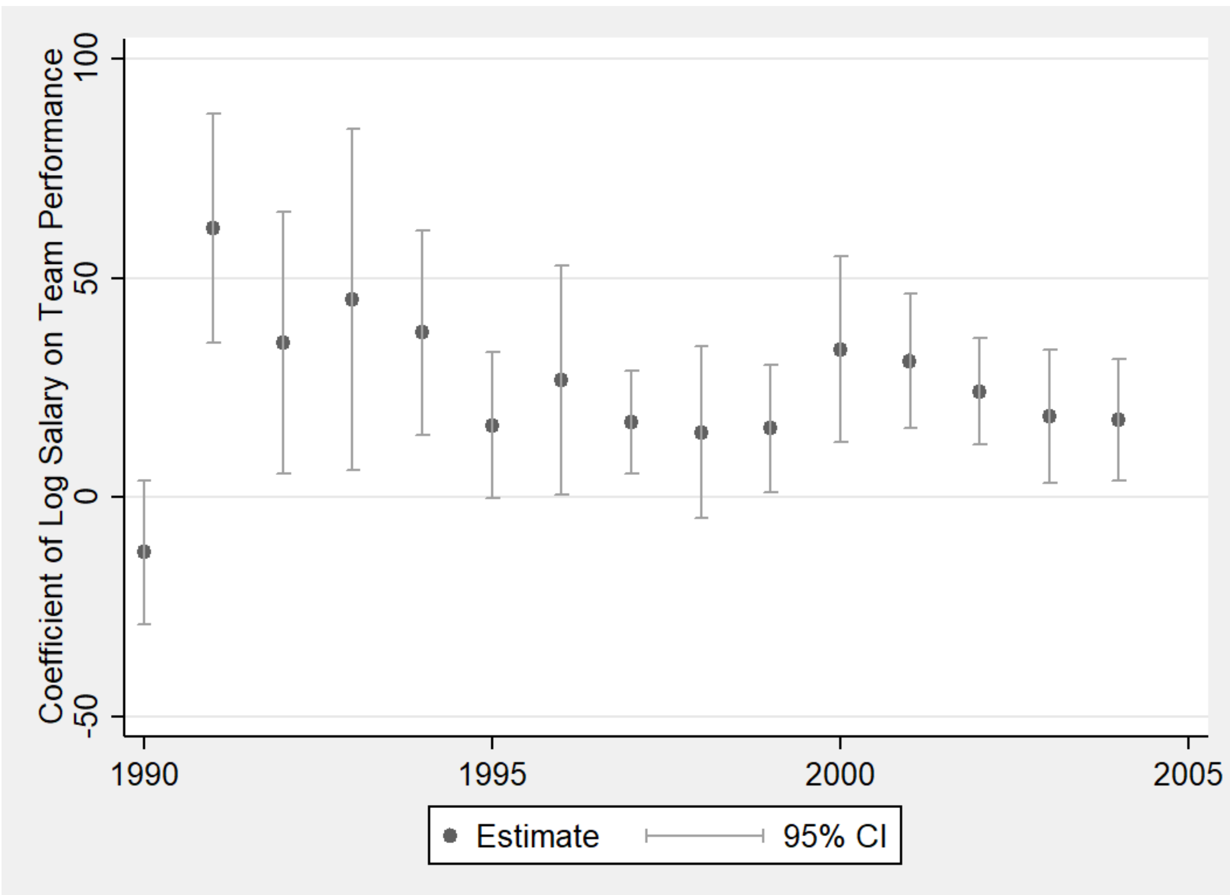
Note: This figure shows the relationship between team payroll and regular-season performance for the 1989-90 season using salary data from markerzone.com and statistical data from hockey-reference.com

Figure 2.5: Salaries and Team Performance: 1990-91



Note: This figure shows the relationship between team payroll and regular-season performance for the 1990-91 season using salary data from markerzone.com and statistical data from hockey-reference.com

Figure 2.6: Coefficient of Team Payroll on Team Performance: 1990-2004



Note: This figure shows the relationship between team payroll and regular-season performance for each season from 1989-90 to 2003-04 using salary data from markerzone.com and statistical data from hockey-reference.com

Chapter 3

Do Sugar-Sweetened Beverage Taxes Improve Public Health for High School Aged Adolescents?

3.1 Introduction

As the prevalence of obesity has been rising worldwide in recent years, Pigouvian sin taxes on the purchase of sugar-sweetened beverages (SSBs) have become an increasingly popular policy instrument. Over 40 nations worldwide have enacted some form of SSB tax, as well as seven American cities[1] , with the vast majority of them being enacted with the intention of decreasing SSB consumption and therefore improving public health[2] . A broad literature has developed in response to these policies, assessing them from several different angles. A series of papers ([83], [285], [78], [261], [260], [211]) has addressed the question of whether or not the taxes are passed through to consumers and effectively raise the cost of consumption, generally finding that they are mostly if not entirely passed through. Others have analyzed the extent to which consumers cross local borders in order to avoid paying the taxes ([259], [211], [280], [50], [269]), with each of these papers finding evidence of cross-border shopping. Still others address whether overall SSB purchases decline in response to the taxes ([122], [258], [259], [261], [260], [211], [139], [165]), estimating own-price elasticities that range from -.71 to -3.87. While they all find evidence of reductions in sales, the wide range of estimates leaves uncertainty about what the overall public health impact of the taxes will be.

[1] Berkeley, Oakland, Albany, and San Francisco, California, as well as Seattle, Washington, Philadelphia, Pennsylvania, and Boulder, Colorado.

[2] One exception to this is Philadelphia, where the tax was sold to voters as a way to raise revenue for pre-K and community infrastructure.

Another strand of the literature looks at what people consume instead of SSBs when they become more expensive ([16], [139], [141], [123]), with mostly ambiguous results which suggest that people may still be making other unhealthy choices. Noting the importance of measuring changes in self-reported consumption in addition to purchasing behavior, others have used repeated cross-sectional surveys to assess whether people are drinking fewer SSBs. [209] and [129] both find substantial reductions in SSB consumption in Berkeley in response to the tax, while papers assessing the consumption response to the Philadelphia tax have found reductions in consumption among adults ([80]), short-term reductions among children ([327]) and a sustained reduction of soda consumption for high school students ([126]). While SSB taxes have been analyzed extensively from many important angles, because most of them have been enacted recently there remains a relative dearth of knowledge about whether they are accomplishing their stated goal of actually making paper healthier, particularly in the United States. [157] uses electronic medical records to show that the SSB tax in Mexico led to small reductions in BMI for girls, while [79] track self-reported SSB consumption and BMI among youths in Mauritius before and after an SSB tax, finding a consumption reduction for males but no discernible change in BMI.

In this paper I make two contributions to this literature using public-use data from the Youth Risk Behavioral Surveillance System (YRBSS) survey. First, I build on the findings of [126] that high school students in Philadelphia[3] substantially decreased their soda consumption after the implementation of the tax. I demonstrate that the reductions in Philadelphia are larger than placebo estimates for each of the untreated areas, suggesting that the change is larger than typical fluctuations in soda consumption, and that they appear to be largest among female and non-white respondents. Second, I find that average BMI decreased in Philadelphia, San Francisco and Oakland after implementation of the taxes, with the impact again appearing largest among non-white and female respondents. To my knowledge, these findings are the first in this literature which demonstrate that the SSB taxes in the United States led to sustained long-term decreases

[3] Philadelphia is the only treated city which has complete data on SSB consumption both before and after the tax went into effect.

in soda consumption as well as improvements in overall public health outcomes. Since the same groups that appear to be consuming fewer sodas are also the ones who see the largest improvements in BMI, this further suggests that the improvements are linked with the SSB tax.

Overall, in response to a question that asks "During the past 7 days, how many times did you drink a can, bottle, or glass of soda or pop, such as Coke, Pepsi, or Sprite?", I find a decline of 1.3 servings of sodas per week, which represents a 24% decline in consumption[4] . One major difference between this paper and several of the previous studies in this literature is that I look at self-reported consumption rather than scanner or other purchase data. Scanner data is useful because it is free of the measurement error and response-bias that can plague survey data, but it also has limitations and it is important that there is concordance among findings across the different types of data. One issue with most scanner data is that it tracks purchases made by a household head, and does not specify which family member is consuming which items. Data on purchases also has a shortcoming in that we do not know whether the items were ever consumed at all. Finally, scanner data which tracks purchases at a given store fails to account for potential tax avoidance unless stores on either side of a taxed border are compared. While the self-reported consumption data I use in this paper may be subject to concerns regarding measurement error and other issues associated with survey data, it is free from the concerns listed above associated with purchase data and thus complements the results already established in this literature.

It could still be the case, however, that substitution towards other unhealthy options is eroding the potential benefits of decreased consumption. For example, if adolescents are now eating candy bars and potato chips instead of soda, it seems unlikely that they would be getting much healthier. While I am unable to speak directly about what is being consumed instead of sodas, by measuring changes in health outcomes I am able to infer whether or not the decline in soda consumption is leading to improvements in health. I find that average BMI declined by .57 points across the treated cities after implementation of the tax, a reduction of .12 standard

[4] As this question asks specifically about soda and not other SSBs, I describe the reductions I find as changes in soda consumption while continuing to describe the taxes as SSB taxes and referring to SSBs when speaking more generally about sugary beverages and the harms they can cause.

deviations. While there are several reasons to be skeptical about the efficacy of SSB taxes, it does appear that they are leading to public health improvements among high school aged adolescents, which can have snowballing effects that could lead them to become healthier adults in the long run. [286] find that 80% of obese adolescents will be obese in adulthood, so preventing cases earlier in life could potentially lead to vast improvements down the road. While there are many potential ways to improve these taxes[5] , this paper represents early evidence that they can be a useful policy instrument in combating the childhood obesity epidemic in the United States.

The remainder of this paper is organized as follows. Section 2 outlines the motivation for taxing SSBs and gives background on the various SSB taxes implemented in the United States. Section 3 describes my identification strategy for estimating the public health impacts of SSB taxes and highlights the strengths and weaknesses of the YRBSS dataset which I use. Section 4 details my results on both consumption and health outcomes and compares these with placebo treatment estimates from all possible combinations of the control cities, while Section 5 concludes.

3.2 Background

3.2.1 Motivation for Taxing SSBs

Excessive sugar consumption has been linked with a number of serious health problems, including weight gain ([113], [124], [233]), heart disease and high blood pressure ([277], [304], [322], [182]), type II diabetes ([179], [268]), and overall mortality ([115]). Beyond the problems associated with the sugar consumption itself, SSBs are a particularly unhealthy delivery mechanism for the calorie-rich sugars because they are not satiating and leave consumers hungry ([250], [232]), suggesting that SSBs may be even worse than sugary solid foods, which at least stave off the hunger of the person consuming them.

While there is clearly potential to reduce harmful outcomes like diabetes, obesity and heart disease by reducing SSB consumption, the question remains as to whether this would be welfare-

[5] See [21] and [164] for discussion on the shortcomings of the SSB taxes in the United State and how they could be improved.

enhancing. If we view SSB drinkers as rational, informed consumers who are maximizing utility by enjoying SSBs even though they understand the long term health outcomes, one could argue that SSB taxes are not socially optimal. A growing body of literature, however, points to the fallacy of this assumption as SSB consumption creates both an externality in public health expenditures and an internality as consumers suffer a lack of self-control. [84] use genetic variation in weight as an IV to get around the endogeneity between obesity and other health outcomes and find that obesity causes annual medical expenses to rise by nearly $3,000 and that 88% of these costs are borne by third parties. [313] estimate that health costs are increased by approximately 1 cent per ounce of SSB consumed, indicating that a 1 cent per ounce tax on them would restore social efficiency.

Additionally, evidence in the medical literature has found that sugar has addictive properties. [317] find that the dopamine response triggered by sugar consumption resembles that of highly addictive drugs, while [31] find that sugar consumption produces the four behaviors that are medically associated with addiction (binging, withdrawal, craving and sensitization) in rats. Clearly, there is a scope for SSB taxes to reduce the social inefficiency generated by SSB consumption, although questions remain about the equity of doing so. For example, a large literature has documented the existence of 'food deserts'[6] , or areas which are usually lower-income and often populated by minority groups, where affordable nutritious food is locally unavailable. In such a setting, taxing one of the relatively few cheap sources of calories that are readily available would be counterproductive unless it produces a clear public health benefit which outweighs the social costs associated with raising taxes on lower-income populations.

3.2.2 SSB Taxes in the United States

The first SSB tax in the United States was implemented in Berkeley, California, when Measure D passed with 75% of the vote on election day in 2014[7] . The measure created a one cent per ounce tax on all beverages containing added sugar, with the exception of diet drinks, milk products,

[6] A full discussion of this literature is beyond the scope of this paper, but see [44], [183], [263], [216] and [19] for a more thorough review of this topic.

[7] The Navajo nation also passed an SSB tax in 2013 but this paper will focus on taxes implemented by local U.S. governments.

100% juice, baby formula, alcohol, and drinks taken for medical reasons. The tax went into effect on January 1, 2015 and was collected directly from distributors, with very small retail stores being exempted. Over the next three years, similar measures were passed in six other cities. Cook County, Illinois, which includes Chicago, also approved an SSB tax on November, 2016, with the tax going into effect on August 2, 2017. Because of a multi-million dollar ad campaign from beverage companies and growing public pressure, lawmakers voted 15-1 to repeal the tax on October 10, 2017, just two months after it went into effect. The important details of each tax which remains in place are included in Appendix Table C.1. The main sources of heterogeneity are in the timing of approval and implementation, the magnitude of the tax per ounce of liquid, and whether or not diet drinks are included in the tax. The majority of the taxes are one cent per ounce except for Philadelphia tax which is 1.5 cents per ounce, Seattle (1.75 cents per ounce), and Boulder (2 cents per ounce). Philadelphia's tax is the only one which does not grant an exception to diet drinks. For the purposes of this paper, the important dates with regard to timing are that Philadelphia approved the tax in July, 2016 with the tax taking effect on January 1, 2017 and that both Oakland and San Francisco approved their taxes in November, 2016 with the taxes going into effect on July 1, 2017 and January 1, 2018, respectively.

The biggest difference between the SSB taxes in the US and those in place in the rest of the world is the geographic coverage of the tax. More than 40 countries have enacted nationwide SSB taxes, while only the United States and Spain have local or regional taxes. Because crossing an international border to avoid the tax is prohibitively costly in most cases, tax avoidance is likely only a concern for these localized versions of the tax. This suggests that the estimates on the impact of these taxes likely understates what the effect would be if they were put in place at the state or national level. Importantly, all SSB taxes in the U.S. and the vast majority of the taxes throughout the world are based on volume of liquid in the beverage as opposed to grams of sugar added, meaning the harshness of the tax is not directly proportional to the harm being caused. For example, a 20 ounce Mountain Dew with 77 grams of sugar is taxed the same (20 cents in most cases) as a 20 ounce Gatorade with 36 grams of sugar. Because of this, [164] estimate that a tax based on the sugar content of the drink could improve the public health benefits by 30%, again suggesting that any improvements that occur because of the current taxes understate the potential impact of a better designed version of the tax.

3.3 Methodology and Data

3.3.1 Data - YRBSS

The Youth Risk Behavioral Surveillance System (YRBSS) is a survey conducted semiannually by the CDC, which interviews tens of thousands of high school aged participants each cycle. It asks questions about a range of risky behaviors, including the soda consumption question mentioned above. It also asks students for their height and weight, which are used to calculate BMI. There are reasons to be skeptical of self-reported height and weight of high school students with regards to reporting accuracy and precision. One concern is that students in SSB tax treated districts might become more self-conscious of their weight after the public debate over the tax and may therefore under-report. Because of this, I prefer to compare outcomes in treated districts with those in other areas where a soda tax was publicly debated and nearly passed.

A total of 39 sites, including 35 separate school districts with each of the five boroughs of New York City reported separately, have opted into implementing the YRBSS survey, but most do not report data in each year. For example, Seattle, which implemented an SSB tax in 2018, provides data for 2013 and 2019 but did not participate in the YRBSS in 2015 or 2017, making it difficult to assess whether any changes between 2013 and 2019 occurred in response to the tax or had already shown up before it was implemented. For this analysis, I include all school districts which have data for each year from at least 2013-2019. This includes three districts where an SSB tax was implemented during this period (Philadelphia, San Francisco and Oakland) and 11 potential control sites: the five boroughs of New York City (Queens, Brooklyn, The Bronx, Manhattan, Staten Island), Los Angeles, San Diego, Jacksonville, Orlando, Fort Lauderdale and Palm Beach. This leaves me with approximately 5,000-6,000 observations per year in treated districts and 20,000-

25,000 observations per year across the control districts. For each of the treated sites, the boundaries of the school district are the same as the city limits, meaning that all of the students in the district should live in an area that is covered by the tax. Within the treated districts, however, there are data limitations. First, while San Francisco and Oakland both track BMI throughout the sample, they do not include the question about SSB consumption that exists in the other 12 districts in both the pre and post period. Therefore, all treatment effects estimated on SSB consumption only include Philadelphia as being treated. Additionally, while the other cities all have data going back further than 2015, Oakland does not, so all the specifications using Oakland are limited in the ability to include pre-treatment lags. My general strategy is to use the full treatment group (Philadelphia, San Francisco and Oakland) in order to understand the magnitude and significance of the treatment effect. I then use a modified treatment group of just Philadelphia and San Francisco, which allows for the inclusion of pre-treatment lags, in order to evaluate the validity of the equal counterfactual trends assumptions inherent to these specifications.

Determining treatment status is also slightly complicated because of the timing of data collection and the staggered implementation of the policy. Though all three treated cities approved the taxes in 2016, the policy went into effect in Philadelphia on January 1, 2017, in Oakland on July 1, 2017 and in San Francisco on January 1, 2018. Since the YRBSS survey is conducted in the spring semester, the 2017 data is treated for Philadelphia, but untreated for Oakland and San Francisco. Since both Oakland and San Francisco had approved SSB taxes before data was collected in 2017, however, it is possible that people had already began altering their consumption in anticipation of the tax. In all event-study graphs, I display the treatment line just before the 2017 data, though for the above-mentioned reasons 2017 should only be considered partially treated. Also, since any BMI response will take time to accumulate, even the Philadelphia data is mostly untreated in 2017 as the respondents had only been interviewed a few months after the tax went into effect. Therefore, while we can expect SSB consumption reductions to take place relatively concurrently with the implementation of the tax, there should be a lag between the consumption change and any subsequent public health response, if in fact the tax has a causal impact on these outcomes.

I use each of the 11 possible control districts to construct three different control groups I will use to estimate treatment effects. With limited data there is no perfect control group, but I select controls based on geographic proximity, economic and demographic similarity, and whether there has been consideration of implementing a soda tax in the district. For a detailed description of how the three groups were selected, see the appendix. Importantly, however, I show results for all three control groups in all specifications which means I use all available observations, and will also estimate synthetic control specifications which do not rely on an arbitrary selection of control sites. The baseline group, C1, includes Brooklyn, Queens, the Bronx, and Los Angeles. C2 includes each of these sites plus Staten Island, Manhattan and San Diego. Finally, C3 adds the four Florida districts of Jacksonville, Orlando, Palm Beach and Fort Lauderdale. Summary statistics on a range of outcomes possibly related to SSB consumption are reported in Table 1 for 2015, the last fully pretreated year , which confirms that C1 is the most similar group to the treated group. For 11 of the 14 variables, equality of means cannot be rejected between the treated group and C1. Only the % of the sample which is Hispanic, the % which is Asian and the age of the respondent can be rejected at 5% for C1. Importantly, both race/ethnicity and age are controlled for in every specification in this paper. For C2, mean equality for six of the 14 variables can be rejected at 5%, while for C3 eight of the 14 can be rejected at 5%, suggesting that the choice of control groups is reasonable. Again, however, it is important to note that I will present estimates for all three groups which means I include estimates on all available observations.

3.3.2 Empirical Strategy

My primary approach for estimating the public health effects of SSB taxes is an event-study design that compares changes in consumption and health outcomes in SSB taxed cities with trends in the same outcomes for cities which had similar public health patterns in the pre-period but which did not implement an SSB tax. The critical assumption underlying this design is that absent the intervention, the treated and control cities would have followed similar trends and that any difference in their trajectories is due to the policy itself and not due to some other factor that changed between the groups around the time of the intervention. My results will then be based on models of the following form:

$$H_{ict} = \sum_{k=-4}^{4} \theta_k SSBTax_{ct,t+k} + \beta_3 X_{ict} + \gamma_c + \delta_t + \epsilon_{ict} \tag{3.1}$$

where H_{ict} is the outcome of interest for individual i in city c at time t, $SSBTax_{ct}$ is an indicator for whether city c implemented an SSB tax during the sample period. X_{ict} is a vector of covariates that are known to influence H, including height and age, as well as indicators for race, gender and grade. Allowing the treatment effect to vary across periods is important because I expect the treatment effect to be cumulative. Even though the treatment itself is constant once switched on, the intended effect is to reduce sugar consumption. For example, if a treated individual responds by having one less SSB per week, and this behavior change remains while the tax is in place, the magnitude of the effect on the individuals health will likely grow over time. Additionally, because of the importance of understanding how the policy impacts different demographic subgroups, I will re-estimate equation (1) on different groups to measure heterogeneous treatment effects across race, ethnicity and gender. If there are certain subgroups with more inelastic demand for SSB products, perhaps because they have high enough income that the SSB taxes are not salient to their day-to-day lives, they may continue consuming them at previous rates and may not experience improvements in health outcomes. Heterogeneous treatment indicators will pick up these differences and allow me to evaluate precisely which subgroups are seeing the largest improvements in health outcomes. This will both provide support for a causal effect of the taxes, as we would expect the groups which respond to the tax by changing their consumption the most to also be the groups which see the largest improvements in health outcomes, and it will also help us understand the distributional consequences of the policy. I estimate the model using weighted least squares, using the survey weights to allow individuals who are more representative of each district to be more influential in the regressions, though the results from using OLS are qualitatively similar. Because of the small number of districts included in the sample, I use robust standard errors instead of clustering, though my main results are robust to clustering at the district level.

In each specification 2015 will serve as the reference year, as it was the last year of data before any of the taxes assessed in this paper were levied. Mechanically, this means that the indicator for 2015 is left out of the regression, meaning that the difference between the treated and control groups for 2015 is absorbed by the fixed effects. Each of the k parameters thus measures how the outcome evolved differently in the treated group versus the control group compared to the difference that existed in 2015. For this reason, the 2015 estimate will always be zero and will not have confidence intervals attached. If the equal counterfactual assumption holds, we should expect to see a 2013 coefficient that is close to zero and insignificant. I estimate the model using weighted least squares, using the survey weights to allow individuals who are more representative of each district to be more influential in the regressions, though the results from using OLS are qualitatively similar. Because of the small number of districts included in the sample, I use robust standard errors instead of clustering, though my main results are robust to clustering at the district level.

In addition to the event-study design, I also estimate the effect of SSB taxes on adolescent BMI using the synthetic control method (SCM) of [6] and [5][8] . SCM provides compelling evidence for the equal counterfactual trends assumption in comparative case studies by creating a control group made up of a weighted average of the potential controls, with the weights being chosen to minimize the pretreated difference in observables between the two groups. This means that SCM will assign non-negative weight to each of the 11 potential control sites in the YRBSS data in order to most closely match the trends in the treated group before the taxes went into effect. The idea is that absent the treatment itself, the synthetic control provides a good counterfactual for

[8] This particular setting, with limited data in the pre-period, is not ideal for the synthetic control method. Abadie (2021) points out the importance of having many years of data prior to the intervention for the treatment and control group to be matched upon. Otherwise, it is unclear whether the control group that is created matches the treatment group closely on observables or if it simply has the same idiosyncratic variation in the periods upon which it is being matched. In such cases, it is often easy to find very close matches on pre-treatment outcomes but then severe variation in outcomes after treatment, because the groups were not actually good comparisons. While this setting is not ideal for the use of the synthetic control method, I include it here to demonstrate that it was considered and implemented and that if anything, it offers evidence in support of the results in the rest of the paper.

what would have happened in the treated group, and any sizeable differences that show up can reasonably be attributed to the treatment. I estimate two versions of the SCM on both the full and modified treatment groups, for a total of four specifications. In the first version, I first run a version of equation (1) excluding the treatment indicators on all of the untreated observations in order to estimate the effect of the demographic and individual controls as well as year and district fixed effects. I then residualize this regression so that each observation contains an estimate of how much higher or lower their BMI is than would be predicted by all of their observable characteristics. I then run a synthetic control which matches the treated group based on average residual BMI from 2013-2015 (or just 2015 for the full treated group). This specification has the benefit of matching based on trends in BMI that is unexplained by observables, which is of primary interest in this study, but it is also complicated so I also run a version which simply matches on average pretreated BMI, without residualizing. While imperfect for this setting because of limited pretreated data, this method has the benefit of not relying on an arbitrary choice of control groups. To conduct inference on these specifications, I follow [5] and divide the mean squared error in the treated period by the mean squared error in the pre-period and compare it to placebo estimates of the 55 combinations of two of the 11 control sites for the modified treatment group and 165 combinations of three of the 11 for the full treatment group.

Because the equal counterfactual trends assumption that underlies these event-study specifications is vital to the validity of my results, I argue for this assumption in three different ways. First, I demonstrate that there is balance between the treated and control groups across a number of different measures of health, consumption and activity choices in 2015, before any of the policies went into place. Second, for each specification I also reestimate the model by dropping Oakland (which only has data going back to 2015) from the treated group, which allows me to include pre-treatment lags. These lagged coefficients demonstrate that the treated and control groups evolved similarly in the years before the intervention and only diverged after the SSB taxes went into effect. Third, because the event-study regressions are effectively comparing residual means across different groups and do little to assess how the distributions of the outcomes compare across time, I also include pre and post treatment kernel density plots which compare the BMI distributions for both the treated and control groups both in 2015 and 2019. If the equal counterfactual trends assumption holds, then the two groups should have similar distributions and not just similar means before the intervention.

3.4 Results

3.4.1 Results - SSB Consumption

First, I consider whether the SSB tax impacted consumption for high school aged adolescents in Philadelphia (the only treated city with SSB consumption data). Figure 3.1 displays coefficient estimates for SSB consumption compared to each of the three potential control groups. Regardless of which group is used, the interpretation is similar. There is a slight but statistically insignificant uptick in consumption in Philadelphia relative to the controls from 2013 to 2015 (p-values of .46, .63 and .70, respectively), followed by a large and significant decline in response to the implementation of the SSB tax in 2017, which increases in magnitude slightly in 2019 (p-values of .002, .000, and .001). By 2019, the treatment effects represent a reduction of over one full serving of SSBs per week, which is a decline of between 21-26% of the 2015 mean for Philadelphia high school students.[9] It does not appear to matter which control group is used for comparison, SSB consumption in Philadelphia appears to have significantly declined relative to each of them in response to the tax.

In order to determine how this decline compares to changes in each of the other cities in the sample, Figure 3.2 compares the sum of the 2017 and 2019 coefficients compared with C3 to those of placebo estimates, where equation (1) is estimated under the assumption that each untreated city had actually implemented an SSB tax instead of Philadelphia. The decline which took place in Philadelphia is by far the largest, nearly double that of the next largest estimate in Fort Lauderdale. The change in Philadelphia is also substantially larger in magnitude than the largest increase, which occurred in Los Angeles, suggesting that the sharp decline which occurred in Philadelphia is in response to the SSB tax and is not simply due to normal fluctuations in

[9] The precise treatment effects are 1.33 servings for C1, 1.24 servings for C2 and 1.07 servings for C3.

consumption which happen to line up with the implementation of the tax.

Next, I turn to look for heteregeneous treatment effects across gender and race/ethnicity. We might expect lower income groups to be more budget-constrained and therefore more sensitive to the price increases that are induced by SSB taxes. Since one of the main objections of SSB taxes is that they hit low income and non-white citizens the hardest, it is important to determine whether the consumption habits of these groups are impacted by the taxes. If they demand SSBs inelastically, then they will continue to consume at the same rates as before and the only outcome of the policy will be that they pay more in taxes. In this case, the policy would be highly regressive and provide no public health benefits to these communities. On the other hand, if groups with lower average income respond to the tax by changing their consumption by more than other groups, it would actually reduce the regressivity of the tax overall and would likely lead to targeted public health benefits that improve the well-being of those who are most likely to be impacted by issues related to obesity. Estimating heterogeneous treatment effects on both SSB consumption and health outcomes separately has the added benefit of providing a falsification test on the results. If reductions in BMI are caused by the SSB taxes, then we should expect the groups which display the largest consumption responses to the taxes to be the same groups which experience the largest public health benefits. If this is not the case, then there is likely another mechanism at play which is driving the improvements.

Figure 3.3 displays SSB treatment estimates separately for male and female students compared with each of the control groups. For females there appears to be a real treatment effect, and the equal counterfactual trends assumption appears to be reasonable, with a p-values of .94, .79, and .77 on the estimate of the change from 2013 to 2015. After the tax goes into effect in 2017 there is an immediate decline of .5-.8 servings per week, but the effect grows to 1.25-1.45 in 2019 (p-values = .002, .003 and .004 for C1-C3, respectively). For males on the other hand, there is a substantial increase from 2013 to 2015, and then a subsequent reduction after the implementation of the tax. It is somewhat unclear whether this reduction represents a real treatment effect or is mostly due to a reversion to the mean. 2019 consumption for the male group was only .6 servings less than 2013 (p-value = .302). A test for equality of the change in males and females from 2015-2019 yields a p-value of .6, but when looking at the change from 2013-2019, females reduced consumption by .97 more servings, with a p-value of .167. This suggests that we may expect to find larger treatment effects in BMI for females, if the reductions are due to the taxes. At the very least, we would not expect to see larger declines for males.

Next, Figure 3.4 displays estimates of the effect of SSB taxes on consumption, broken out by race/ethnicity. The most compelling treatment effect shows up in the Black population. There is virtually no change in consumption between 2013-2015 (p-values of .82, .74 and .95 for C1-C3), with consumption declining by 1.05-1.45 servings per week in each specification, with p-values of .018, .038 and .056 on the 2019 coefficients. There appear to be declines in response to the tax for the Hispanic subgroup as well, though there is also movement from 2013 to 2015. The coefficients on the 2019 treatment effect are -1.67 (p=.014), -1.54 (p=.021) and -1.05 (p=.056). On the other hand, both the White and Asian subgroups mostly appear to continue on trends they were already on. The White group decreases consumption from 2013-2017 in all three specifications before leveling off and increasing somewhat by 2019. While the equal counterfactual trends assumption does not look credible on the Asian subgroup, there is some evidence of a decline in 2019 after increasing steadily from 2013-2017. As the sample sizes have gotten smaller, confidence intervals have gotten bigger, so I am not able to rule out equality of the change in consumption from 2015-2019 for any of the groups, but the fact that the Black and Hispanic subgroups show a relatively clear response to treatment while the White group does not suggests that lower income groups may respond more elastically in response to SSB taxes than higher income groups, which would reduce the regressivity of the tax. Clearly, SSB consumption declined in Philadelphia after the tax went into effect, and the impact appears to be greatest in females, as well as non-white respondents. I now switch gears in order to look at the public health outcomes of SSB taxes for high school students.

3.4.2 Do SSB Taxes Lead to Lower BMI?

Figure 3.5 displays estimates from equation (1) with BMI replacing SSB consumption on the left-hand side. As a reminder, while Philadelphia is the only treated city with SSB consumption data, San Francisco and Oakland both have data on both BMI and obesity, though Oakland's data only goes back to 2015. Therefore, my strategy is to include estimates for all three cities from 2015 onward, and then present a modified treatment group of just Philadelphia and San Francisco in order to estimate pre-treatment lags, allowing me to assess the equal counterfactual trends assumption that equation (1) relies upon. While the decline in consumption associated with the implementation of the tax occurred immediately after the tax went into effect, changes to BMI and obesity take longer to show up, which is precisely what one would expect if the tax led to a consistent reduction in SSB consumption over a long period of time.

Compared with C1, both the full and modified treatment groups show large, statistically significant reductions in BMI in response to the SSB tax. The full treated group has a decline of .57 points (p=.002) while the modified group has a decline of .56 points (p=.006), an approximately .12 standard deviation reduction. The modified treatment group also provides evidence for the equal counterfactual trends assumption, as there is virtually no movement from 2013 to 2015, where BMI only changes by .01 points (p=.96). The story is similar for C2, though the results are smaller in magnitude. BMI in the full treated group decreases by .35 points (p=.046), compared with .34 points (p=.079) for the modified group. Finally, when the Florida districts are added to the control group, the results shrink once again, with treatment effects of .26 (p=.116) and .25 (p=.175). When the treated group is compared with the most similar districts, the results are large and significant, though they decrease in both magnitude and significance when more districts are added.

Figure 3.6 shows the BMI estimates broken out by gender. Since females displayed a more compelling decrease in consumption in response to the tax than males, we might expect to see bigger improvements in BMI for them as well. Compared with C1, BMI for females in the full and modified treatment groups decreased by .68 (p=.007) and .71 (p=.010), respectively. BMI also appears to have gone down in treated males, though by smaller amounts of .46 (p=.095) and .41 (p=.179) points. I cannot, however, rule out that the two groups decreased BMI by the same amount in either specification. The story is similar in C2, where the effect appears larger in females and is still significant at .1, with a decrease of .42 (p=.07) for the full treated group and .45 (p=.08) for the modified group. The corresponding estimates for males are .27 (p=.296) and .22 (p=.45), though I cannot reject equality of the coefficients across genders. Compared with C3, females still show borderline significant treatment effects of .41 (p=.066) and .44 (p=.076), while males show almost no change at all, with 2019 coefficients of -.11 (p=.64) and .06 (p=.83). For females, all six of the estimates are significant at .1, while this is true of only one of the six for males.

While the conditional mean of BMI went down relative to the control groups for females, the above figures do very little to shed light on which parts of the distribution these BMI reductions are coming from. To gain insight into this, Figure 3.7 displays pre and post treatment BMI kernel density plots for both males and females in both the treated and untreated groups against C1. The left side of the figure compares BMI distributions in 2015, where no taxes were in effect and none had even been voted on in the eventual treated cities. For females, the two distributions look virtually identical, with means of 23.18 (treated) and 23.25 (control), respectively, a difference of 0.01 standard deviations. There is slightly more mass in the treated group around the mode, but aside from that, the two are very similar. On the other hand, for males the treated group has less mass around the 25-30 portion of the distribution and more around the mode of 21. This could by why there appears to be little or no treatment effect for males, as treated males appear to have a healthier distribution before the tax even went into effect. Alternatively, the right side of Figure 3.7 shows the same distributions in 2019. While BMI was rising across the U.S. in this period, the mean BMI for the female treated group actually fell to 23.14. In 2019, the female control group now has less mass around the mode and more in the right half, with more mass than the female treated group for BMI scores of 26 to 32, which are all in the overweight/obese portion of the distribution[10] . This suggests that the relative gains which accrued to the treatment group are happening in the

[10] For adults, BMI's over 25 are considered overweight, while BMI's over 30 are considered obese. The calculations

upper half of the distribution, which is precisely the intent of the policy. For males, on the other hand, the two distributions actually look more similar in 2019 than they did in 2015.

Figures 3.8 and 3.9 explore heterogeneous treatment effects by race/ethnicity, estimating equation (1) on the Black, Hispanic, White and Asian subgroups separately using both the full and modified treatment groups. Earlier, the Black and Hispanic subgroups both displayed strong consumption declines in response to treatment, while the Asian group showed a mild decline and the White subgroup showed virtually no change. Therefore, if any decreases in BMI show up, we would expect them to occur mainly in the non-White populations if they are being caused by decreased SSB consumption in response to the taxes. As with the consumption response, the confidence intervals are growing larger as the respective samples are growing smaller, but the same general trends emerge. Treatment effects appear to show up in the Black, Hispanic, and Asian groups, while the White group appears to be mostly continuing on the trends they were already on. In each of the six specifications, the Black subgroup displays a economically meaningful decline of between .39-.57 points, though the lack of precision with smaller samples causes none of these estimates to be significant (p-values range from .168 to .358). While there appear to be some minor treatment effects for the Black subgroup overall, there is a slight increase from 2013 to 2015 which casts doubt on the equal counterfactual trends assumption. For the Hispanic group, the estimates are slightly larger (estimates range from .40-.71) and more significant (p-values range from .03 to .255), with the equal counterfactual trends assumption appearing to be much more reasonable. The treated Asian subgroup also appears to have experienced a decline by 2019, with the estimates being relatively similar to those for the Black and Hispanic groups, ranging from .33-.59 with p-values ranging from .15-.32. Finally, the White subgroup displays small and highly insignificant treatment effects ranging from .08-.27 points, with the smallest p-value being .69.

Overall, the results in this section demonstrate that there were large and statistically significant declines in both SSB consumption and BMI in the treated cities, with suggestive evidence that these improvements are concentrated among females and non-white respondents. I now explore the

for adolescents are slightly more complicated, but are highly correlated with the measures for adults robustness of the results in this section by implementing a synthetic control.

3.4.3 Synthetic Control

Figure 3.10 displays estimates of the four SCM specifications described above. In each case there is a close match between the treated group and its synthetic control until after the first tax goes into effect, then there is a sharp divergence by 2019, with the treated group having substantially lower BMI, both in the raw data and when residualized after controlling for demographic and individual characteristics as well as secular trends and time-invariant differences between areas. In the top two specifications, residualized BMI decreases by .23 BMI points (p=.37, sd=.046) for the modified treatment group and .27 (.23, sd=.054) for the full treatment group. In the bottom two specifications the gap between the groups is .44 points (p=.036, sd=.089) for the modified treatment group and .55 points (p=.066, sd=.111) for the full treatment group.

3.5 Conclusion

There are several reasons why the predicted effect of SSB taxes on public health outcomes is unclear. In order for the tax to make any difference, the majority of it must be passed on to the consumer in the form of higher prices and the consumer must not be able to easily avoid it, something which is not always true of local taxes like the ones found in the U.S. cities that have enacted these measures. Consumers must also respond to the higher prices by consuming less, and must also not substitute their SSB consumption for other untaxed yet equally unhealthy options. Even if all these things take place, it's also important that they occur on a large enough scale so that the impacted consumers actually get healthier as a result. Because of these ambiguities, it is important to measure the impact of SSB taxes directly on public health outcomes.

This paper does this by comparing SSB consumption and average BMI in treated U.S. school districts to similar untreated ones. Using public-access data from the Youth Risk Behavioral Surveillance System survey (YRBSS), I find substantial reductions in both categories for high school students in response to SSB tax treatment. While only suggestive, the fact that both the SSB consumption and BMI reduction treatment effects appear to be strongest in females and non-White respondents has two important implications. First, it suggests that the two effects are linked, and that the reduced consumption in SSB's is what is causing the subsequent BMI reductions. In addition, the fact that the BMI response lags behind the consumption response also suggests that it is part of the cumulative effect of reduced SSB consumption. Second, because Black and Hispanic Americans have lower incomes on average, if they have larger consumption responses than the White subgroup, this would suggest that perhaps SSB taxes are not as regressive as they were originally thought to be.

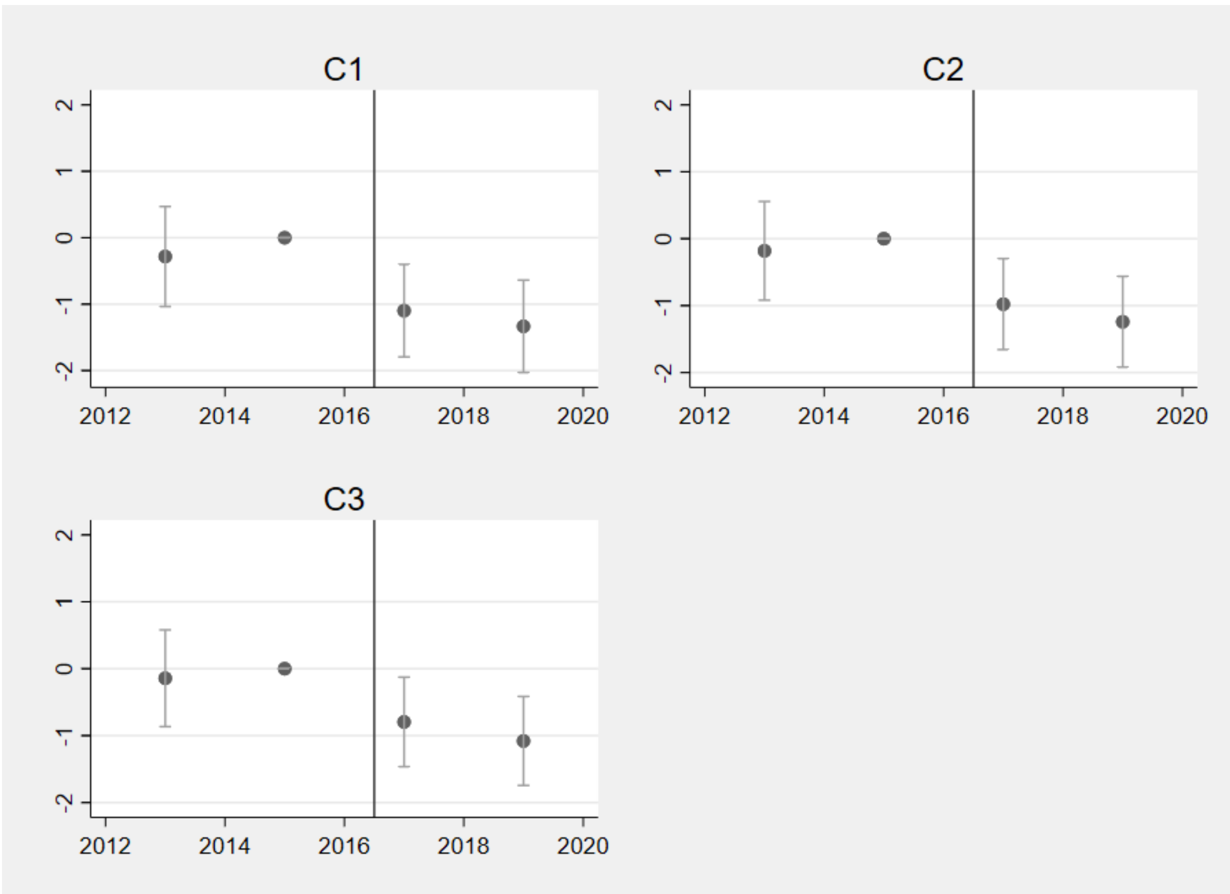
While many of the results in this paper are statistically insignificant, it is telling that each of the 40 estimates on BMI presented show relative declines in BMI relative to the control groups. Going forward, it will be important to continue to track these outcomes to see if they persist or revert to the mean, and also whether other adverse health outcomes related to sugar consumption like type II diabetes and high blood sugar are reduced as well.

Table 3.1: Balance on Observables 2015

Variable	Treatment	C1	p	C2	p	C3	p
SSBs per week	3.94	3.92	0.88	3.81	0.24	4.01	0.56
BMI	23.22	23.30	0.37	23.22	0.98	23.13	0.25
% Obese	0.13	0.14	0.23	0.13	0.43	0.13	0.72
% Black	0.22	0.21	0.24	0.18	0.00	0.22	0.16
% White	0.08	0.08	0.34	0.13	0.00	0.19	0.00
% Hispanic	0.32	0.49	0.00	0.47	0.00	0.39	0.00
% Asian	0.25	0.11	0.00	0.11	0.00	0.08	0.00
% Female	0.50	0.50	0.76	0.50	0.23	0.50	0.59
Cigs per day	0.23	0.23	0.94	0.24	0.78	0.27	0.27
Age	15.97	15.74	0.00	15.72	0.00	15.81	0.00
Height (m)	1.67	1.67	0.92	1.67	0.09	1.68	0.00
Average sleep	6.53	6.54	0.67	6.53	0.84	6.37	0.00
Hrs TV per day	1.73	1.68	0.12	1.60	0.00	1.62	0.00
Hrs video games per day	2.32	2.33	0.90	2.32	0.98	2.21	0.00

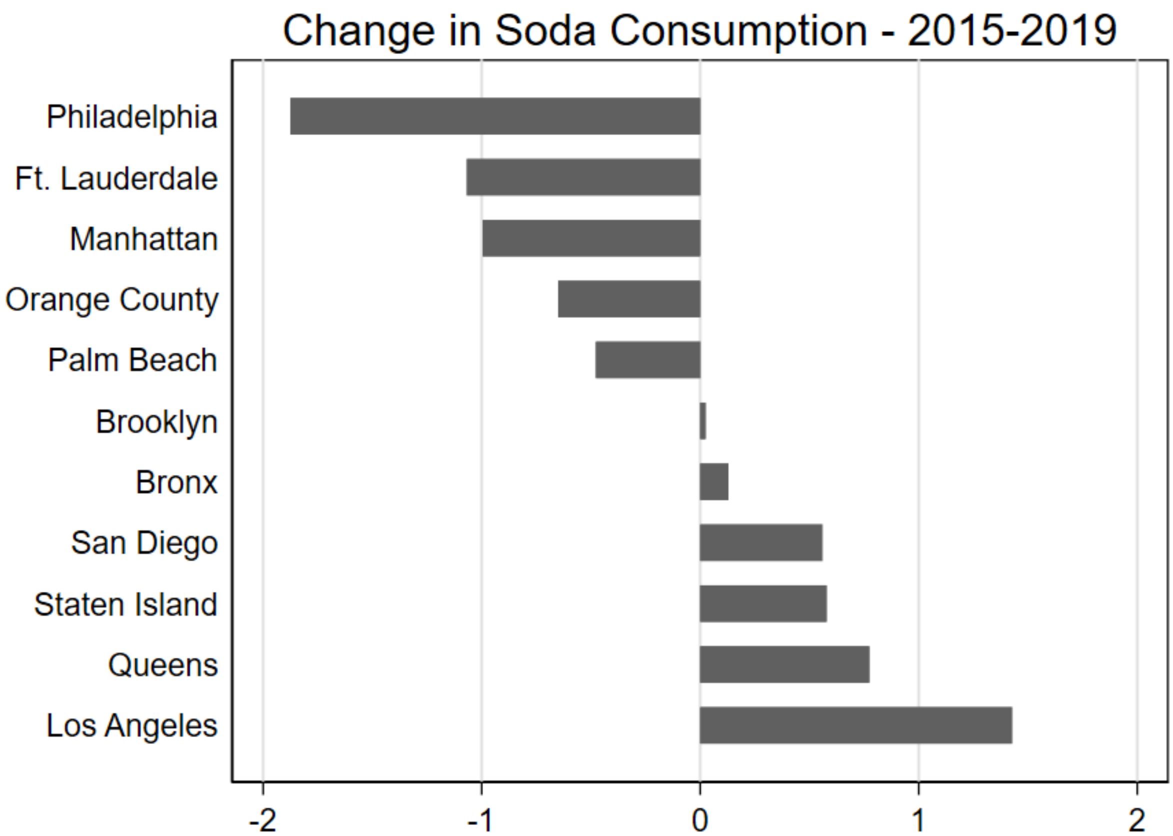
This table compares 2015 pre-treatment means for 14 variables between the treated group (Philadelphia, Oakland and San Francisco) against three potential control groups. C1 includes Queens, Brooklyn, the Bronx, and Los Angeles. C2 includes C1 plus Manhattan, Staten Island and San Diego, while C3 also includes Orlando, Jacksonville, Fort Lauderdale and Palm Beach. The column following each control group gives the p-value for a test of equality between that group and the treatment group. For C1, equality cannot be rejected for 11 of the 14 variables. For C2 and C3, equality cannot be rejected for only 8 and 6 of the 17 variables, respectively.

Figure 3.1: SSB Consumption Impact



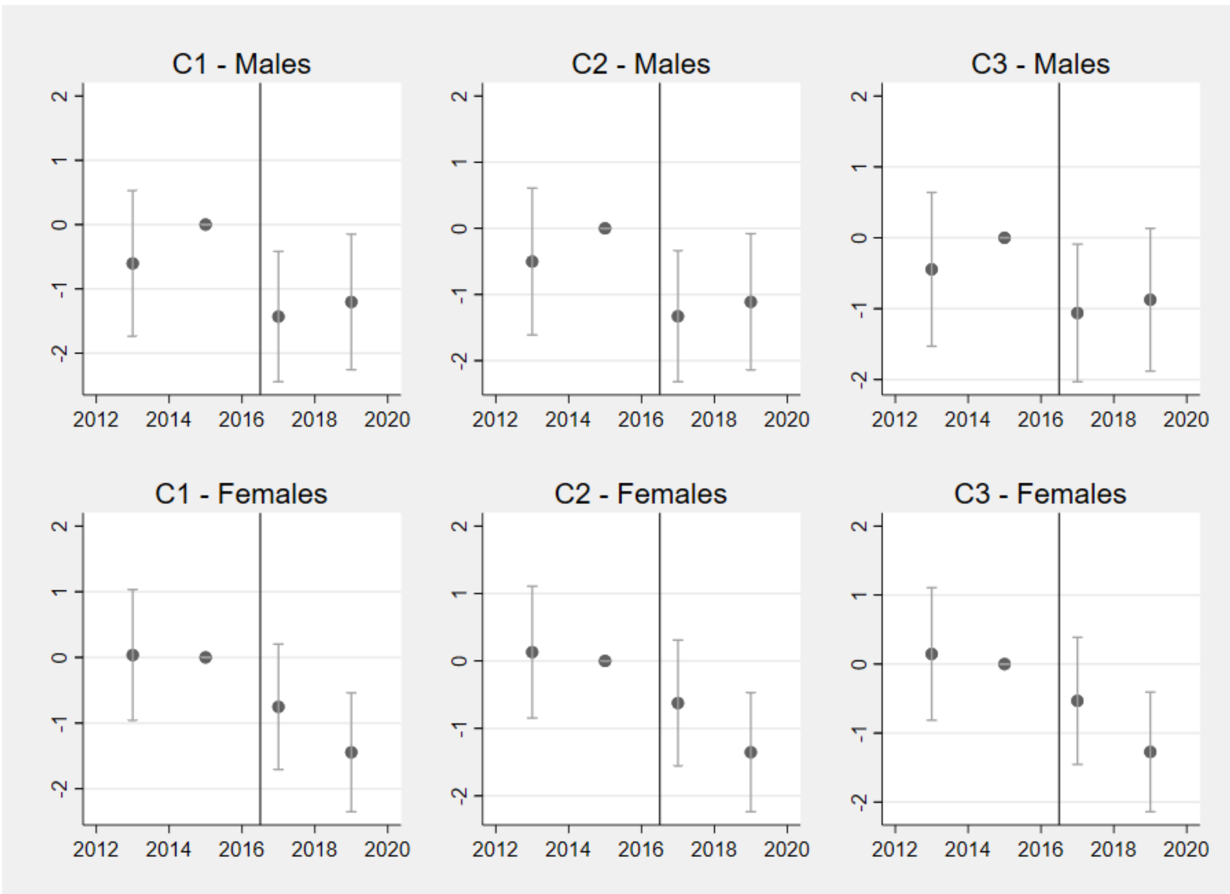
This figure displays the event-study treatment effect on SSB consumption for Philadelphia against each of the potential control groups. In 2017, there is an immediate decline in consumption (p-values of .002, .012, and .020, respectively), which increases in magnitude and significance in 2019 (p-values = .000, .001 and .001).

Figure 3.2: Treatment Effect - Philadelphia vs. Placebos



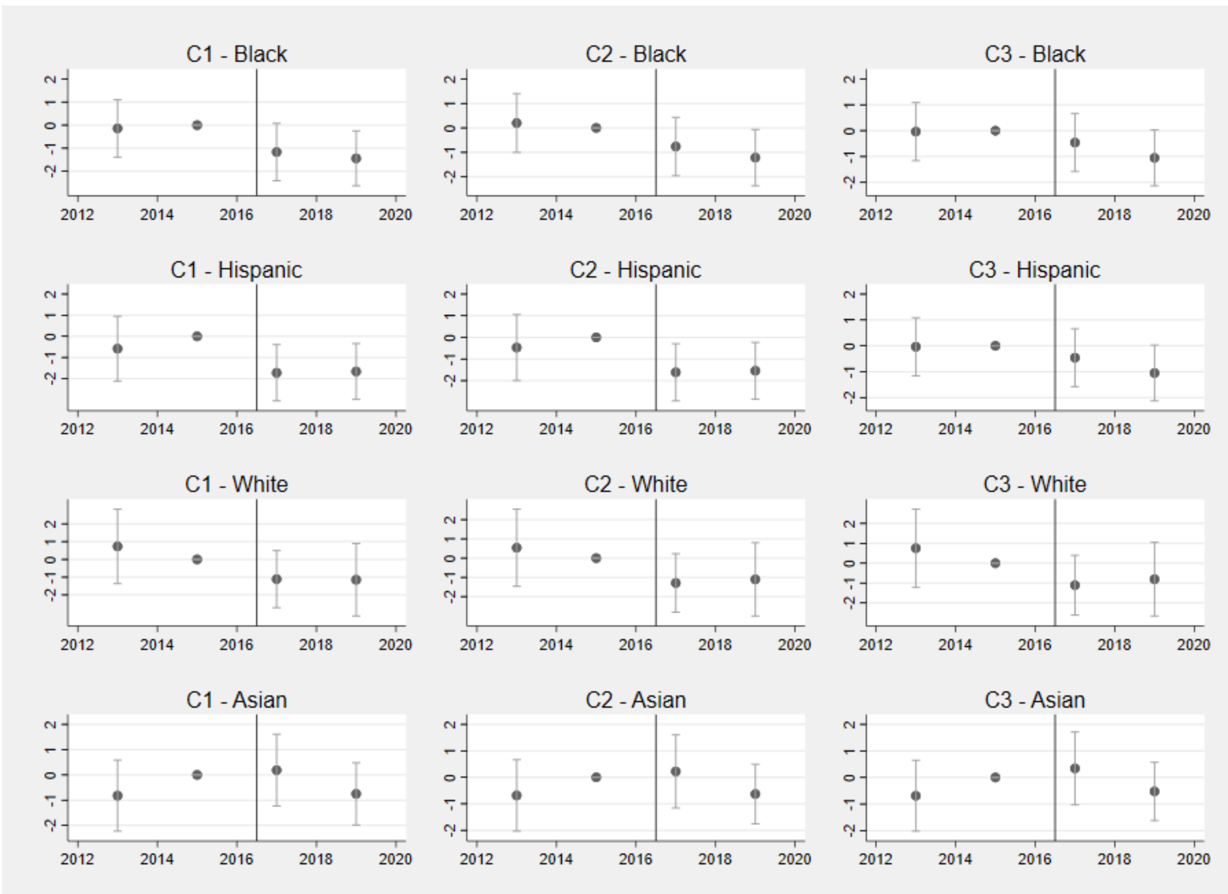
This figure compares the actual treatment effect (defined as the sum of the 2017 and 2019 coefficients in equation (1)) in Philadelphia with placebo treatments which iteratively assume that each of the untreated cities had implemented the tax instead of Philadelphia. The actual treatment effect in Philadelphia is by far the largest in magnitude and is nearly double the size of the next largest reduction.

Figure 3.3: SSB Treatment - Males vs. Females



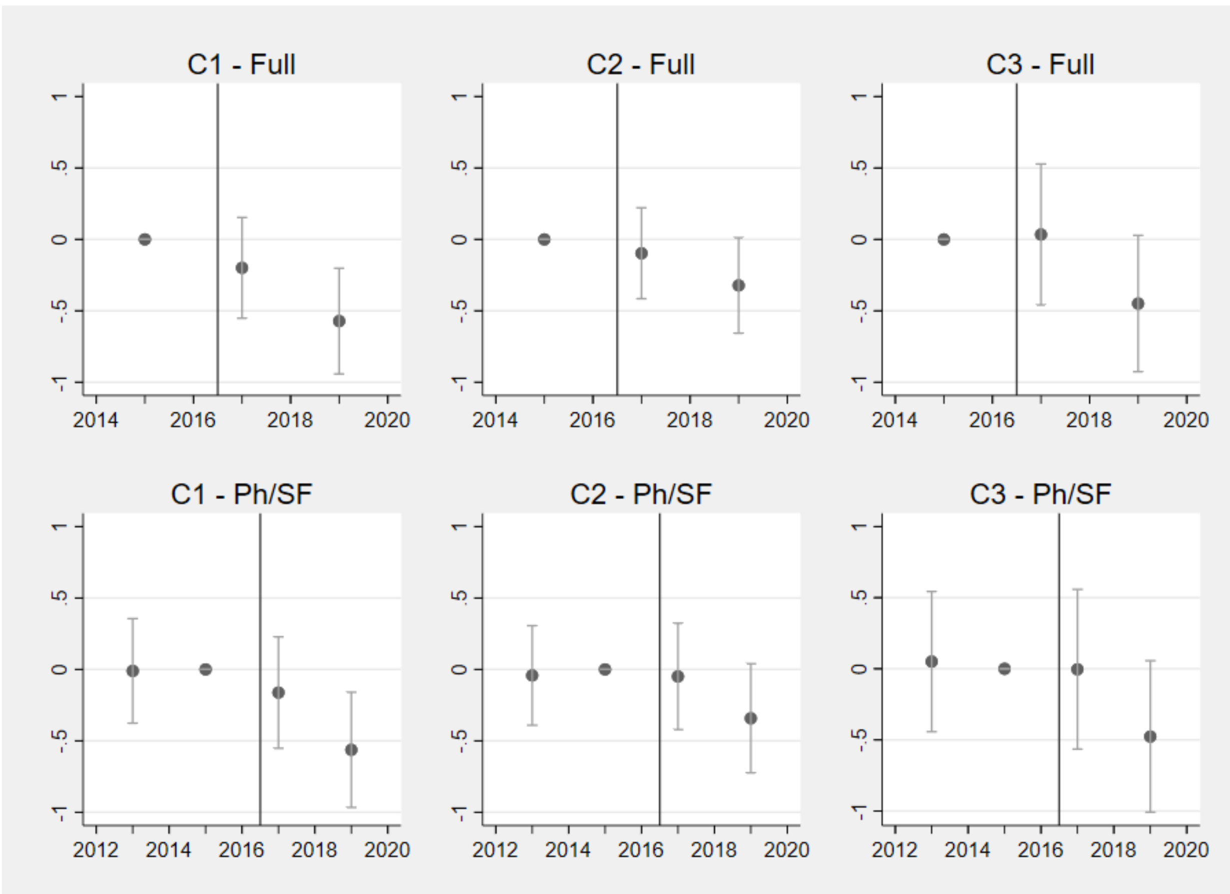
This figure compares the consumption response of male and female high school students in Philadelphia to the SSB tax that went into effect on January 1, 2017. While both groups appear to have decreased consumption between 2015 and 2019, the equal counterfactual trends assumption appears to be much more reasonable for females.

Figure 3.4: SSB Treatment By Race



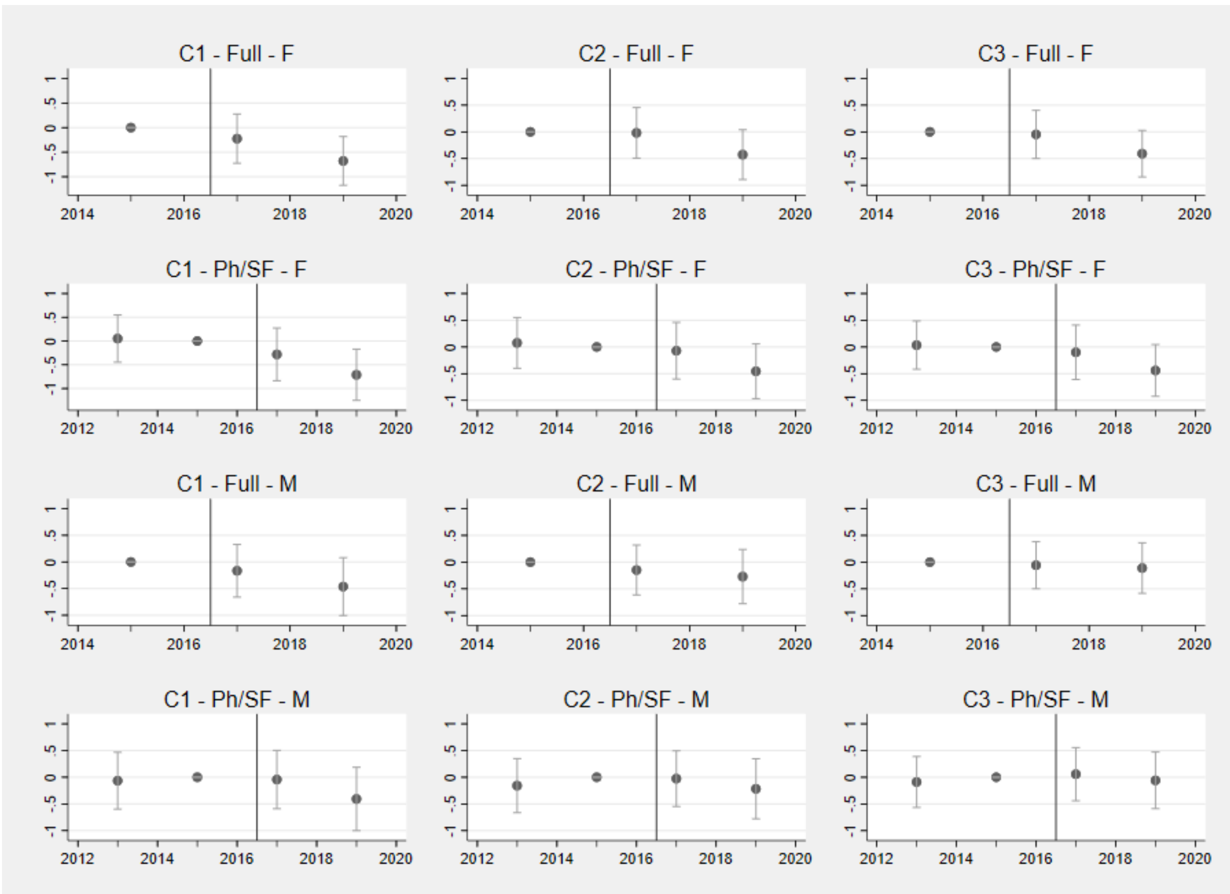
This figure displays the SSB consumption response, broken out by race. Both the Black and Hispanic groups decreased their consumption in 2017 and 2019, while the White and Asian groups appear to continue along the trends they were already on before the tax went into effect

Figure 3.5: SSB Tax Treatment Effect on BMI



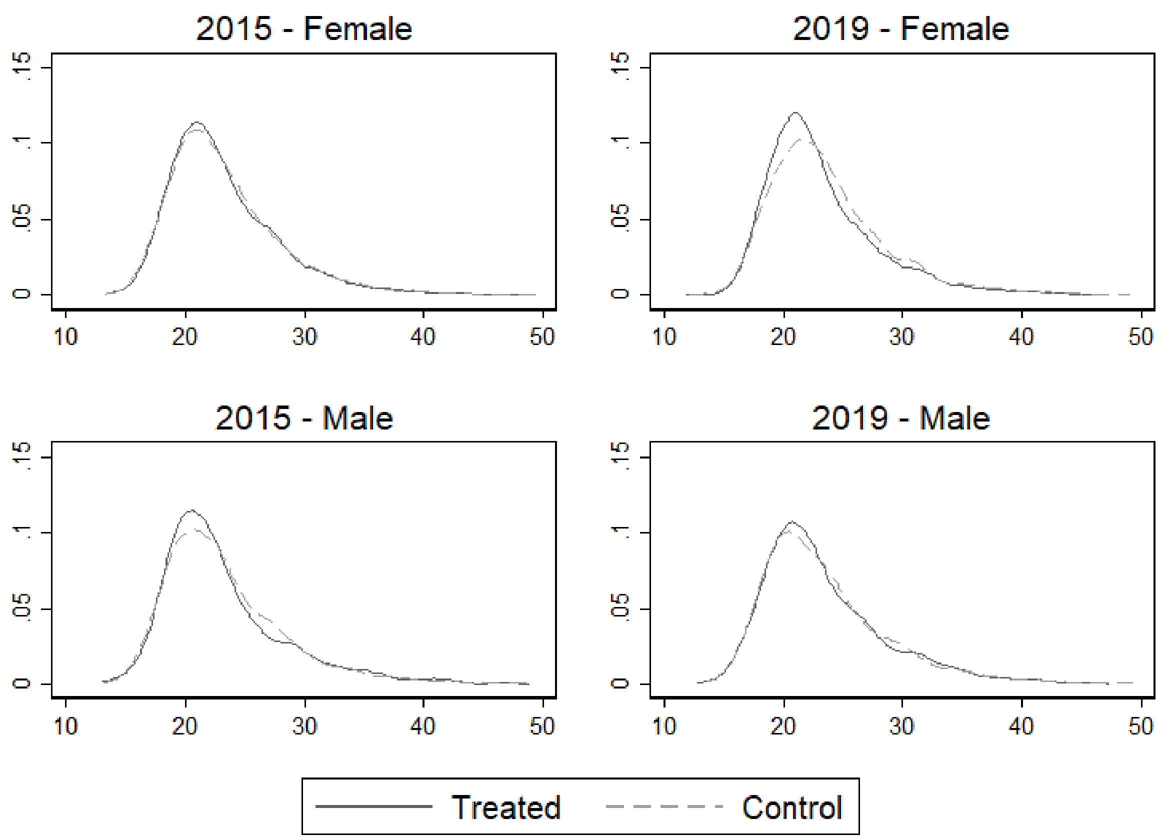
This figure displays estimates from equation (1) using the baseline control group (C1) for both BMI and obesity rates. Across the three treated cities, BMI displayed a minor decline of .2 points from 2015 to 2017, with a much larger decline of .57 points by 2019 (p-value = .002). The modified control group displays a similar treatment effect, with almost no movement from 2013 to 2015, which supports the equal counterfactual trends assumption. Similarly, obesity rates declined by 2.4 percentage points by 2019.

Figure 3.6: SSB Treatment on BMI by Gender



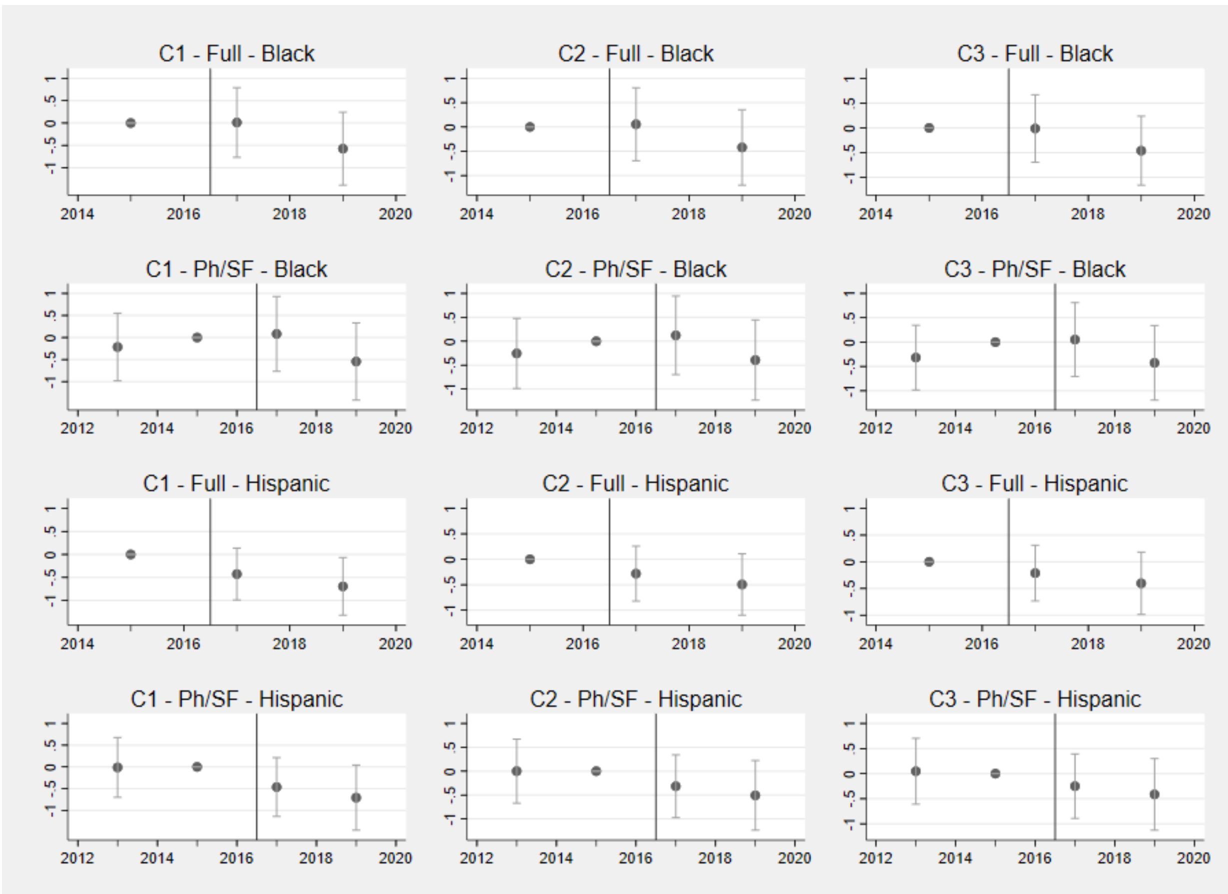
This figure compares the BMI response of male and female high school students across the three treated cities to control group C1. For females, the equal counterfactual trends assumption appears reasonable and then there is a decline of .68 BMI points by 2019. For males, the equal counterfactual treands assumption also appears reasonable, but there is an insignificant decline of only .34 points by 2019.

Figure 3.7: Pre and Post Treatment BMI Kernel Density Plots - By Gender



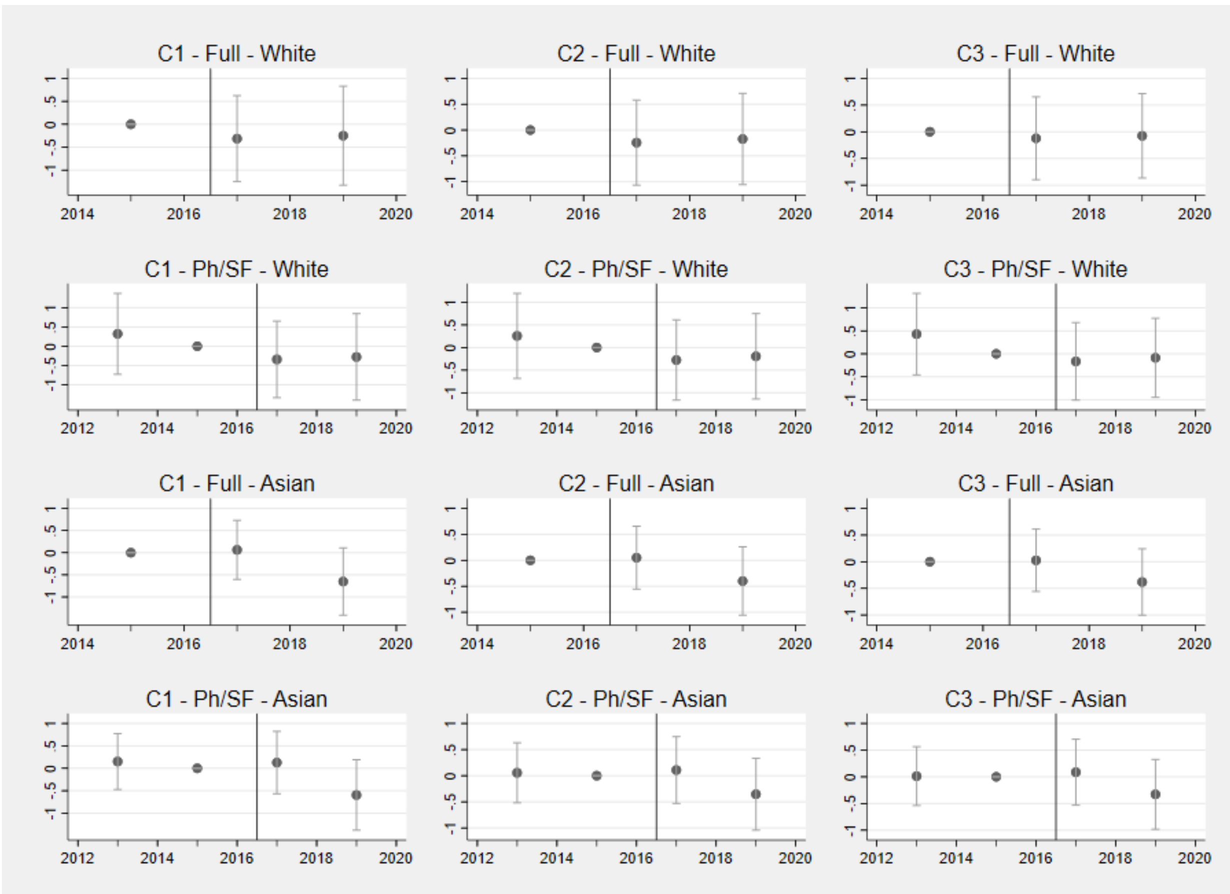
This figure compares BMI kernel densities before and after the SSB taxes went into effect. While the two appear to be almost identical in 2015, there is a sharp divergence in 2019, with significant improvements in the treated group relative to the control group.

Figure 3.8: BMI Treatment Effect by Race - Black and Hispanic



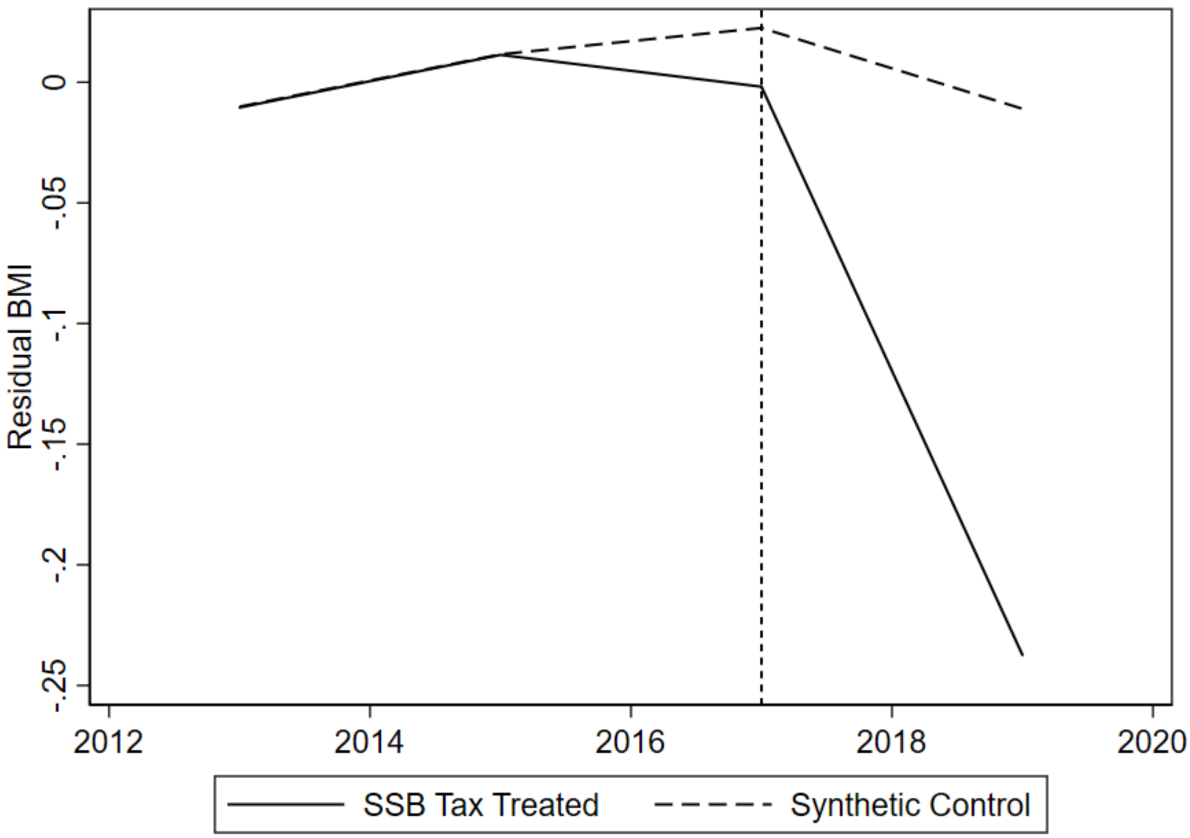
This figure compares the BMI response to SSB taxes across the three treated cities for the Black and Hispanic subpopulations. For the Black group, there is an uptick from 2013 to 2015 but then a decline of .54 points by 2019. For the Hispanic group, there is virtually no change from 2013 to 2015 but then a decline of .7 points by 2019.

Figure 3.9: BMI Treatment Effect by Race - White and Asian



This figure compares the BMI response to SSB taxes across the three treated cities for the White and Asian subpopulations. For the White group, there is an decrease from 2013 to 2015 but then little change from 2015 to 2019. For the Asian group, there is virtually no change from 2013 to 2017 but then a decline of .65 points by 2019.

Figure 3.10: Synthetic Control Specification Matched on Pre-Treatment Residual BMI



This figure displays the synthetic control method which estimates the effect of SSB tax treatment on BMI, estimated on the average residualized BMI after running equation (1) with all covariates but the treatment indicators on untreated observations, using the modified treatment group of San Francisco and Philadelphia.

Bibliography

[1] Tax on soda? jacksonville organization pushes the proposed legislation. Action News Jax, Jan 2017.

[2] California blocks local soda taxes until 2031. The Associated Press, Jun 2018.

[3] The history of salary disclosure for nhl players, 2018.

[4] Alberto Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. Journal of Economic Literature, 59(2):391–425, June 2021.

[5] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. Journal of the American Statistical Association, 105(490):493–505, 2010.

[6] Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. American Economic Review, 93(1):113–132, March 2003.

[7] Johannes Abeler, Armin Falk, Lorenz Goette, and David Huffman. Reference points and effort provision. American Economic Review, 101(2):470–92, April 2011.

[8] ACOG. Acog committee opinion no. 450: Increasing use of contraceptive implants and intrauterine devices to reduce unintended pregnancy. Obstet Gynecol, 114(6):1434–1438, 2009.

[9] Alan Adams. Players vote for salary disclosure. Edmonton Journal, Nov 1989.

[10] J. S. Adams. Towards an understanding of inequity. The Journal of Abnormal and Social Psychology, 67:422–436, 1963.

[11] J. S. Adams and W. B. Rosenbaum. The relationship of worker productivity to cognitive dissonance about wage inequities. Journal of Applied Psychology, 46(3):161–164, 1962.

[12] J. Stacy Adams. Wage inequities, productivity and work quality. Industrial Relations: A Journal of Economy and Society, 3(1):9–16, 1963.

[13] J. Stacy Adams. Inequity in social exchange. volume 2 of Advances in Experimental Social Psychology, pages 267–299. Academic Press, 1965.

[14] John Adams and PR Jacobsen. Effects of wage inequities on work quality. The Journal of Abnormal and Social Psychology, 69:19–25, 1964.

[15] Arturo Aguilar, Emilio Gutierrez, and Enrique Seira. The Effectiveness of Sin Food Taxes: Evidence from Mexico. Documentos de Trabajo LACEA 016421, The Latin American and Caribbean Economic Association - LACEA, July 2018.

[16] Arturo Aguilar, Emilio Gutierrez, and Enrique Seira. The effectiveness of sin food taxes: Evidence from mexico. Journal of Health Economics, 77:102455, 2021.

[17] George A. Akerlof and Janet L. Yellen. The fair wage-effort hypothesis and unemployment. The Quarterly Journal of Economics, 105(2):255–283, 1990.

[18] Greg R. Alexander, Martha S. Wingate, Deren Bader, and Michael D. Kogan. The increasing racial disparity in infant mortality rates: Composition and contributors to recent us trends. American Journal of Obstetrics and Gynecology, 198(1):51.e1–51.e9, 2008.

[19] Hunt Allcott, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell. Food Deserts and the Causes of Nutritional Inequality*. The Quarterly Journal of Economics, 134(4):1793–1844, 05 2019.

[20] Hunt Allcott, Benjamin Lockwood, and Dmitry Taubinsky. Regressive Sin Taxes, With an Application to the Optimal Soda Tax. NBER Working Papers 25841, National Bureau of Economic Research, Inc, May 2019.

[21] Hunt Allcott, Benjamin Lockwood, and Dmitry Taubinsky. Should we tax sugar-sweetened beverages? an overview of theory and evidence. Journal of Economic Perspectives, 33(3):202–27, August 2019.

[22] Hunt Allcott, Benjamin B Lockwood, and Dmitry Taubinsky. Regressive Sin Taxes, with an Application to the Optimal Soda Tax*. The Quarterly Journal of Economics, 134(3):1557–1626, 05 2019.

[23] Douglas Almond and Janet Currie. Killing me softly: The fetal origins hypothesis. Journal of Economic Perspectives, 25(3):153–72, September 2011.

[24] Douglas Almond, Jr. Doyle, Joseph J., Amanda E. Kowalski, and Heidi Williams. Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns*. The Quarterly Journal of Economics, 125(2):591–634, 05 2010.

[25] Elizabeth Oltmans Ananat, Jonathan Gruber, Phillip B Levine, and Douglas Staiger. Abortion and Selection. The Review of Economics and Statistics, 91(1):124–136, 02 2009.

[26] Elizabeth Oltmans Ananat and Daniel M. Hungerman. The Power of the Pill for the Next Generation: Oral Contraception's Effects on Fertility, Abortion, and Maternal and Child Characteristics. The Review of Economics and Statistics, 94(1):37–51, 02 2012.

[27] Tatiana Andreyeva, Michael W. Long, and Kelly D. Brownell. The impact of food prices on consumption: A systematic review of research on the price elasticity of demand for food. American Journal of Public Health, 100(2):216–222, 2010. PMID: 20019319.

[28] Raquel D Arias. Compelling reasons for recommending iuds to any woman of reproductive age. International journal of fertility and women's Medicine, 47(2):87—95, 2002.

[29] Marcus Asplund, Richard Friberg, and Fredrik Wilander. Demand and distance: Evidence on cross-border shopping. Journal of Public Economics, 91(1-2):141–157, February 2007.

[30] Styliadis I Athanasakis E, Karavasiliadou S. The factors contributing to the risk of sudden infant death syndrome. Hippokratia., 15(2):127–131, Apr 2011.

[31] Nicole Avena, Pedro Rada, and Bartley Hoebel. Evidence for sugar addiction: Behavioral and neurochemical effects of intermittent, excessive sugar intake. Neuroscience and Biobehavioral Reviews, 32(1):20 – 39, 2008.

[32] Simcha Avugos, Ofer H. Azar, Eran Sher, Nadav Gavish, and Michael Bar-Eli. The right-oriented bias in soccer penalty shootouts. Journal of Behavioral and Experimental Economics, 89:101546, 2020.

[33] Bailey. More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply*. The Quarterly Journal of Economics, 121(1):289–320, 02 2006.

[34] Bailey. Reexamining the impact of family planning programs on us fertility: Evidence from the war on poverty and the early years of title x. American Economic Journal: Applied Economics, 4(2):62–97, April 2012.

[35] Martha Bailey, Karen Clay, Price Fishback, Michael Haines, Shawn Kantor, Edson Severnini, and Anna. Wentz. U.s. county-level natality and mortality data, 1915-2007. Technical report, 2018.

[36] Martha J. Bailey, Brad Hershbein, and Amalia R. Miller. The opt-in revolution? contraception and the gender gap in wages. American Economic Journal: Applied Economics, 4(3):225–54, July 2012.

[37] Jim Baillie. An investigation into the collective bargaining relationship between the nhl and the nhlpa, 1994-1995. Technical Report 2005-02, Queens University, 2005.

[38] Michael Baker, Yosh Halberstam, Kory Kroft, Alexandre Mas, and Derek Messacar. Pay transparency and the gender gap. Working Paper 25834, National Bureau of Economic Research, May 2019.

[39] John A. Ballweg. Unwanted pregnancies and unwanted fertility: Conceptual variations. Population and Environment, 9(3):138–147, 1987.

[40] Stephen J. Bartlett. Contract negotiations and salary arbitration in the nhl...an agent's view. Marq. Sports L. Rev., 4(1), 1993.

[41] Björn Bartling and Ferdinand A. von Siemens. Wage inequality and team production: An experimental analysis. Journal of Economic Psychology, 32(1):1–16, 2011.

[42] Charles L. Baum. The effects of cigarette costs on bmi and obesity. Health Economics, 18(1):3–19, 2009.

[43] Jonathan Marc Bearak, Anna Popinchalk, Cynthia Beavin, Bela Ganatra, Ann-Beth Moller, Özge Tunçalp, and Leontine Alkema. Country-specific estimates of unintended pregnancy and abortion incidence: a global comparative analysis of levels in 2015–2019. BMJ Global Health, 7(3), 2022.

[44] Julie Beaulac, Elizabeth Kristjansson, and Steven Cummins. A systematic review of food deserts, 1966-2007. Preventing chronic disease, 6:A105, 08 2009.

[45] Chintan B. Bhatt and Consuelo M. Beck-Sagué. Medicaid expansion and infant mortality in the united states. American Journal of Public Health, 108(4):565–567, 2018. PMID: 29346003.

[46] Jay Bhattacharya, Kate Bundorf, Noemi Pace, and Neeraj Sood. Does health insurance make you fat? Working Paper 15163, National Bureau of Economic Research, July 2009.

[47] M.A. Biggs, C.H. Rocca, C.D. Brindis, H. Hirsch, and D. Grossman. Did increasing use of highly effective contraception contribute to declining abortions in iowa? Contraception, 91(2):167–173, 2015.

[48] Natalia E. Birgisson, Qiuhong Zhao, Gina M. Secura, Tessa Madden, and Jeffrey F. Peipert. Preventing unintended pregnancy: The contraceptive choice project in review. Journal of Women's Health, 24(5):349–353, 2015. PMID: 25825986.

[49] P.D. Blumenthal, A. Voedisch, and K. Gemzell-Danielsson. Strategies to prevent unintended pregnancy: increasing use of long-acting reversible contraception. Human Reproduction Update, 17(1):121–137, 07 2010.

[50] Bryan Bollinger and Steven E. Sexton. Local excise taxes, sticky prices, and spillovers: Evidence from berkeley's soda tax. Political Economy: Fiscal Policies and Behavior of Economic Agents eJournal, 2018.

[51] Patrick Bolton, Jose Scheinkman, and Wei Xiong. Executive compensation and short-termist behaviour in speculative markets. Review of Economic Studies, 73(3):577–610, 2006.

[52] John Bongaarts. Does family planning reduce infant mortality rates? Population and Development Review, 13(2):323–334, 1987.

[53] Martha Bornstein, Marion Carter, Lauren Zapata, Loretta Gavin, and Susan Moskosky. Access to long-acting reversible contraception among us publicly funded health centers. Contraception, 97:405–410, 2018.

[54] Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event study designs: Robust and efficient estimation. Technical report, 2021.

[55] Anat Bracha, Uri Gneezy, and George Loewenstein. Relative pay and labor supply. Journal of Labor Economics, 33(2):297 – 315, 2015.

[56] Kristyn Brandi and Liza Fuentes. The history of tiered-effectiveness contraceptive counseling and the importance of patient-centered family planning care. American Journal of Obstetrics and Gynecology, 222(4, Supplement):S873–S877, 2020. Long-Acting Reversible Contraception (LARC): Controversies Evidence.

[57] Emily Breza, Supreet Kaur, and Yogita Shamdasani. The Morale Effects of Pay Inequality*. The Quarterly Journal of Economics, 133(2):611–663, 10 2017.

[58] Tamara Broderick, Ryan Giordano, and Rachael Meager. An automatic finite-sample robustness metric: Can dropping a little data change conclusions?, 2020.

[59] Jennifer Brown. Quitters never win: The (adverse) incentive effects of competing with superstars. Journal of Political Economy, 119(5):982–1013, 2011.

[60] Jennifer Brown and Dylan B. Minor. Selecting the best? spillover and shadows in elimination tournaments. Management Science, 60(12):3087–3102, 2014.

[61] Kai J. Buhling, Nikki B. Zite, Pamela Lotke, and Kirsten Black. Worldwide use of intrauterine contraception: a review. Contraception, 89(3):162–173, 2014.

[62] Paolo Buonanno, Francesco Drago, Roberto Galbiati, and Giulio Zanella. Crime in Europe and the United States: dissecting the 'reversal of misfortunes'. Economic Policy, 26(67):347–385, 08 2014.

[63] Larry Burd. Maternal alcohol use increases risk of infant mortality. BMJ Evidence-Based Medicine, 19(1):27–27, 2014.

[64] Allison Burdo. N.y.c. couldn't get a soda tax passed, but this city just did. New York Business Journal, Jul 2016.

[65] Patricia Ritter Burga. Essays on obesity and oral health in developing countries. ProQuest Dissertations and These Global, 2016.

[66] Kristen Lagasse Burke, Joseph E. Potter, and Kari White. Unsatisfied contraceptive preferences due to cost among women in the united states. Contraception: X, 2, 2020.

[67] Steve Burtch. Intro to advanced hockey statistics -nbsp;fenwick, Jul 2012.

[68] M N Bustan and A L Coker. Maternal attitude toward pregnancy and the risk of neonatal death. American Journal of Public Health, 84(3):411–414, 1994. PMID: 8129057.

[69] Kyle Butts. Differences-in-differences with spatial spillovers. Technical report, 2021.

[70] Luigi Buzzacchi and Stefano Pedrini. Does player specialization predict player actions? evidence from penalty kicks at fifa world cup and uefa euro cup. Applied Economics, 46(10):1067–1080, 2014.

[71] Brantly Callaway and Pedro H.C. Sant'Anna. Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2):200–230, 2021. Themed Issue: Treatment Effect 1.

[72] Colin Camerer, Linda Babcock, George Loewenstein, and Richard Thaler. Labor supply of new york city cabdrivers: One day at a time. The Quarterly Journal of Economics, 112(2):407–441, 1997.

[73] David Card and Gordon B. Dahl. Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior*. The Quarterly Journal of Economics, 126(1):103–143, 02 2011.

[74] David Card, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. Inequality at work: The effect of peer salaries on job satisfaction. American Economic Review, 102(6):2981–3003, May 2012.

[75] Alexander Cardazzi, Bryan C McCannon, Brad R Humphreys, and Zachary Rodriguez. Emotional Cues and Violent Behavior: Unexpected Basketball Losses Increase Incidents of Family Violence. The Journal of Law, Economics, and Organization, 07 2022. ewac014.

[76] Eric Cardella and Alex Roomets. Pay distribution preferences and productivity effects: An experiment. Journal of Behavioral and Experimental Economics, 96:101814, 2022.

[77] Anne Case, Darren Lubotsky, and Christina Paxson. Economic status and health in childhood: The origins of the gradient. American Economic Review, 92(5):1308–1334, December 2002.

[78] John Cawley, Chelsea Crain, David Frisvold, and David Jones. The Pass-Through of the Largest Tax on Sugar-Sweetened Beverages: The Case of Boulder, Colorado. NBER Working Papers 25050, National Bureau of Economic Research, Inc, September 2018.

[79] John Cawley, Michael R Daly, and Rebecca Thornton. The effect of beverage taxes on youth consumption and bmi: Evidence from mauritius. Working Paper 28960, National Bureau of Economic Research, 6 2021.

[80] John Cawley, David Frisvold, Anna Hill, and David Jones. The impact of the philadelphia beverage tax on purchases and consumption by adults and children. Journal of Health Economics, 67:102225, 2019.

[81] John Cawley, David Frisvold, Anna Hill, and David Jones. The impact of the philadelphia beverage tax on purchases and consumption by adults and children. Journal of Health Economics, 67:102225, 08 2019.

[82] John Cawley, David Frisvold, Anna Hill, and David Jones. Oakland's sugar-sweetened beverage tax: Impacts on prices, purchases and consumption by adults and children. Econ Hum Biol, 2020.

[83] John Cawley, David Frisvold, David Jones, and Chelsea Lensing. The pass-through of a tax on sugar-sweetened beverages in boulder, colorado. American Journal of Agricultural Economics, 103(3):987–1005, 2021.

[84] John Cawley and Chad Meyerhoefer. The medical care costs of obesity: An instrumental variables approach. Journal of Health Economics, 31(1):219 – 230, 2012.

[85] CDPHE. Taking the unintended out of pregnancy: Colorado's success with long-acting reversible contraceptives. Technical report, 2017.

[86] National Women's Law Center. Contraceptive equity laws in your state: Know your rights - use your rights, a consumer guide. 2012.

[87] Frank Chaloupka and Kenneth E. Warner. The economics of smoking. In A. J. Culyer and J. P. Newhouse, editors, Handbook of Health Economics, volume 1, chapter 29, pages 1539–1627. Elsevier, 1 edition, 2000.

[88] Jonathan Chapman, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. Willingness to pay and willingness to accept are probably less correlated than you think. Working Paper 23954, National Bureau of Economic Research, October 2017.

[89] Gary Charness and Peter Kuhn. Does pay inequality affect worker effort? experimental evidence. Journal of Labor Economics, 25(4):693–723, 2007.

[90] Gary Charness and Peter Kuhn. Does pay inequality affect worker effort? experimental evidence. Journal of Labor Economics, 25(4):693–723, 2007.

[91] Kenneth Y. Chay and Michael Greenstone. The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. The Quarterly Journal of Economics, 118(3):1121–1167, 2003.

[92] Alice Chen, Emily Oster, and Heidi Williams. Why is infant mortality higher in the united states than in europe? American Economic Journal: Economic Policy, 8(2):89–124, May 2016.

[93] Alice Chen, Emily Oster, and Heidi Williams. Why is infant mortality higher in the united states than in europe? American Economic Journal: Economic Policy, 8(2):89–124, May 2016.

[94] P.-A. Chiappori, S. Levitt, and T. Groseclose. Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. American Economic Review, 92(4):1138–1151, September 2002.

[95] Dan Chisholm, Daniela Moro, Melanie Bertram, Carel Pretorius, Gerrit Gmel, Kevin Shield, and Jürgen Rehm. Are the “best buys” for alcohol control still valid? an update on the comparative cost-effectiveness of alcohol control strategies at the global level. Journal of Studies on Alcohol and Drugs, 79(4):514–522, 2018. PMID: 30079865.

[96] L Chola, S McGee, A Tugendhaft, E Buchmann, and K Hofman. Scaling Up Family Planning to Reduce Maternal and Child Mortality: The Potential Costs and Benefits of Modern Contraceptive Use in South Africa. PLoS ONE, 10(6), 2015.

[97] Shin-Yi Chou, Michael Grossman, and Henry Saffer. An economic analysis of adult obesity: results from the behavioral risk factor surveillance system. Journal of Health Economics, 23(3):565 – 587, 2004.

[98] Andrew E. Clark and Andrew J. Oswald. Satisfaction and comparison income. Journal of Public Economics, 61(3):359 – 381, 1996.

[99] Damian Clarke and Hanna Mühlrad. Abortion laws and women’s health. Journal of Health Economics, 76:102413, 2021.

[100] Alain Cohn, Ernst Fehr, and Lorenz Goette. Fair wages and effort provision: Combining evidence from a choice experiment and a field experiment. Management Science, 61(8):1777–1794, 2015.

[101] Alain Cohn, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. Social Comparison and Effort Provision: Evidence from a Field Experiment. Journal of the European Economic Association, 12(4):877–898, 08 2014.

[102] Simon Condliffe and Charles R. Link. The relationship between economic status and child health: Evidence from the united states. American Economic Review, 98(4):1605–18, September 2008.

[103] Russ Conway. Game misconduct: Alan Eagleson and the corruption of Hockey. MacFarlane Walter amp; Ross, 1997.

[104] Hope Corman and Michael Grossman. Determinants of neonatal mortality rates in the U.S. : A reduced form model. Journal of Health Economics, 4(3):213–236, September 1985.

[105] Charles Courtemanche. Rising cigarette prices and rising obesity: Coincidence or unintended consequence? Journal of Health Economics, 28(4):781 – 798, 2009.

[106] Vincent P. Crawford and Juanjuan Meng. New york city cab drivers' labor supply revisited: Reference-dependent preferences with rational-expectations targets for hours and income. American Economic Review, 101(5):1912–32, August 2011.

[107] Janet Currie and Matthew Neidell. Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?*. The Quarterly Journal of Economics, 120(3):1003–1030, 2005.

[108] Janet Currie and Mark Stabile. Socioeconomic status and child health: Why is the relationship stronger for older children? American Economic Review, 93(5):1813–1823, December 2003.

[109] Janet Currie and Reed Walker. Traffic congestion and infant health: Evidence from e-zpass. American Economic Journal: Applied Economics, 3(1):65–90, 2011.

[110] Kathryn M. Curtis and Jeffrey F. Peipert. Long-acting reversible contraception. New England Journal of Medicine, 376(5):461–468, 2017. PMID: 28146650.

[111] David M. Cutler and Adriana Lleras-Muney. Understanding differences in health behaviors by education. Journal of Health Economics, 29(1):1–28, 2010.

[112] Gus De Franco, Ole-Kristian Hope, and Stephannie Larocque. The effect of disclosure on the pay-performance relation. Journal of Accounting and Public Policy, 32(5):319 – 341, 2013.

[113] Janne C. de Ruyter, Margreet R. Olthof, Jacob C. Seidell, and Martijn B. Katan. A trial of sugar-free or sugar-sweetened beverages and body weight in children. New England Journal of Medicine, 367(15):1397–1406, 2012. PMID: 22998340.

[114] Lowen A. Cobley S. Deaner, R. O. Born at the wrong time: selection bias in the nhl draft. PloS one, 8(2), 2013.

[115] Mahshid Dehghan, Andrew Mente, Xiaohe Zhang, Sumathi Swaminathan, Wei Li, Viswanathan Mohan, Romaina Iqbal, Rajesh Kumar, Edelweiss Wentzel-Viljoen, Annika Rosengren, et al. Associations of fats and carbohydrate intake with cardiovascular disease and mortality in 18 countries from five continents (pure): a prospective cohort study. The Lancet, 390(10107):2050–2062, 2017.

[116] Linda M. Dinerman, Michele D. Wilson, Anne K. Duggan, and Alain Joffe. Outcomes of Adolescents Using Levonorgestrel Implants vs Oral Contraceptives or Other Contraceptive Methods. Archives of Pediatrics & Adolescent Medicine, 149(9):967–972, 09 1995.

[117] John Donohue, Jeffrey Grogger, and Steven Levitt. The Impact of Legalized Abortion on Teen Childbearing. American Law and Economics Review, 11(1):24–46, 06 2009.

[118] John Donohue and Steven Levitt. The Impact of Legalized Abortion on Crime*. The Quarterly Journal of Economics, 116(2):379–420, 05 2001.

[119] John Donohue and Steven Levitt. Measurement error, legalized abortion, and the decline in crime: A response to foote and goetz. The Quarterly Journal of Economics, 123(1):425–440, 2008.

[120] John Donohue and Steven Levitt. The Impact of Legalized Abortion on Crime over the Last Two Decades. American Law and Economics Review, 22(2):241–302, 11 2020.

[121] Arindrajit Dube, Laura Giuliano, and Jonathan Leonard. Fairness and frictions: The impact of unequal raises on quit behavior. American Economic Review, 109(2):620–63, February 2019.

[122] Pierre Dubois, Rachel Griffith, and Martin O'Connell. How well targeted are soda taxes? American Economic Review, 110(11):3661–3704, 11 2020.

[123] Kiyah J. Duffey, Penny Gordon-Larsen, James M. Shikany, David Guilkey, Jr Jacobs, David R., and Barry M. Popkin. Food Price and Diet and Health Outcomes: 20 Years of the CARDIA Study. Archives of Internal Medicine, 170(5):420–426, 3 2010.

[124] Cara B. Ebbeling, Janis F. Swain, Henry A. Feldman, William W. Wong, David L. Hachey, Erica Garcia-Lago, and David S. Ludwig. Effects of Dietary Composition on Energy Expenditure During Weight-Loss Maintenance. JAMA, 307(24):2627–2634, 06 2012.

[125] Marc Edge. Red Line, Blue Line, Bottom line: How push came to shove between the National Hockey League and its players. New Star Books, 2004.

[126] Emma K. Edmondson, Christina A. Roberto, Emily F. Gregory, Nandita Mitra, and Senbagam Virudachalam. Association of a Sweetened Beverage Tax With Soda Consumption in High School Students. JAMA Pediatrics, 175(12):1261–1268, 12 2021.

[127] Keith Ericson and Andreas Fuster. Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. The Quarterly Journal of Economics, 126(4):1879–1907, 2011.

[128] Cabrera Escobar, J.L Veerman, and S.S. et al. Tollman. Evidence that a tax on sugar sweetened beverages reduces the obesity rate: a meta-analysis. BMC Public Health, 13, 2013.

[129] Jennifer Falbe, Hannah R. Thompson, Christina M. Becker, Nadia Rojas, Charles E. McCulloch, and Kristine A. Madsen. Impact of the berkeley excise tax on sugar-sweetened beverage consumption. American Journal of Public Health, 106(10):1865–1871, 2016. PMID: 27552267.

[130] Henry S. Farber. Is tomorrow another day? the labor supply of new york city cabdrivers. Journal of Political Economy, 113(1):46–82, 2005.

[131] Robert Fatchet. Lemieux no. 1 in nhl with $2 million salary. Washington Post, Jan 1990.

[132] Robert Fatchet. Salary figures raise passions. Washington Post, Jan 1990.

[133] Ernst Fehr and Lorenz Goette. Do workers work more if wages are high? evidence from a randomized field experiment. American Economic Review, 97(1):298–317, March 2007.

[134] Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. The Quarterly Journal of Economics, 114(3):817–868, 1999.

[135] Gerald A. Feltham and Jim Xie. Performance measure congruity and diversity in multi-task principal/agent relations. The Accounting Review, 69(3):429–453, 1994.

[136] Lou Festinger. A theory of cognitive dissonance. Stanford University Press, 1957.

[137] Brian Karl Finch. Early origins of the gradient: the relationship between socioeconomic status and infant mortality in the United States. Demography, 40(4):675–699, 11 2003.

[138] Lawrence B. Finer and Mia R. Zolna. Shifts in intended and unintended pregnancies in the united states, 2001–2008. American Journal of Public Health, 104(S1):S43–S48, 2014. PMID: 24354819.

[139] EA Finkelstein, C Zhen, M Bilger, J Nonnemaker, AM Farooqui, and Todd JE. Implications of a sugar-sweetened beverage (ssb) tax when substitutions to non-beverage items are considered. Journal of Health Economics, 32:219–39, 1 2013.

[140] Jocelyn E Finlay, Emre Özaltin, and David Canning. The association of maternal age with infant mortality, child anthropometric failure, diarrhoea and anaemia for first births: evidence from 55 low- and middle-income countries. BMJ Open, 1(2), 2011.

[141] Jason Fletcher, David Frisvold, and Nathan Tefft. The effects of soft drink taxes on child and adolescent consumption and weight outcomes. Journal of Public Economics, 94(11-12):967–974, 2010.

[142] JASON M. FLETCHER, DAVID FRISVOLD, and NATHAN TEFFT. Can soft drink taxes reduce population weight? Contemporary Economic Policy, 28(1):23–35, 2010.

[143] K. Fliessbach, B. Weber, P. Trautner, T. Dohmen, U. Sunde, C. E. Elger, and A. Falk. Social comparison affects reward-related brain activity in the human ventral striatum. Science, 318(5854):1305–1308, 2007.

[144] A. Flower, J. Shawe, and J. et al. Stephenson. Pregnancy planning, smoking behaviour during pregnancy, and neonatal outcome: Uk millennium cohort study. BMC Pregnancy Childbirth, (13), 2013.

[145] Christopher L. Foote and Christopher F. Goetz. The impact of legalized abortion on crime: Comment. The Quarterly Journal of Economics, 123(1):407–423, 2008.

[146] Diana Greene Foster, M. Antonia Biggs, Sarah Raifman, Jessica Gipson, Katrina Kimport, and Corinne H. Rocca. Comparison of Health, Development, Maternal Bonding, and Poverty Among Children Born After Denial of Abortion vs After Pregnancies Subsequent to an Abortion. JAMA Pediatrics, 172(11):1053–1060, 11 2018.

[147] Matthew Fournier and Dominic Roux. Labor relations in the national hockey league: A model of transnational collective bargaining. Marq. Sports L. Rev., 20(1), 2009.

[148] Abel François, Raul Magni-Berton, and Laurent Weill. Abortion and crime: Cross-country evidence from europe. International Review of Law and Economics, 40:24–35, 2014.

[149] John Gardner. Two stage difference-in-differences. Technical report, 2021.

[150] Olivier Godechot and Claudia Senik. Wage comparisons in and out of the firm. evidence from a matched employer–employee french database. Journal of Economic Behavior and Organization, 117:395 – 410, 2015.

[151] Claudia Goldin and Lawrence F. Katz. The power of the pill: Oral contraceptives and women's career and marriage decisions. Journal of Political Economy, 110(4):730–770, 2002.

[152] Thomas Goldring. ddtiming: Stata module to perform a goodman-bacon decomposition of difference-in-differences estimation. Technical report, 2019.

[153] Lisa M. Goldthwaite, Lindsey Duca, Randi K. Johnson, Danielle Ostendorf, and Jeanelle Sheeder. Adverse birth outcomes in colorado: Assessing the impact of a statewide initiative to prevent unintended pregnancy. American Journal of Public Health, 105(9):e60–e66, 2015. PMID: 26180990.

[154] Anu Manchikanti Gomez, Liza Fuentes, and Amy Allina. Women or larc first? reproductive autonomy and the promotion of long-acting reversible contraceptive methods. Perspectives on Sexual and Reproductive Health, 46(3):171–175, 2014.

[155] Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2):254–277, 2021. Themed Issue: Treatment Effect 1.

[156] William R. Grady, Mark D. Hayward, and Junichi Yagi. Contraceptive failure in the united states: Estimates from the 1982 national survey of family growth. Family Planning Perspectives, 18(5):200–209, 1986.

[157] Tadeja Gračner, Fernanda Marquez-Padilla, and Danae Hernandez-Cortes. Changes in Weight-Related Outcomes Among Adolescents Following Consumer Price Increases of Taxed Sugar-Sweetened Beverages. JAMA Pediatrics, 176(2):150–158, 02 2022.

[158] J. Greenberg. Equity and workplace status: a field experiment. Applied Psychology, 73(4):606–613, 1988.

[159] Michael Grossman and Steven Jacobowitz. Variations in Infant Mortality Rates Among Counties of the United States: The Roles of Public Policies and Programs. Demography, 18(4):695–713, 11 1981.

[160] Michael Grossman and Theodore J Joyce. Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City. Journal of Political Economy, 98(5):983–1007, October 1990.

[161] Jonathan Gruber and Michael Frakes. Does falling smoking lead to rising obesity? Journal of Health Economics, 25(2):183 – 197, 2006.

[162] Jonathan Gruber and Botond Koszegi. Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. Journal of Public Economics, 88(9-10):1959–1987, 2004.

[163] Jonathan Gruber, Phillip Levine, and Douglas Staiger. Abortion legalization and child living circumstances: Who is the "marginal child"? The Quarterly Journal of Economics, 114(1):263–291, 1999.

[164] Anna H. Grummon, Benjamin B. Lockwood, Dmitry Taubinsky, and Hunt Allcott. Designing better sugary drink taxes. Science, 365(6457):989–990, 2019.

[165] CM Guerrero-López, M Unar-Munguía, and MA Colchero. Price elasticity of the demand for soft drinks, other sugar-sweetened beverages and energy dense food in chile. BMC Public Health, 17(180), 2017.

[166] J.A. Hall, L. Benton, and A. et al. Copas. Pregnancy intention and pregnancy outcome: Systematic review and meta-analysis. Matern Child Health J, (21):670–704, 2017.

[167] Matthew Harding, Ephraim Leibtag, and Michael F. Lovenheim. The heterogeneous geographic and socioeconomic incidence of cigarette taxes: Evidence from nielsen homescan data. American Economic Journal: Economic Policy, 4(4):169–98, May 2012.

[168] Robert A. Hatcher, James Trussell, Anita L. Nelson, Willard Cates, Deborah Kowal, and Michael S. Policar. Contraceptive Technology. Ardent Media, 20th edition, 2011.

[169] W. Hayman, S. Leuthner, N. Laventhal, D. Brousseau, and J Lagatta. Cost comparison of mechanically ventilated patients across the age span. J Perinatol, 35:1020–1026, 2015.

[170] Ori Heffetz and John A. List. Is the endowment effect and expectations effect? Journal of the European Economic Association, 12(5):1396–1422, 2014.

[171] Thomas Hellmann and Veikko Thiele. Incentives and innovation: A multitasking approach. American Economic Journal: Microeconomics, 3(1):78–128, February 2011.

[172] Leah Henke, Summer Martins, and Christy Boraas. Barriers to obtaining long-acting reversible contraception among low-income women. Obstetrics & Gynecology, 135, 2020.

[173] Bengt Holmstrom and Paul Milgrom. Aggregation and linearity in the provision of intertemporal incentives. Econometrica, 55(2):303–328, 1987.

[174] Bengt Holmstrom and Paul Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. Journal of Law, Economics, Organization, 7:24–52, 1991.

[175] G. C. Homans. Social behavior: Its elementary forms. Harcourt, Brace., 1961.

[176] Hilary Hoynes, Doug Miller, and David Simon. Income, the earned income tax credit, and infant health. American Economic Journal: Economic Policy, 7(1):172–211, February 2015.

[177] Hilary Hoynes, Diane Whitmore Schanzenbach, and Douglas Almond. Long-run impacts of childhood access to the safety net. American Economic Review, 106(4):903–34, April 2016.

[178] EJ Hradek. More to plus-minus than meets the eye. ESPN: The Magazine, Dec 2003.

[179] Fumiaki Imamura, Laura O'Connor, Zheng Ye, Jaakko Mursu, Yasuaki Hayashino, Shilpa N Bhupathiraju, and Nita G Forouhi. Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: systematic review, meta-analysis, and estimation of population attributable fraction. BMJ, 351, 2015.

[180] Mireille Jacobson and Heather Royer. Aftershocks: The impact of clinic violence on abortion services. American Economic Journal: Applied Economics, 3(1):189–223, January 2011.

[181] Elliott Jaques. Measurement of responsibility. Travistock, 1956.

[182] David J.A. Jenkins, Mahshid Dehghan, Andrew Mente, Shrikant I. Bangdiwala, Sumathy Rangarajan, Kristie Srichaikul, Viswanathan Mohan, Alvaro Avezum, Rafael Díaz, Annika Rosengren, Fernando Lanas, Patricio Lopez-Jaramillo, Wei Li, Aytekin Oguz, Rasha Khatib, Paul Poirier, Noushin Mohammadifard, Andrea Pepe, Khalid F. Alhabib, Jephat Chifamba, Afzal Hussein Yusufali, Romaina Iqbal, Karen Yeates, Khalid Yusoff, Noorhassim Ismail, Koon Teo, Sumathi Swaminathan, Xiaoyun Liu, Katarzyna Zatońska, Rita Yusuf, and Salim Yusuf. Glycemic index, glycemic load, and cardiovascular disease and mortality. New England Journal of Medicine, 384(14):1312–1322, 2021. PMID: 33626252.

[183] Karen M. Jetter and Diana L. Cassady. The availability and cost of healthier food alternatives. American Journal of Preventive Medicine, 30(1):38–44, 2006.

[184] Anna L. V.a Johansson, Paul W.a Dickman, Michael S.b Kramer, and Svena Cnattingius. Maternal smoking and infant mortality. Epidemiology, 20, 2009.

[185] Jessica C. Jones-Smith, Melissa A. Knox, Norma B. Coe, Lina P. Walkinshaw, John Schoof, Deven Hamilton, Philip M. Hurvitz, and James Krieger. Sweetened beverage taxes: Economic benefits and costs according to household income. Food Policy, 110:102277, 2022.

[186] Ted Joyce. Did legalized abortion lower crime? The Journal of Human Resources, 39(1):1–28, 2004.

[187] Ted Joyce. A Simple Test of Abortion and Crime. The Review of Economics and Statistics, 91(1):112–123, 02 2009.

[188] Theodore Joyce. The impact of induced abortion on black and white birth outcomes in the United States. Demography, 24(2):229–244, 05 1987.

[189] Daniel T. Kaffine and Graham A. Davis. A multi-row deletion diagnostic for influential observations in small-sample regressions. Computational Statistics & Data Analysis, 108(C):133–145, 2017.

[190] Leo H. Kahane, David Paton, and Rob Simmons. The abortion–crime link: Evidence from england and wales. Economica, 75(297):1–21, 2008.

[191] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. Econometrica, 47(2):263–291, 1979.

[192] Brad Kamp and Philip Porter. Asymmetric information in wage negotiations: Hockey's natural experiment. Working Papers 1413, University of South Florida, Department of Economics, 2013.

[193] Mark Kassis, Sascha L. Schmidt, Dominik Schreyer, and Matthias Sutter. Psychological pressure and the right to determine the moves in dynamic tournaments – evidence from a natural field experiment. Games and Economic Behavior, 126:278–287, 2021.

[194] Megan L. Kavanaugh, Jenna Jerman, and Lawrence B. Finer. Changes in use of long-acting reversible contraceptive methods among u.s. women, 2009-2012. Obstretrics & Gynecology, 126:917–927, 2015.

[195] Melissa S. Kearney and Phillip B. Levine. Why is the teen birth rate in the united states so high and why does it matter? Journal of Economic Perspectives, 26(2):141–63, May 2012.

[196] Melissa S. Kearney, Phillip B. Levine, and Luke Pardue. The puzzle of falling us birth rates since the great recession. Journal of Economic Perspectives, 36(1):151–76, February 2022.

[197] Andrea Kelly, Jason M. Lindo, and Analisa Packham. The power of the iud: Effects of expanding access to contraception through title x clinics. Journal of Public Economics, 192:104288, 2020.

[198] Mike Kiley. Nhl checks fall far short. Chicago Tribune, Jan 1990.

[199] Mike Kiley. Blues pay dearly for free agent - captain stevens. Chicago Tribune, Sep 1991.

[200] M. Kim, S. Lee, and SH. et al. Bae. Socioeconomic status can affect pregnancy outcomes and complications, even with a universal healthcare system. Int J Equity Health, 17(2), 2018.

[201] Marlene Kim. Pay secrecy and the gender wage gap in the united states. Industrial Relations: A Journal of Economy and Society, 54(4):648–667, 2015.

[202] Jack Knetsch and Wei-Kang Wong. The endowment effect and the reference state: Evidence and manipulations. Journal of Economic Behavior Organization, 71(2):407–413, 2009.

[203] Kathryn Kost and Laura Lindberg. Pregnancy Intentions, Maternal Behaviors, and Infant Health: Investigating Relationships With New Measures and Propensity Score Analysis. Demography, 52(1):83–111, 01 2015.

[204] James Krieger, Kiran Magee, Tayler Hennings, John Schoof, and Kristine A. Madsen. How sugar-sweetened beverage tax revenues are being used in the united states. Preventive Medicine Reports, 23:101388, 2021.

[205] Shantha Kumari. Permanent sterilisation to long-acting reversible contraception: Is a paradigm shift necessary? J Obstet Gynaecol India., 66(3):149–153, June 2016.

[206] Botond Kőszegi and Matthew Rabin. A Model of Reference-Dependent Preferences*. The Quarterly Journal of Economics, 121(4):1133–1165, 11 2006.

[207] Mario Lackner and Hendrik Sonnabend. Coping with advantageous inequity—field evidence from professional penalty kicking. Journal of Behavioral and Experimental Economics, 91:101678, 2021.

[208] Charles P Larson. Poverty during pregnancy: Its effects on child health outcomes. Paediatrics Child Health, 12(8):673–677, 10 2007.

[209] Matthew M. Lee, Jennifer Falbe, Dean Schillinger, Sanjay Basu, Charles E. McCulloch, and Kristine A. Madsen. Sugar-sweetened beverage consumption 3 years after the berkeley, california, sugar-sweetened beverage tax. American Journal of Public Health, 109(4):637–639, 2019. PMID: 30789776.

[210] Lars J. Lefgren, David P. Sims, and Olga B. Stoddard. Effort, luck, and voting for redistribution. Journal of Public Economics, 143:89–97, 2016.

[211] Julien Leider and Powell. Longer-term impacts of the oakland, california, sugar-sweetened beverage tax on prices and volume sold at two-years post-tax. Social Science Medicine, 292:114537, 2022.

[212] Phillip Levine, D Staiger, T Kane, and D Zimmerman. Roe v wade and american fertility. American Journal of Public Health, 89(2):199–203, 1999. PMID: 9949749.

[213] Steven D. Levitt. Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. Journal of Economic Perspectives, 18(1):163–190, March 2004.

[214] Shirlee Lichtman-Sadot. Does banning carbonated beverages in schools decrease student consumption? Journal of Public Economics, 140:30–50, 2016.

[215] Jason M. Lindo and Analisa Packham. How much can expanding access to long-acting reversible contraceptives reduce teen birth rates? American Economic Journal: Economic Policy, 9(3):348–76, August 2017.

[216] Jens Ludwig, Lisa Sanbonmatsu, Lisa Gennetian, Emma Adam, Greg J. Duncan, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Stacy Tessler Lindau, Robert C. Whitaker, and Thomas W. McDade. Neighborhoods, obesity, and diabetes — a randomized social experiment. New England Journal of Medicine, 365(16):1509–1519, 2011. PMID: 22010917.

[217] Erzo Luttmer. Neighbors as negatives: Relative earnings and well-being. Quarterly Journal of Economics, 120:963–1002, 08 2005.

[218] Marianna Makri, Peter J. Lane, and Luis R. Gomez-Mejia. Ceo incentives, innovation, and performance in technology-intensive firms: a reconciliation of outcome and behavior-based incentive schemes. Strategic Management Journal, 27(11):1057–1080, 2006.

[219] Vasanti S. Malik, Barry M. Popkin, George A. Bray, Jean-Pierre Després, Walter C. Willett, and Frank B. Hu. Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes. Diabetes Care, 33(11):2477–2483, 2010.

[220] Michael Marmot. Status syndrome. Significance, 1(4):150–154, 2004.

[221] James Marsh. National hockey league. The Canadian Encyclopedia, 2006.

[222] M C Marshall. Diabetes in african americans. Postgraduate Medical Journal, 81(962):734–740, 2005.

[223] Alexandre Mas. Pay, Reference Points, and Police Performance*. The Quarterly Journal of Economics, 121(3):783–821, 08 2006.

[224] Alexandre Mas. Does transparency lead to pay compression? Journal of Political Economy, 125(5):1683–1721, 2017.

[225] Jane Mauldon, Diana Greene Foster, and Sarah C. M. Roberts. Effect of abortion vs. carrying to term on a woman's relationship with the man involved in the pregnancy. Perspectives on Sexual and Reproductive Health, 47(1):11–18, 2015.

[226] Colleen McNicholas, Tessa Madden, Gina Secura, and Jeffrey Peipert. The contraceptive choice project round up. Clinical Obstetrics and Gynecology, 57(4):635–643, December 2014.

[227] Sarah Miller, Laura R Wherry, and Diana Greene Foster. The economic consequences of being denied an abortion. Working Paper 26662, National Bureau of Economic Research, January 2020.

[228] Chandra S Mishra, Daniel L McConaughy, and David H Gobeli. Effectiveness of ceo pay-for-performance. Review of Financial Economics, 9(1):1–13, 2000.

[229] Y.A. Mohamoud, R.S. Kirby, and D.B. Ehrenthal. Poverty, urban-rural classification and term infant mortality: a population-based multilevel analysis. BMC Pregnancy Childbirth, 19(40), 2019.

[230] Loretta M. Moore and Reuben M. Baron. Effects of wage inequities on work attitudes and performance. Journal of Experimental Social Psychology, 9(1):1–16, 1973.

[231] Katherine W. Morgan. The intrauterine device: Rethinking old paradigms. Journal of Midwifery Women's Health, 51(6):464–470, 2006. Special Continuing Education Issue.

[232] DM Mourao, J Bressan, W Campbell, and RD Mattes. Effects of food form on appetite and energy intake in lean and obese young adults. Int J Obes, 2007.

[233] Dariush Mozaffarian, Tao Hao, Eric B. Rimm, Walter C. Willett, and Frank B. Hu. Changes in diet and lifestyle and long-term weight gain in women and men. New England Journal of Medicine, 364(25):2392–2404, 2011. PMID: 21696306.

[234] Alistair Munro and Robert Sugden. On the theory of reference-dependent preferences. Journal of Economic Behavior Organization, 50(4):407–428, 2003.

[235] Caitlin Knowles Myers. The power of abortion policy: Reexamining the effects of young women's access to reproductive control. Journal of Political Economy, 125(6):2178–2224, 2017.

[236] Sharon Nakhimovsky, AB Feigl, C Avila, G O'Sullivan, E Macgregor-Skinner, and Spranca M. Taxes on sugar-sweetened beverages to reduce overweight and obesity in middle-income countries: A systematic review. PLoS One, 11(9), 2016.

[237] Susan K Neely. Tax makes wallets thinner but doesn't shrink waistlines: Beverage advocate. The Orlando Sentinel, Jul 2016.

[238] NFPRHA. Title x family planning annual report: 2013 national summary. National Family Planning and Reproductive Health Association, 2013.

[239] NN Nour. Obesity in resource-poor nations. Rev Obstet Gynecol., 3:180–4, 2010.

[240] Shakked Noy and Isabelle Sin. The effects of neighbourhood and workplace income comparisons on subjective wellbeing. Journal of Economic Behavior Organization, 185:918–945, 2021.

[241] Tomasz Obloj and Todd R. Zenger. The influence of pay transparency on inequity, inequality, and the performance-basis of pay. SSRN, 2020.

[242] OECD. Oecd health statistics, 2019. Organisation for Economic Co-operation and Development, 2019.

[243] Gerald S. Oettinger. An empirical analysis of the daily labor supply of stadium vendors. Journal of Political Economy, 107(2):360–392, 1999.

[244] National Conference of State Legislatures. Insurance coverage for contraception laws. 2012.

[245] David L. Olds, Harriet Kitzman, Carole Hanks, Robert Cole, Elizabeth Anson, Kimberly Sidora-Arcoleo, Dennis W. Luckey, Charles R. Henderson, John Holmberg, Robin A. Tutt, Amanda J. Stevenson, and Jessica Bondy. Effects of nurse home visiting on maternal and child functioning: Age-9 follow-up of a randomized trial. Pediatrics, 120(4):e832–e845, 2007.

[246] CM O'Leary, PJ Jacoby, A Bartu, D'Antoine H, and C. Bower. Maternal alcohol use and sudden infant death syndrome and infant mortality excluding sids. Pediatrics, 131(3), 2013.

[247] Emily Oster. Diabetes and diet: Purchasing behavior change in response to health information. American Economic Journal: Applied Economics, 10(4):308–48, October 2018.

[248] Mary A. Ott, Gina S. Sucato, COMMITTEE ON ADOLESCENCE, Paula K. Braverman, William P. Adelman, Elizabeth M. Alderman, Cora C. Breuner, David A. Levine, Arik V. Marcell, and Rebecca F. O'Brien. Contraception for Adolescents. Pediatrics, 134(4):e1257–e1281, 10 2014.

[249] Roman Pabayo, Amy Ehntholt, Daniel M. Cook, Megan Reynolds, Peter Muennig, and Sze Y. Liu. Laws restricting access to abortion services and infant mortality risk in the united states. International Journal of Environmental Research and Public Health, 17(11), 2020.

[250] An Pan and Frank Hu. Effects of carbohydrates on satiety: Differences between liquid and solid food. Current opinion in clinical nutrition and metabolic care, 14:385–90, 07 2011.

[251] Prakash Dev Pant. Effect of education and household characteristics on infant and child mortality in urban nepal. Journal of Biosocial Science, 23(4):437–443, 1991.

[252] Ravi M. Patel, Sarah Kandefer, Michele C. Walsh, Edward F. Bell, Waldemar A. Carlo, Abbot R. Laptook, Pablo J. Sánchez, Seetha Shankaran, Krisa P. Van Meurs, M. Bethany Ball, Ellen C. Hale, Nancy S. Newman, Abhik Das, Rosemary D. Higgins, and Barbara J. Stoll. Causes and timing of death in extremely premature infants from 2000 through 2011. New England Journal of Medicine, 372(4):331–340, 2015. PMID: 25607427.

[253] Karen Pazol, Sonya Gamble, Wilda Parker, Douglas Cook, Suzanne Zane, and Hamdan Saeed. Abortion surveillance, united states, 2006. Technical report, 2006.

[254] Jeffrey F. Peipert, Qiuhong Zhao, Jenifer E. Allsworth, Emiko Petrosky, Tessa Madden, David Eisenberg, and Gina Secura. Continuation and satisfaction of reversible contraception. Obstetrics and Gynecology, 117(5):1105–1113, 5 2011.

[255] Véronique Pierrat, Laetitia Marchand-Martin, Stéphane Marret, Catherine Arnaud, Valérie Benhammou, Gilles Cambonie, Thierry Debillon, Marie-Noëlle Dufourg, Catherine Gire, François Goffinet, Monique Kaminski, Alexandre Lapillonne, Andrei Scott Morgan, Jean-Christophe Rozé, Sabrina Twilhaar, Marie-Aline Charles, and Pierre-Yves Ancel. Neurodevelopmental outcomes at age 5 among children born preterm: Epipage-2 cohort study. BMJ, 373, 2021.

[256] Joseph Potter, Celia Hubert, Amanda Stevenson, Kristine Hopkins, Abigail Aiken, Kari White, and Daniel Grossman. Postpartum contraception in texas and pregnancy within 2 years of delivery. Obstetrics Gynecology, 127:289–296, 2016.

[257] Joseph E. Potter, Kristine Hopkins, Abigail R.A. Aiken, Celia Hubert, Amanda J. Stevenson, Kari White, and Daniel Grossman. Unmet demand for highly effective postpartum contraception in texas. Contraception, 90(5):488–495, 2014.

[258] Powell, J Chriqui, T Khan, R Wada, and F Chaloupka. Assessing the potential effectiveness of food and beverage taxes and subsidies for improving public health: a systematic review of prices, demand and body weight outcomes. Obesity Reviews, 14(2):110–128, 2013.

[259] Powell and Julien Leider. Evaluation of Changes in Beverage Prices and Volume Sold Following the Implementation and Repeal of a Sweetened Beverage Tax in Cook County, Illinois. JAMA Network Open, 3(12):e2031083–e2031083, 12 2020.

[260] Powell and Julien Leider. The impact of seattle's sweetened beverage tax on beverage prices and volume sold. Economics Human Biology, 37:100856, 2020.

[261] Powell, Julien Leider, and Pierre Leger. The impact of a sweetened beverage tax on beverage volume sold in cook county, illinois, and its border area. Annals of Internal Medicine, 172(6):390–397, 2020. PMID: 32092766.

[262] Lisa M. Powell, Julien Leider, and Vanessa M. Oddo. Evaluation of Changes in Grams of Sugar Sold After the Implementation of the Seattle Sweetened Beverage Tax. JAMA Network Open, 4(11):e2132271–e2132271, 11 2021.

[263] Lisa M. Powell, Sandy Slater, Donka Mirtcheva, Yanjun Bao, and Frank J. Chaloupka. Food store availability and neighborhood characteristics in the united states. Preventive Medicine, 44(3):189–195, 2007.

[264] Inas Rashad, Michael Grossman, and Shin-Yi Chou. The Super Size of America: An Economic Estimation of Body Mass Index and Obesity in Adults. Eastern Economic Journal, 32(1):133–148, 12 2006.

[265] Ed Ratushny. Dispute Resolution in the National Hockey League. Labour Arbitration Yearbook, 1992.

[266] Emily Rauscher and David E. Rangel. Rising inequality of infant health in the u.s. SSM - Population Health, 12:100698, 2020.

[267] Sue Ricketts, Greta Klingler, and Renee Schwalberg. Game change in colorado: Widespread use of long-acting reversible contraceptives and rapid decline in births among young, low-income women. Perspectives on Sexual and Reproductive Health, 46(3):125–132, 2014.

[268] James M. Rippe and Theodore J. Angelopoulos. Relationship between added sugars consumption and chronic disease risk factors: Current understanding. Nutrients, 8(11), 2016.

[269] Christina A. Roberto, Hannah G. Lawman, Michael T. LeVasseur, Nandita Mitra, Ana Peterhans, Bradley Herring, and Sara N. Bleich. Association of a Beverage Tax on Sugar-Sweetened and Artificially Sweetened Beverages With Changes in Beverage Prices and Sales at Chain Retailers in a Large Urban Setting. JAMA, 321(18):1799–1810, 05 2019.

[270] S.C. Roberts, M.A. Biggs, and K.S. et al. Chibber. Risk of violence from the man involved in the pregnancy after receiving or being denied an abortion. BMC Med, 12(144), 2014.

[271] Jennifer A. Russo, Elizabeth Miller, and Melanie A. Gold. Myths and misconceptions about long-acting reversible contraception (larc). Journal of Adolescent Health, 52, 2013.

[272] Marjorie R. Sable, John C. Spencer, Joseph W. Stockbauer, Wayne F. Schramm, Vicky Howell, and Allen A. Herman. Pregnancy wantedness and adverse pregnancy outcomes: Differences by race and medicaid status. Family Planning Perspectives, 29(2):76–81, 1997.

[273] Sally Sadoff, Anya Samek, and Charles Sprenger. Dynamic Inconsistency in Food Choice: Experimental Evidence from Two Food Deserts. The Review of Economic Studies, 87(4):1954–1988, 05 2019.

[274] Unnati Rani Saha and Arthur van Soest. Contraceptive use, birth spacing, and child survival in matlab, bangladesh. Studies in Family Planning, 44(1):45–66, 2013.

[275] H.M. Salihu, M.H. Aliyu, and B.J. et al. Pierre-Louis. Levels of excess infant deaths attributable to maternal smoking during pregnancy in the united states. Matern Child Health, 7, 2003.

[276] Sue Santa. Skating on this ice: Nhl owners and players clash over free agency. Wash U. L. Rev., 915, 1992.

[277] F. L. Santos, S. S. Esteves, A. da Costa Pereira, W. S. Yancy Jr, and J. P. L. Nunes. Systematic review and meta-analysis of clinical trials of the effects of low carbohydrate diets on cardiovascular risk factors. Obesity Reviews, 13(11):1048–1066, 2012.

[278] Pedro H.C. Sant'Anna and Jun Zhao. Doubly robust difference-in-differences estimators. Journal of Econometrics, 219(1):101–122, 2020.

[279] Leonard R. Sayles. Behavior of industrial work groups prediction and Control. Wiley, 1958.

[280] Stephan Seiler, Anna Tuchman, and Song Yao. The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects. Journal of Marketing Research, 58(1):22–49, 2021.

[281] Claudia Senik. Wage Satisfcaction and Reference Wages, pages 1–13. Springer International Publishing, Cham, 2020.

[282] Fredrik Serenius, Uwe Ewald, Aijaz Farooqi, Vineta Fellman, Maria Hafström, Kerstin Hellgren, Karel Maršál, Andreas Ohlin, Elisabeth Olhager, Karin Stjernqvist, Bo Strömberg, Ulrika Ådén, Karin Källén, and for the Extremely Preterm Infants in Sweden Study Group. Neurodevelopmental Outcomes Among Extremely Preterm Infants 6.5 Years After Active Perinatal Care in Sweden. JAMA Pediatrics, 170(10):954–963, 10 2016.

[283] P.S. Shah, T. Balkhair, and A. et al. Ohlsson. Intention to Become Pregnant and Low Birth Weight and Preterm Birth: A Systematic Review. Matern Child Health, 15:205–216, 2011.

[284] Ankita Shukla, Abhishek Kumar, Arupendra Mozumdar, Kumudha Aruldas, Rajib Acharya, F. Ram, and Niranjan Saggurti. Association between modern contraceptive use and child mortality in india: A calendar data analysis of the national family health survey (2015-16). SSM - Population Health, 11:100588, 2020.

[285] L. Silver, S. Ng, Suzanne Ryan-Ibarra, L. S. Taillie, Marta Induni, D. Miles, J. Poti, and B. Popkin. Changes in prices, sales, consumer spending, and beverage consumption one year after a tax on sugar-sweetened beverages in berkeley, california, us: A before-and-after study. PLoS Medicine, 14, 2017.

[286] M. Simmonds, A. Llewellyn, C. G. Owen, and N. Woolacott. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. Obesity Reviews, 17(2):95–107, 2016.

[287] GK Singh and SM Yu. Infant Mortality in the United States, 1915-2017: Large Social Inequalities have Persisted for Over a Century. Int J MCH AIDS, 8(1):19–31, 2019.

[288] Wiedaad Slemming, Braimoh Bello, Haroon Saloojee, and Linda Richter. Maternal risk exposure during pregnancy and infant birth weight. Early Human Development, 99:31–36, 2016.

[289] Alec Smith. Lagged beliefs and reference-dependent utility. Journal of Economic Behavior Organization, 167:331–340, 2019.

[290] Sara Solnick and David Hemenway. Is more always better?: A survey on positional concerns. Journal of Economic Behavior and Organization, 37(3):373–383, 1998.

[291] Adam Sonfield, Kinsey Hasstedt, and Rachel Benson Gold. Moving forward: Family planning in the era of health reform. The Guttmacher Institute, 2014.

[292] A. J. Spector. Expectations, fulfillment, and morale. The Journal of Abnormal and Social Psychology, 52:51–56, 1956.

[293] Charles Sprenger. An endowment effect for risk: Experimental tests of stochastic reference points. Journal of Political Economy, 123(6):1456–1499, 2015.

[294] Steve Springer. The nhl: Gretzky's contract is worth $22.3 million over 16 years. Los Angeles Times, Jan 1990.

[295] Markus J. Steiner. Contraceptive effectiveness. JAMA, (15):1405–1407, 10 1999.

[296] Amanda J. Stevenson, Katie R. Genadek, Sara Yeatman, Stefanie Mollborn, and Jane A. Menken. The impact of contraceptive access on high school graduation. Science Advances, 7(19), 2021.

[297] Amy Stoddard, Colleen McNicholas, and Jeffrey F. Peipert. Efficacy and safety of long-acting reversible contraception. Drugs, 71:969–980, 2011.

[298] Julia Strasser, Liz Borkowski, Megan Couillard, Amy Allina, and Susan Wood. Long-acting reversible contraception: Overview of research and policy in the united states. Jacobs Institute of Women's Health, 2016.

[299] Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2):175–199, 2021. Themed Issue: Treatment Effect 1.

[300] Aparna Sundaram, Barbara Vaughan, Kathryn Kost, Akinrinola Bankole, Lawrence Finer, Susheela Singh, and James Trussell. Contraceptive failure in the united states: Estimates from the 2006–2010 national survey of family growth. Perspectives on Sexual and Reproductive Health, 49(1):7–16, 2017.

[301] Sumits T, Bennett R, and Gould J. Maternal risks for very low birth weight infant mortality. Pediatrics, 2, 1996.

[302] Janice Hopkins Tanne. Problems with contraception play big part in unplanned pregnancies, study says. BMJ, 336(7653):1095–1095, 2008.

[303] Yu-Li Tao, Hwei-Lin Chuang, and Eric S. Lin. Compensation and performance in major league baseball: Evidence from salary dispersion and team performance. International Review of Economics & Finance, 43:151–159, 2016.

[304] Lisa A Te Morenga, Alex J Howatson, Rhiannon M Jones, and Jim Mann. Dietary sugars and cardiometabolic risk: systematic review and meta-analyses of randomized controlled trials of the effects on blood pressure and lipids. The American Journal of Clinical Nutrition, 100(1):65–79, 05 2014.

[305] Andrea M. Teng, Amanda C. Jones, Anja Mizdrak, Louise Signal, Murat Genç, and Nick Wilson. Impact of sugar-sweetened beverage taxes on purchases and dietary intake: Systematic review and meta-analysis. Obesity Reviews, 20(9):1187–1204, 2019.

[306] John Thibaut. An experimental study of the cohesiveness of underprivileged groups. Human Relations, 3(3):251–278, 1950.

[307] James Trussell. Contraceptive failure in the united states. Contraception, 2004.

[308] James Trussell and Barbara Vaughan. Contraceptive failure, method-related discontinuation and resumption of use: Results from the 1995 national survey of family growth. Family Planning Perspectives, 31(2):64–93, 1999.

[309] Nick Turner, Kaveh Danesh, and Kelsey Moran. The evolution of infant mortality inequality in the united states, 1960–2016. Science Advances, 6(29):eaba5908, 2020.

[310] Amos Tversky and Daniel Kahneman. Loss Aversion in Riskless Choice: A Reference-Dependent Model*. The Quarterly Journal of Economics, 106(4):1039–1061, 11 1991.

[311] Arthur van Soest and Unnati Saha. Relationships between infant mortality, birth spacing and fertility in Matlab, Bangladesh. PLoS ONE, 13(4):695–713, 2018.

[312] Mark Walker and John Wooders. Minimax play at wimbledon. American Economic Review, 91(5):1521–1538, December 2001.

[313] Y. Claire Wang, Pamela Coxson, Yu-Ming Shen, Lee Goldman, and Kirsten Bibbins-Domingo. A penny-per-ounce tax on sugar-sweetened beverages would cut health and cost burdens of diabetes. Health Affairs, 31(1):199–207, 2012. PMID: 22232111.

[314] Joseph M. Weiler. Legal analysis of the nhl player's contract. Marq. Sports L. Rev., 3(1), 2009.

[315] Allen J. Wilcox. Birth Weight and Perinatal Mortality: The Effect of Maternal Smoking. American Journal of Epidemiology, 137(10):1098–1104, 05 1993.

[316] A.N. Wise and B.S. Meyer. International Sports Law and Business. Number v. 2 in International Sports Law and Business. Springer Netherlands, 1997.

[317] David A. Wiss, Nicole Avena, and Pedro Rada. Sugar addiction: From evolution to revolution. Frontiers in Psychiatry, 9:545, 2018.

[318] Janet M. Wojcicki and Melvin B. Heyman. Healthier choices and increased participation in a middle school lunch program: Effects of nutrition policy changes in san francisco. American Journal of Public Health, 96(9):1542–1547, 2006. PMID: 16873747.

[319] E. L. Wolfe, T. Davis, J. Guydish, and K. L. Delucchi. Mortality risk associated with perinatal drug and alcohol use in california. official journal of the California Perinatal Association, 25(2):93–100, 2005.

[320] World Bank. Taxes on sugar-sweetened beverages: International evidence and experiences. The World Bank, 2020.

[321] Nicholas H. Wright. Family Planning and Infant Mortality Rate Decline in the United States. American Journal of Epidemiology, 101(3):182–187, 03 1975.

[322] Bo Xi, Yubei Huang, Kathleen Heather Reilly, Shuangshuang Li, Ruolong Zheng, Maria T. Barrio-Lopez, Miguel A. Martinez-Gonzalez, and Donghao Zhou. Sugar-sweetened beverages and risk of hypertension and cvd: a dose–response meta-analysis. British Journal of Nutrition, 113(5):709–717, 2015.

[323] J. Yang, S. L. Carmichael, M. Canfield, J. Song, G. M. Shaw, and the National Birth Defects Prevention Study. Socioeconomic Status in Relation to Selected Birth Defects in a Large Multicentered US Case-Control Study. American Journal of Epidemiology, 167(2):145–154, 10 2007.

[324] Sara Yeatman, James M. Flynn, Amanda Stevenson, Katie Genadek, Stefanie Mollborn, and Jane Menken. Expanded contraceptive access linked to increase in college completion among women in colorado. Health Affairs, 41(12):1754–1762, 2022. PMID: 36469823.

[325] Todd R. Zenger. Explaining organizational diseconomies of scale in r&d: Agency problems and the allocation of engineering talent, ideas, and effort by firm size. Management Science, 40(6):708–729, 1994.

[326] Chen Zhen, Eric A. Finkelstein, James M. Nonnemaker, Shawn A. Karns, and Jessica E. Todd. Predicting the effects of sugar-sweetened beverage taxes on food and beverage demand in a large demand system. American Journal of Agricultural Economics, 96(1):1–25, 2014.

[327] Yichen Zhong, Amy H Auchincloss, Brian K Lee, and Genevieve P Kanter. The short-term impacts of the philadelphia beverage tax on beverage consumption. Am J Prev Med, 2018.

[328] Yichen Zhong, Amy H. Auchincloss, Brian K. Lee, Ryan M. McKenna, and Brent A. Langellier. Sugar-sweetened and diet beverage consumption in philadelphia one year after the beverage tax. International Journal of Environmental Research and Public Health, 17(4), 2020.

[329] Andrew Zuppann. The impact of emergency contraception on dating and marriage. Technical report, 2011.

Appendix A

Appendix for Chapter 1 - Can Expanding Contraceptive Access Reduce Adverse Infant Health Outcomes?

A.1 Appendix Figures

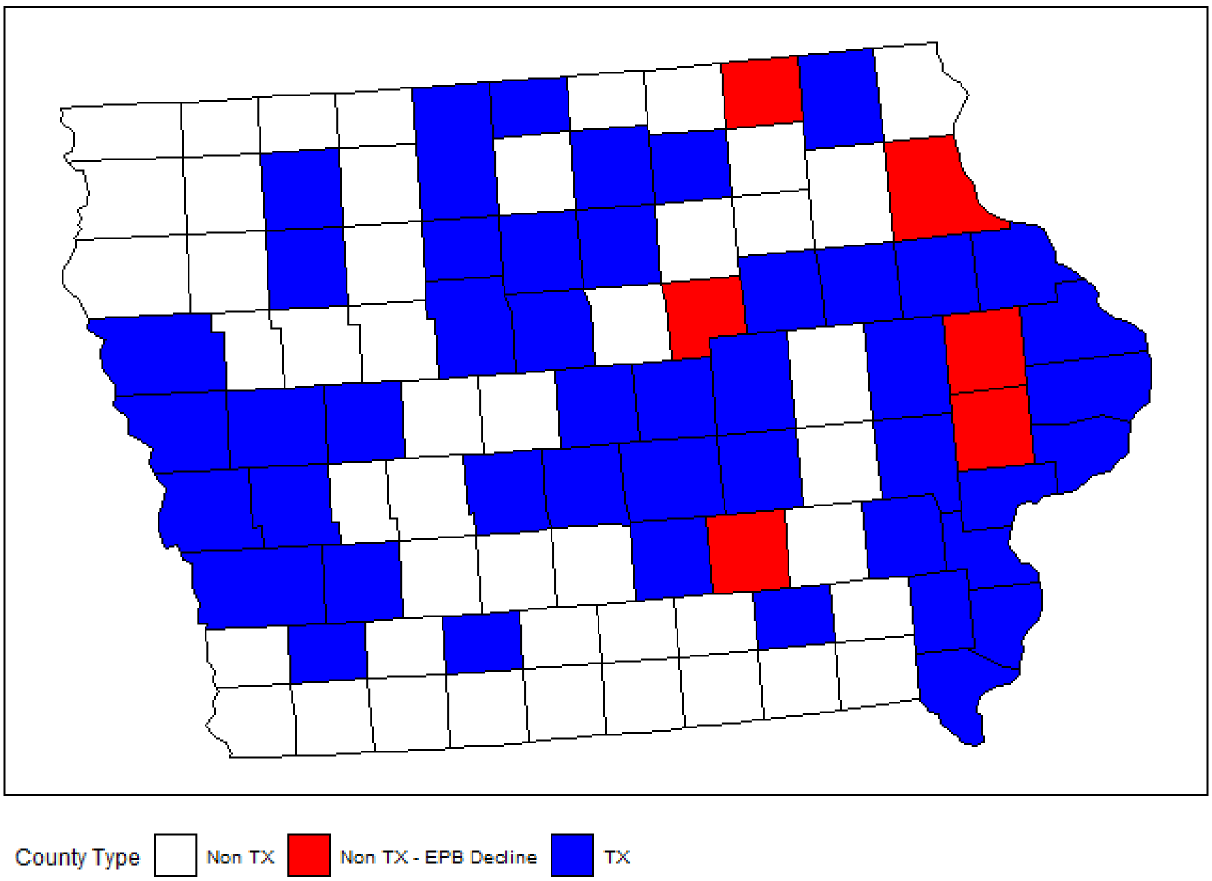
Table A.1: Goodman-Bacon Decomposition Table - TWFE Model

DD Comparison	Weight	Avg DD Est
Earlier T vs. Later C	.003	-.157
Later T vs. Earlier C	.004	.171
T vs. Never-treated	.993	-.484

T = Treatment; C = Comparison

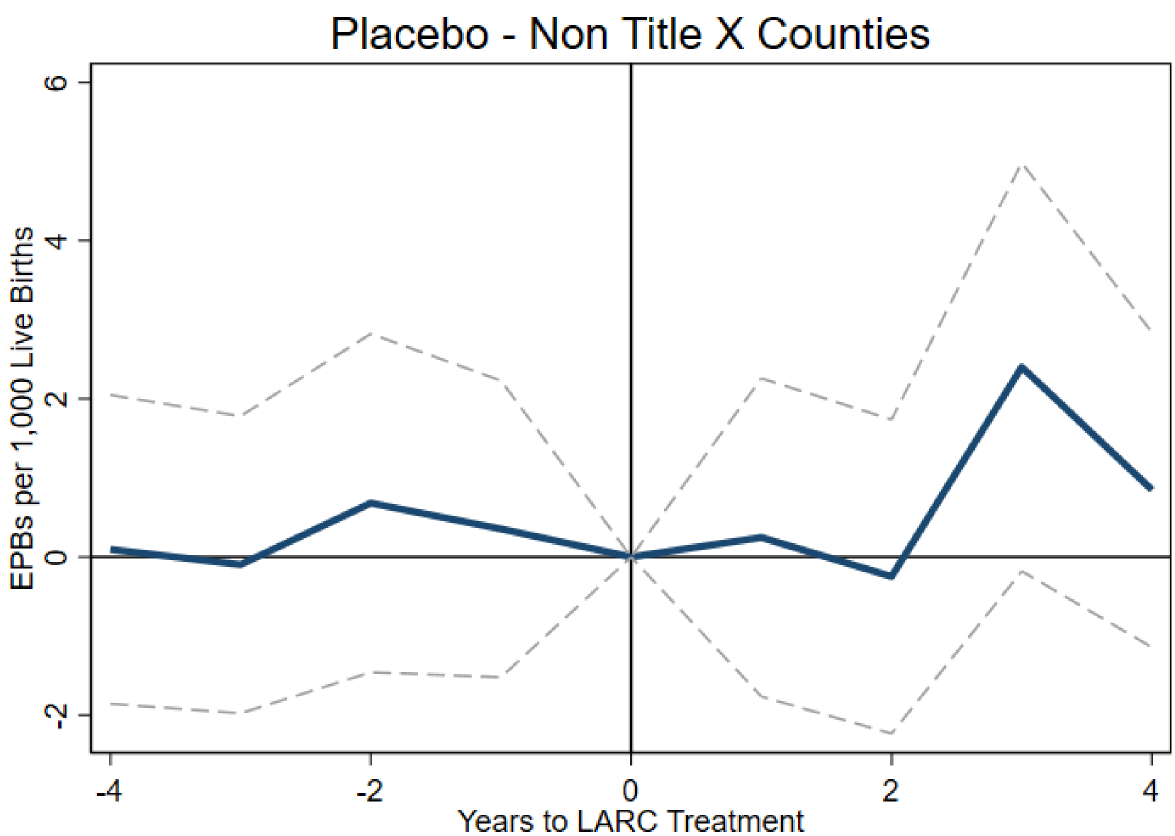
Notes: This table displays the weights from the regressions in Table 1 going to each type of comparison. The first row shows that only 0.3% of the weight is going to the Earlier Treated vs. Later Control comparison, where Iowa and St. Louis are considered treated and Colorado is the control. The second row shows that only 0.4% of the weight is going to the Later Treated vs. Earlier Control comparison, where Colorado is treated and Iowa and St. Louis are controls. Finally, the third row shows that 99.3% of the weight is going to the Treated vs. Never-Treated comparison, where Colorado, Iowa, and St. Louis are being compared to never-treated counties across the United States. This table was created using Goldring$_1 9' sddtimingpackageinStata$.

Figure A.1: Map of Untreated Iowa Counties Which Experienced Reductions in Extremely Preterm Births Following the Iowa Initiative to Reduce Unintended Pregnancies.



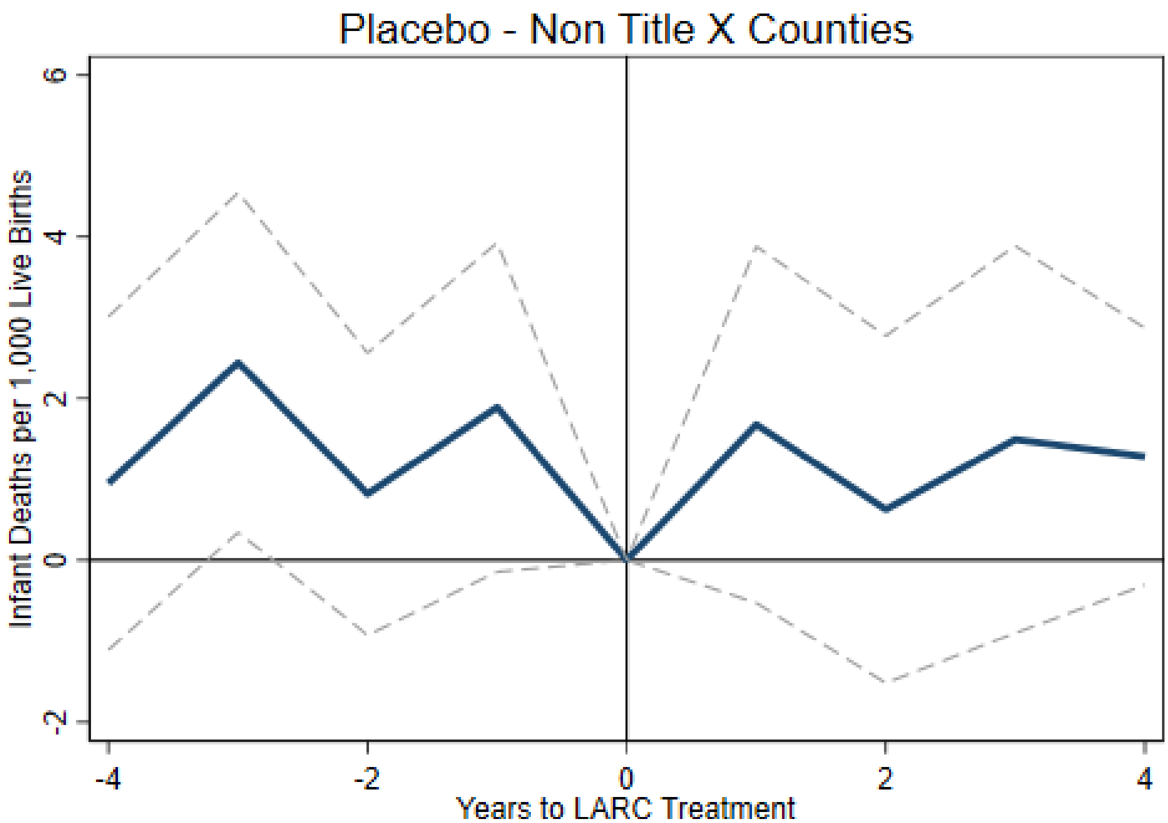
Note: This map displays the 99 counties of Iowa. The blue counties represent the 46 Iowa counties with a Title X family planning clinic. The white counties represent the 47 untreated Iowa counties which did not experience a decline in extremely preterm births following the Iowa Initiative to Reduce Unintended Pregnancies. The six red counties are untreated Iowa counties which did experience a reduction in extremely preterm births.

Figure A.2: Placebo Event-Study for Extremely Preterm Births - Non Title X Counties



Note: This figure displays estimates from a placebo event-study, which compares the extremely preterm birth rate in untreated counties in Colorado and Iowa with other counties across the U.S. which similarly do not have a Title X family planning clinic.

Figure A.3: Placebo Event-Study for Infant Mortality - Non Title X Counties



Note: This figure displays estimates from a placebo event-study, which compares the infant mortality rate in untreated counties in Colorado and Iowa with other counties across the U.S. which similarly do not have a Title X family planning clinic.

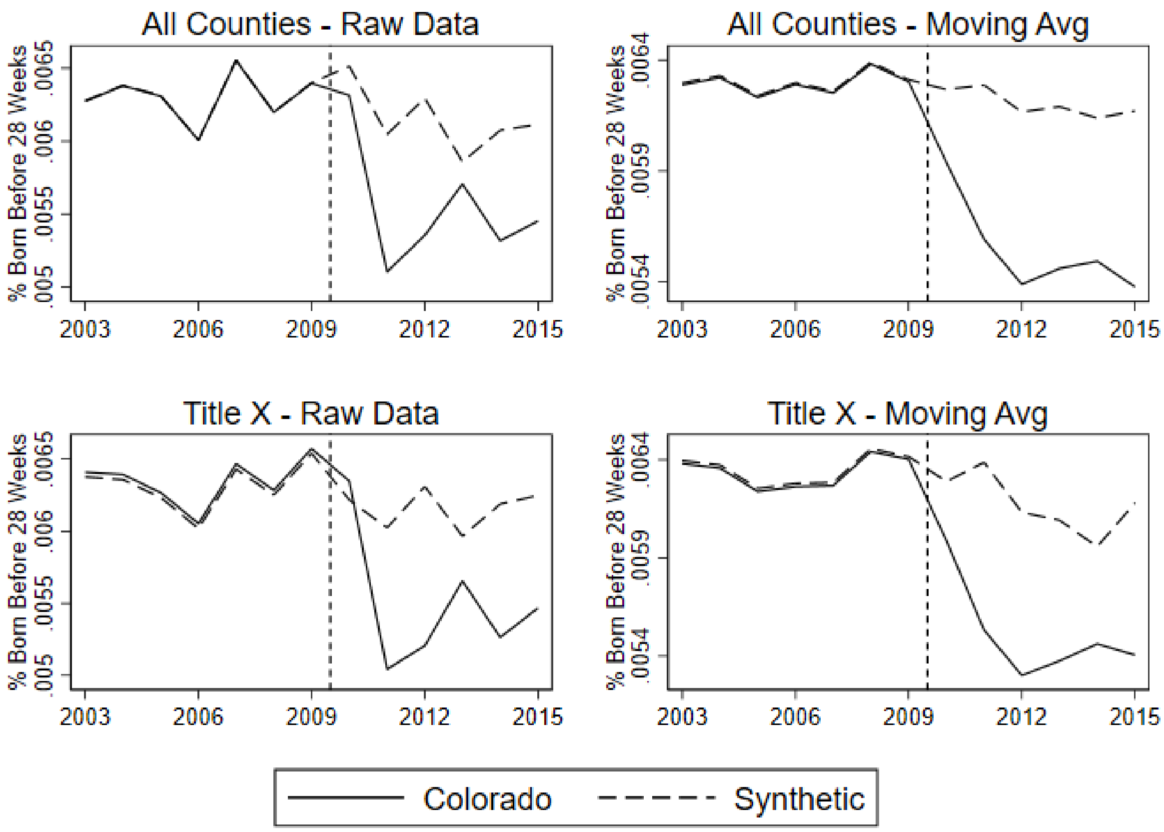
A.2 Synthetic Controls

A.2.1 Extremely Preterm Births

In order to get a better understanding of the separate effect of each intervention, and to more adequately address the concerning pretrends which appeared in the Iowa and St. Louis specification, I now reestimate the effect of these LARC interventions using the synthetic control method pioneered by [5] and [6]. For both Colorado and Iowa, where some counties were treated and some were not, I estimate the synthetic control specification both on the entire state and on the entire state that remains after dropping all of the counties which do not have a Title X clinic. If the LARC interventions were the reason behind the decline we saw in Table 1, then we should see larger impacts when the specifications only include Title X counties. Additionally, for each treated region, I estimate synthetic controls for both the raw EPB data as well as on three-year moving averages of the rates of EPB. I do this because these rates are inherently noisy, and this can cause synthetic controls to match on idiosyncratic noise than on the latent variables which are causing differences in trends.

Appendix Figure A.4 displays all four specifications for Colorado, while Appendix Figures A.5 and A.6 display the standard graphs for inference with synthetic control specifications. Beginning with the top left, which estimates the model using all counties in Colorado and raw data instead of moving averages, there is a close match prior to the CFPI. After 2009, there is a slight drop in the first year, but then a large decline in 2011, down 1.3 EPBs per 1,000 births from 2009. There is a slight rebound, but overall there still appears to be a large change in levels of between .5 and 1.0 EPBs per 1,000 births. When compared to the 49 placebo specifications, the ratio of post versus pretreatment root mean squared error for Colorado is the largest, more than double the next highest. Because of the noise that occurs in relatively rare outcomes like this one, the top right panel of Figure A.4 reestimates the same specification on a three-year moving average of the rate of EPB. Now, the change in levels is far more obvious, as there is a decline of about .9 EPBs per 1,000 live births by 2012 which shrinks slightly in the later years as the levels drop in

Figure A.4: Synthetic Control - Colorado

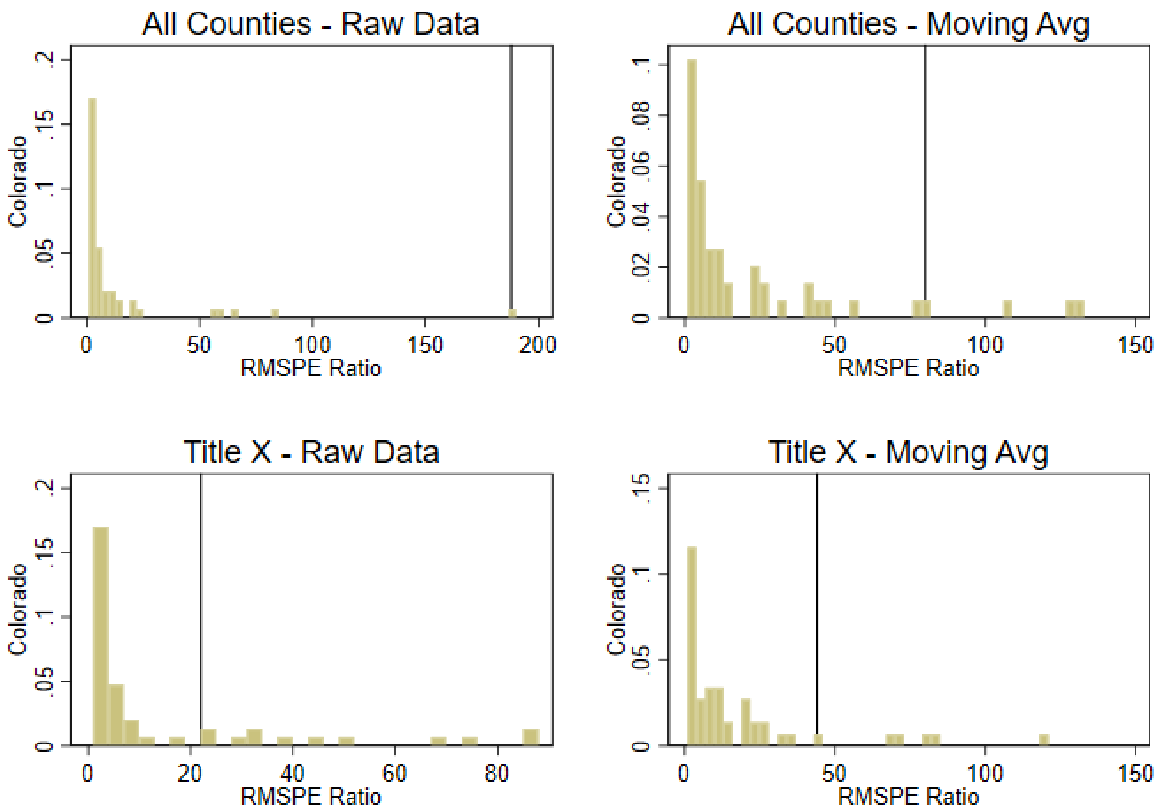


Note: This figure displays outcomes from applying the SCM to Colorado, estimating the effect of the CFPI on extremely preterm births. The top row estimates the SCM on all counties, while the bottom row only includes counties with a Title X clinic. The left column estimates the SCM on the raw data, while the right column uses a three-year moving average to reduce noise. p-values (moving from top left to bottom right) = .00, .06, .22, .1

the synthetic control as well.

The bottom left panel again estimates the same specification, this time dropping data from all counties in the United States which did not have a Title X clinic in 2008. In Colorado, around 92% of births occur in counties with such a clinic, so there is not a large difference between the top left and bottom left panels, but the treatment effect is in fact larger in the bottom panel as the rate of EPB declined by 1.5 per 1,000 live births in the bottom panel (compared to 1.3 in the top). This suggests that the reductions were largest in areas with Title X clinics, which is further evidence that the reduction we see was in fact caused by the CFPI. The story is similar in the

Figure A.5: Synthetic Controls - RMSPE Distributions - Colorado

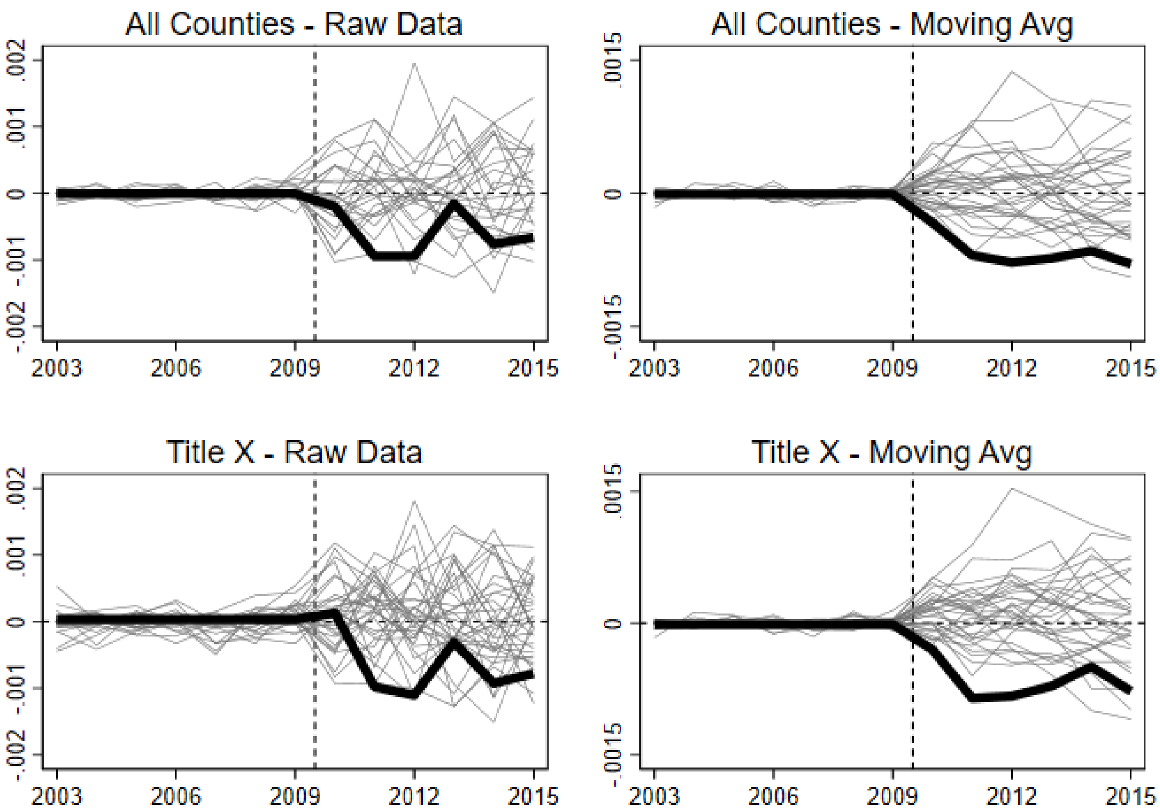


Note: This figure displays the distribution of root mean squared predicted error (RMSPE) ratios for Colorado compared with placebo ratios for each of the other 49 states. The top row estimates the SCM on all counties, while the bottom row only includes counties with a Title X clinic.

bottom right graph, which shows the three-year moving average for only counties with a Title X clinic. Again, there is a close pretreatment match and then a large decline of over 1.0 EPB between 2009 and 2011. The two main takeaways are that the decline which occurred in Colorado in the years following the CFPI did not occur in other states which had been evolving similarly up to that point, and that the treatment effect is larger in counties with Title X clinics than elsewhere.

Appendix Figure A.7 repeats this exercise for Iowa, and the results are similar though not quite as compelling. In all four graphs, there is a close pretreatment match between treated counties and their synthetic control. In 2009, in the raw data, there is a large decline of about 1.2 EPBs per 1,000 live births, followed by a rebound in 2010. This mirrors the experience of Colorado, where

Figure A.6: Synthetic Controls - Placebo Treatment Effects - Colorado

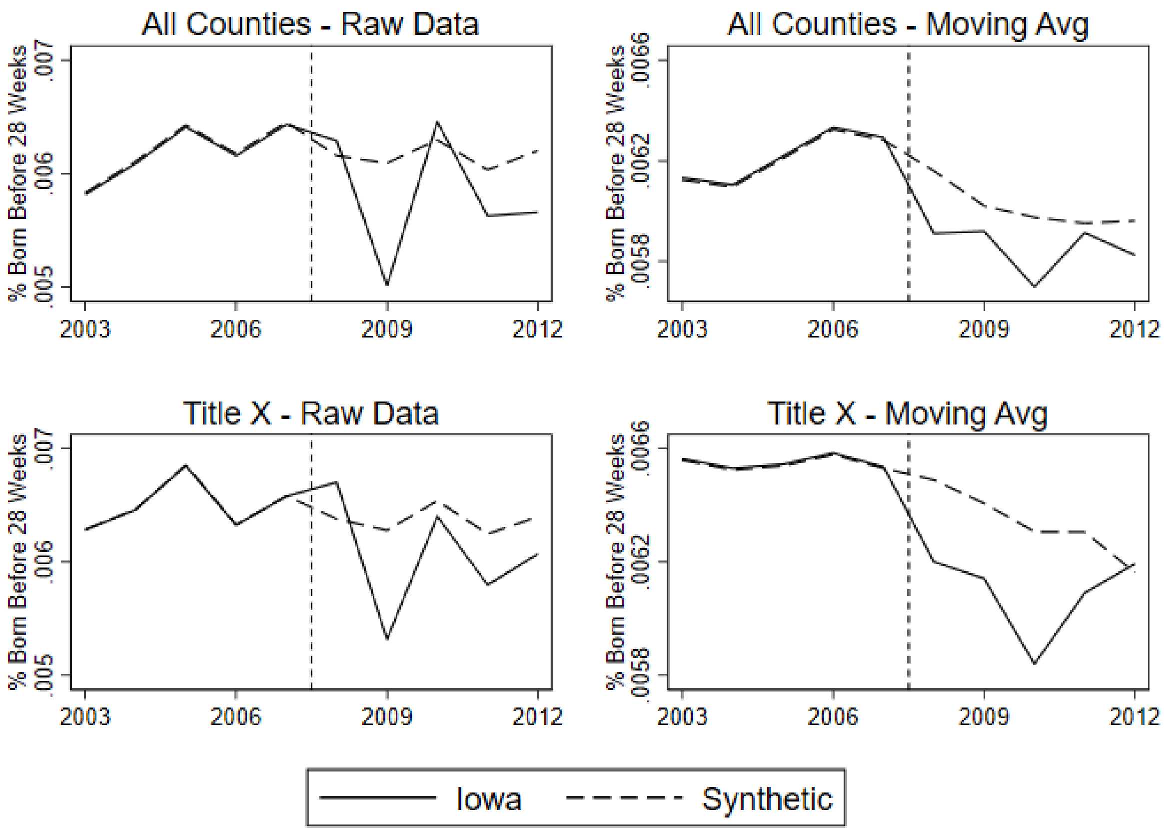


Note: This figure displays the difference between each state and its synthetic control for each period from 2003-2015. The bold line represents Colorado, while each of the other lines represents one of the 49 placebo states. The top row estimates the SCM on all counties, while the bottom row only includes counties with a Title X clinic.

there was a large decline in EPB in 2011, two years after the CFPI went into effect, followed by a smaller rebound. While the initial decline appears equally large for both the Title X and non Title X counties, the rebound is much larger in the top left, suggesting that this rebound effect was stronger in the untreated Iowa counties. Because of the smaller rebound in Title X counties, the moving average effect is much larger in the Title X counties than in the sample overall, which is consistent with a causal impact of the LARC intervention. Appendix Figures A.8 and A.9 display the respective graphs of the distribution of RMSPE ratios and placebo treatment comparisons. When conducting inference, none of the Iowa specifications is individually significant[1] , though

[1] The p-value on all counties estimated with raw data is .30, for the all county moving average the p-value is also

Figure A.7: Synthetic Control - Iowa



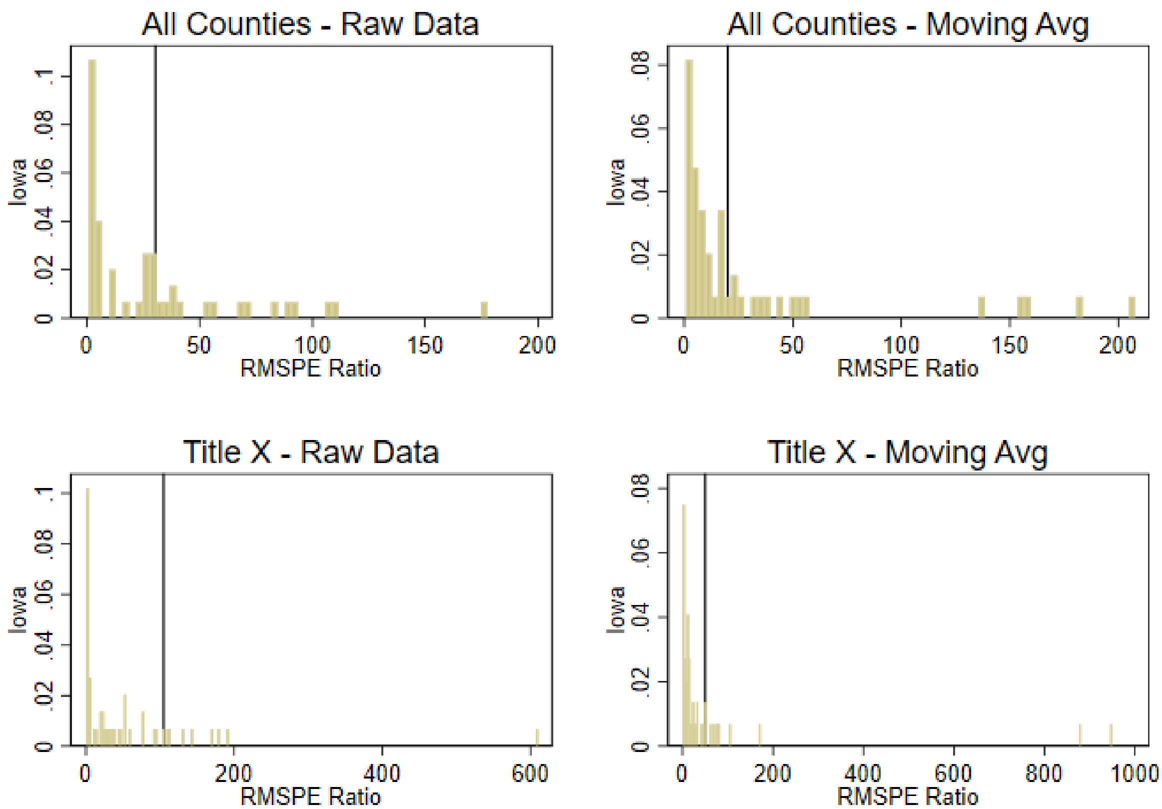
Note:This figure displays outcomes from applying the SCM to Iowa, estimating the effect of the IIRUP on extremely preterm births. The top row estimates the SCM on all counties, while the bottom row only includes counties with a Title X clinic. The left column estimates the SCM on the raw data, while the right column uses a three-year moving average to reduce noise. p-values (moving from top left to bottom right) = .30, .30, .16, .18

they all have p-values of .30 or smaller.

Finally, Figure A.10 displays synthetic control estimate for St. Louis. Since this is estimated at a local level, I compare St. Louis to other counties with at least 3,000 births per year, of which there are 242. Since there are no 'untreated' units within St. Louis, I only estimate the model on the raw data and the moving average, so the bottom half of Appendix Figure A.10 also includes the placebo treatment plot. As in Colorado and Iowa, St. Louis appears to be relatively steady in the few years leading up to treatment, before declining rapidly 1-2 years after the LARC intervention

.3, for the Title X raw data it is .16 and for the Title X moving average it is .18

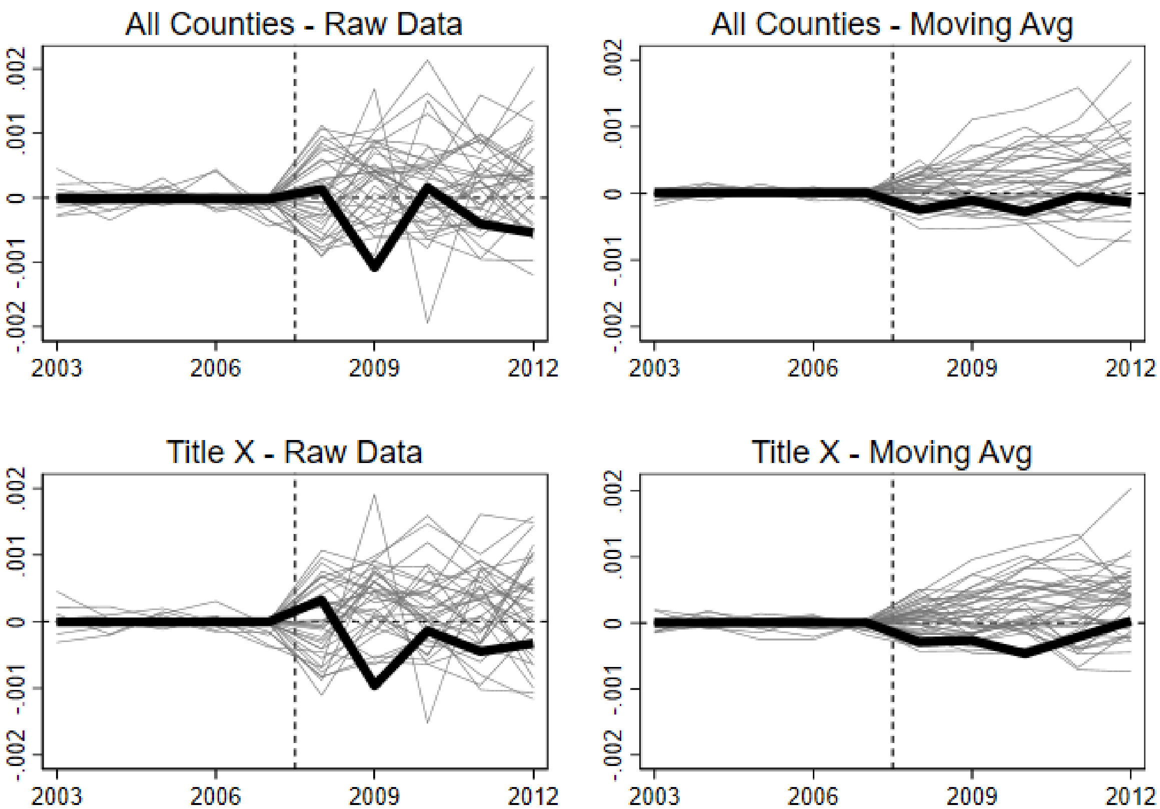
Figure A.8: Synthetic Controls - RMSPE Distributions - Iowa



Note: This figure displays the distribution of root mean squared predicted error (RMSPE) ratios for Iowa compared with placebo ratios for each of the other 49 states. The top row estimates the SCM on all counties, while the bottom row only includes counties with a Title X clinic.

takes place. Similar to Colorado and Iowa, there is a slight rebound after the drop, but overall there appears to be a substantial change in levels, from about 11 EPBs per 1,000 live births to somewhere between 8 and 9. Comparing the RMSPE ratio of St. Louis with the placebos returns a p-value of .02 on the raw data and .14 on the three year moving average. Taken together, however, the results across the three separate interventions provide evidence for a causal impact of LARC access on the rate of EPB's, with the largest impact occurring 1-2 years after treatment. In order for that not to be the case, some other factor would have had to cause large declines in EPBs in all three treated areas within two years of the LARC interventions, which seems unlikely.

Figure A.9: Synthetic Controls - Placebo Treatment Effects - Iowa

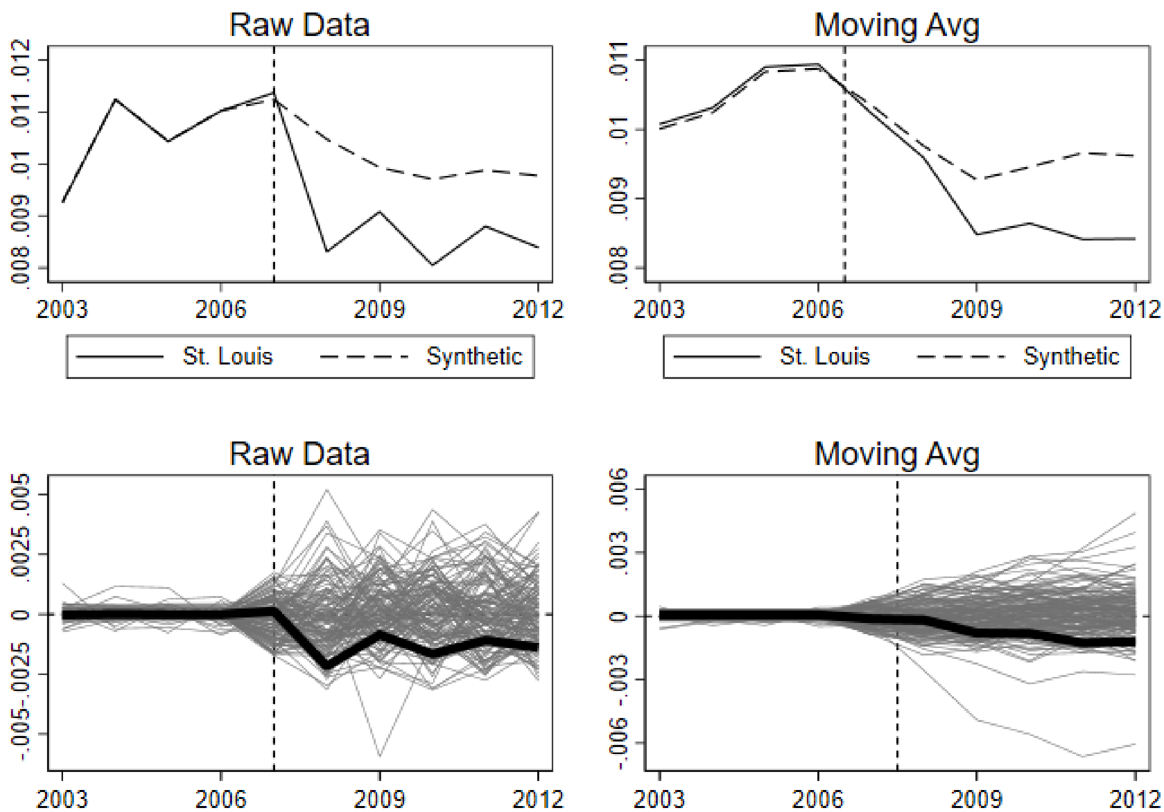


Note: This figure displays the difference between each state and its synthetic control for each period from 2003-2015. The bold line represents Iowa, while each of the other lines represents one of the 49 placebo states. The top row estimates the SCM on all counties, while the bottom row only includes counties with a Title X clinic.

A.2.2 Infant Mortality

In order to compare changes in infant mortality which are showing in Colorado and Iowa with other states who were evolving similarly in the pretreatment years, Appendix Figures A.11 and A.12 plot three-year moving averages in infant mortality in Colorado and Iowa against their synthetic controls and compares these treatment effects with placebo estimates for the other 48 states and Washington D.C. Beginning with Colorado and the top left graph in Figure A.11, the synthetic control matches closely in the pretreatment period, hovering between 6.0 and 6.3 without any large changes. Both groups begin to decline in 2009, though the drop in Colorado is substantially more

Figure A.10: Synthetic Control - St. Louis



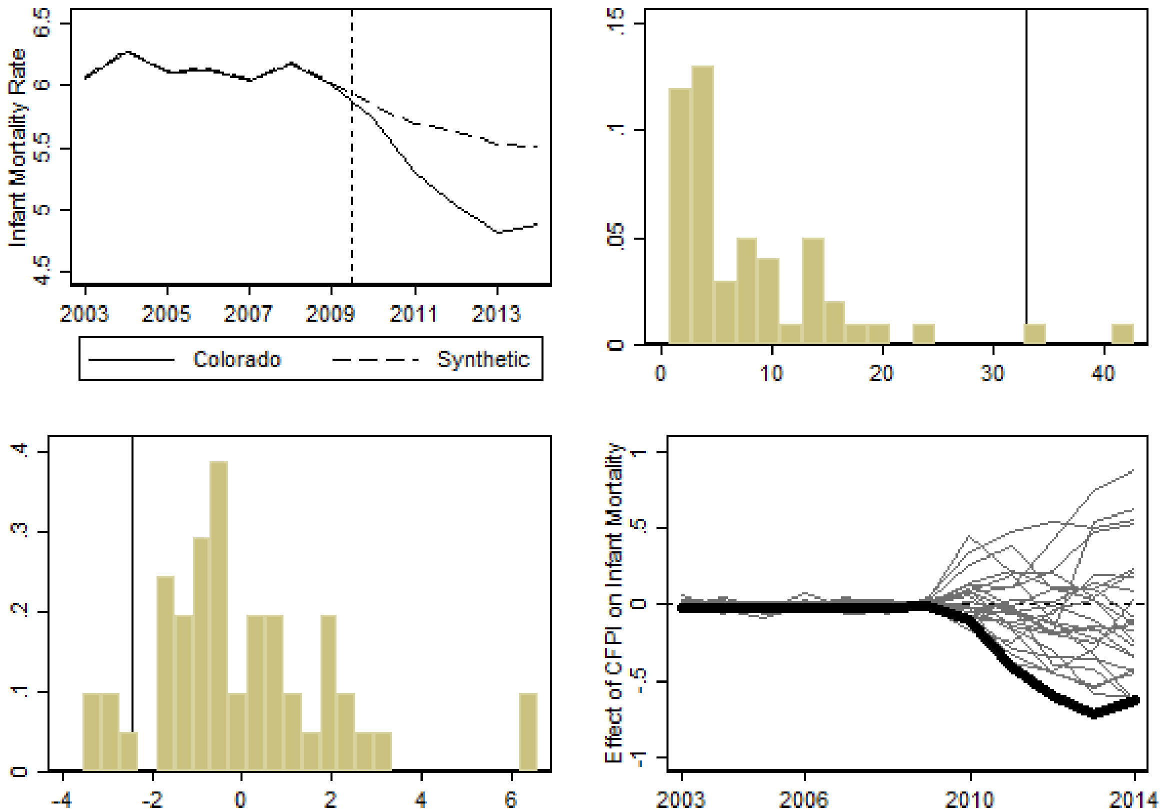
Note: This figure displays outcomes from applying the SCM to St. Louis as well as other counties with at least 3,000 births per year. The top left graph displays the synthetic control estimated on raw data, while the top right graph displays the same estimate on a three-year moving average. The bottom row displays the treated estimate compared with all of the placebo estimates which fit reasonably well in the pretreated period. The p-value for the raw data is .02, while on the weighted average it is .14

pronounced than its control. The divergence does not occur right away, which makes sense because of the staggered treatment of the CFPI as well as the lag between receiving a LARC and the avoidance of an unintended pregnancy. In 2010 there is only a difference of 0.1 infant deaths per 1,000 births, but this grows to 0.41 in 2011 and 0.60 in 2012, before remaining between 0.62-0.73 in 2013 and 2014. While infant mortality was declining after the CFPI, the decline in Colorado after its implementation was substantially larger than in the states that most closely matched Colorado's trends in the pretreatment period. The top-right graph in Figure A.11 plots the distribution of

RMSPE ratios of Colorado compared with the 50 placebo estimates. Colorado is the third largest ratio, corresponding with a p-value of .04, and is clearly out in the right tail of the distribution. The two placebos with larger ratios than Colorado are Kentucky and Oklahoma. In both cases, the ratio is larger than Colorado's because the states match in the pre-treatment period substantially more closely than Colorado, even though they show less variation in the posttreatment period. This is further illustrated by the bottom-left graph in Figure A.11, which plots the distribution of total treatment effects, or the sum of the difference between the treated group and it's synthetic control for the years 2010-2014. Here, Colorado has the 5th largest negative treatment effect at -2.5, with both Kentucky and Oklahoma sitting in the middle of the distribution at 1.5 and 0.8, respectively. The four states with more negative sums (Wyoming, Vermont, Virginia and Massachusetts) all match poorly in the pretreatment period and have much smaller RMPSE ratios than Colorado. Finally, the graph on the bottom-right of Figure A.11 plots the treatment effects of Colorado against the placebo estimates. In line with [6], I iteratively drop placebo estimates which match poorly in the pretreatment period. The graph in the figure display all the placebos with a root mean squared error in the pretreatment period no less than four times as large as Colorado's. This leaves 25 placebo estimates, of which Colorado is clearly the most negative. Overall, the decline in infant mortality in Colorado after the CFPI appears to be much larger than what could have been expected to happen by chance, and the staggered timing of the drop fits closely with what would be expected if it were caused directly by the CFPI.

Looking at Iowa in Figure A.12, there is also a clear decline in infant mortality around the time of the IIRUP, but it is much more closely matched by its synthetic control than Colorado. Both Iowa and its synthetic control remain between 5.2 and 5.6 from 2002-2007 dropping rapidly for two years and then rebounding slightly. Overall, Iowa still declines by more than its control by between 0.1-0.2 infant deaths per 1,000 births for 2008-2010 before this rebound. After this, Iowa continues to decline while its synthetic control remains steady at around 5. Since Title X clinics in Iowa were still giving out free LARCs in 2012, the continued drop could still be attributed to the IIRUP even after the synthetic control rebounded and then remained steady. Iowa displays the 35th

Figure A.11: Synthetic Control - Colorado

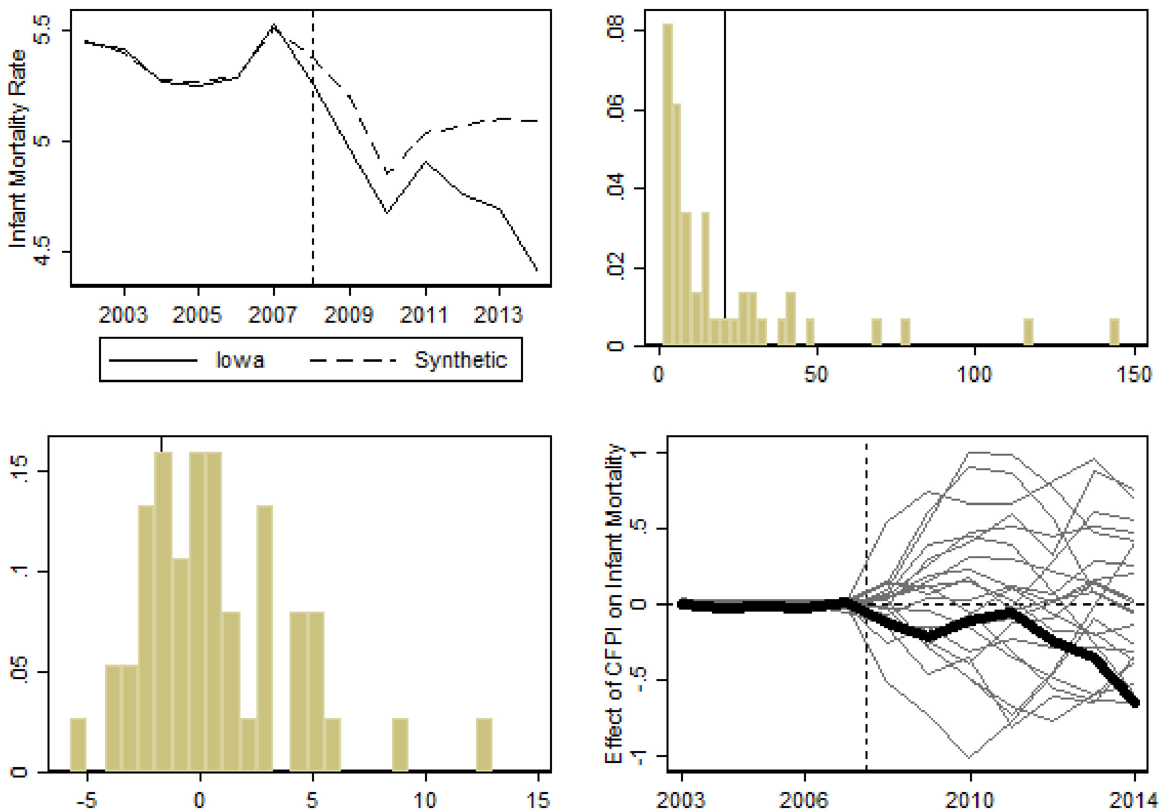


Note: This figure compares synthetic control outcomes for Colorado against placebo specifications when measuring the impact of the CFPI on infant mortality. The top-left graph plots the evolution of infant mortality in Colorado versus its synthetic control. The top-right graph plost the distribution of RMSPE ratios with a vertical line where the true treatment effect lies, with one extreme outlier dropped. The bottom-left graph displays the distribution of total treatment effects, which are the sum of the difference between the designated treatment group and its synthetic control for the posttreatment period. Finally, the bottom-right graph displays the difference between evolution of the actual treated group versus its synthetic control in black against all placebos which match closely in the pretreatment period in grey.

largest RMSPE ratio out of 50 for a p-value of .28, while its total treatment effect of -1.73 is the 13th most negative. Finally, comparing trends in Iowa against placebos which fit well prior to 2008 shows a modest and insignificant decline. While the results for Iowa are not nearly as compelling as those for Colorado, the fact that both states show improvements in infant mortality shortly after increasing LARC access to low-income women suggests that there is a causal relationship. This interpretation is supported by the fact that Colorado saw a larger decline as the intervention in

Colorado occurred on a larger scale than the one in Iowa.

Figure A.12: Synthetic Control - Iowa



Note: This figure compares synthetic control outcomes for Iowa against placebo specifications when measuring the impact of the IIRUP on infant mortality. The top-left graph plots the evolution of infant mortality in Iowa versus its synthetic control. The top-right graph plots the distribution of RMSPE ratios with a vertical line where the true treatment effect lies, with one extreme outlier dropped. The bottom-left graph displays the distribution of total treatment effects, which are the sum of the difference between the designated treatment group and its synthetic control for the posttreatment period. Finally, the bottom-right graph displays the difference between evolution of the actual treated group versus its synthetic control in black against all placebos which match closely in the pretreatment period in grey.

Appendix B

Appendix for Chapter 2 - Salary Disclosure and Individual Effort: Evidence from the National Hockey League

Table B.1: Linkage Rates Between Salary and Statistical Data

Year	Matched	Salary	Games	Stats Only	Games	Salary Only	Salary
1989-90	453	200,715	56.4	199	14.0	0	N/A
1990-91	465	236,329	53.4	193	16.1	3	83,385
1991-92	485	309,807	51.8	222	17.3	4	139,210
1992-93	542	337,019	51.0	171	38.3	6	143,704
1993-94	558	495,777	57.8	231	16.7	4	232,223
1994-95	574	372,798	33.6	162	9.6	5	197,623
1995-96	547	824,876	55.5	229	18.3	7	458,165
1996-97	695	785,199	47.8	73	35.4	10	346,396
1997-98	576	1,052,768	57.0	185	15.0	7	483,853
1998-99	659	1,116,316	51.5	153	15.7	5	581,326

Note: This table displays linkage rates across the salary data from www.markerzone.com and the statistical performance data from www.hockey-reference.com. For each year from 1990-1999, the table displays the number of players who are linked across both datasets, as well as average salaries and number of games played in columns two through four. Columns five and six display the total number of players who appear only in the statistical data along with their average number of games played. Finally, columns seven and eight display the number of players who appear only in the salary data along with their average salaries.

Table B.2: Lasso Regression Predicting Salary Based on Performance

	(1)
	Log Salary
Goals*Assists	0.000461**
	(0.000161)
Shots on Goal	0.00196***
	(0.000519)
Games Played	-0.00317**
	(0.00104)
Goals	-0.00257
	(0.00598)
Goals Squared	-0.000220
	(0.000160)
Assists	-0.000799
	(0.00337)
Plus-Minus	-0.000694
	(0.00134)
Penalty Minutes	0.000234
	(0.000246)
Age	0.0622***
	(0.00689)
Age Fixed Effects	Y
Position Fixed Effects	Y
Observations	453

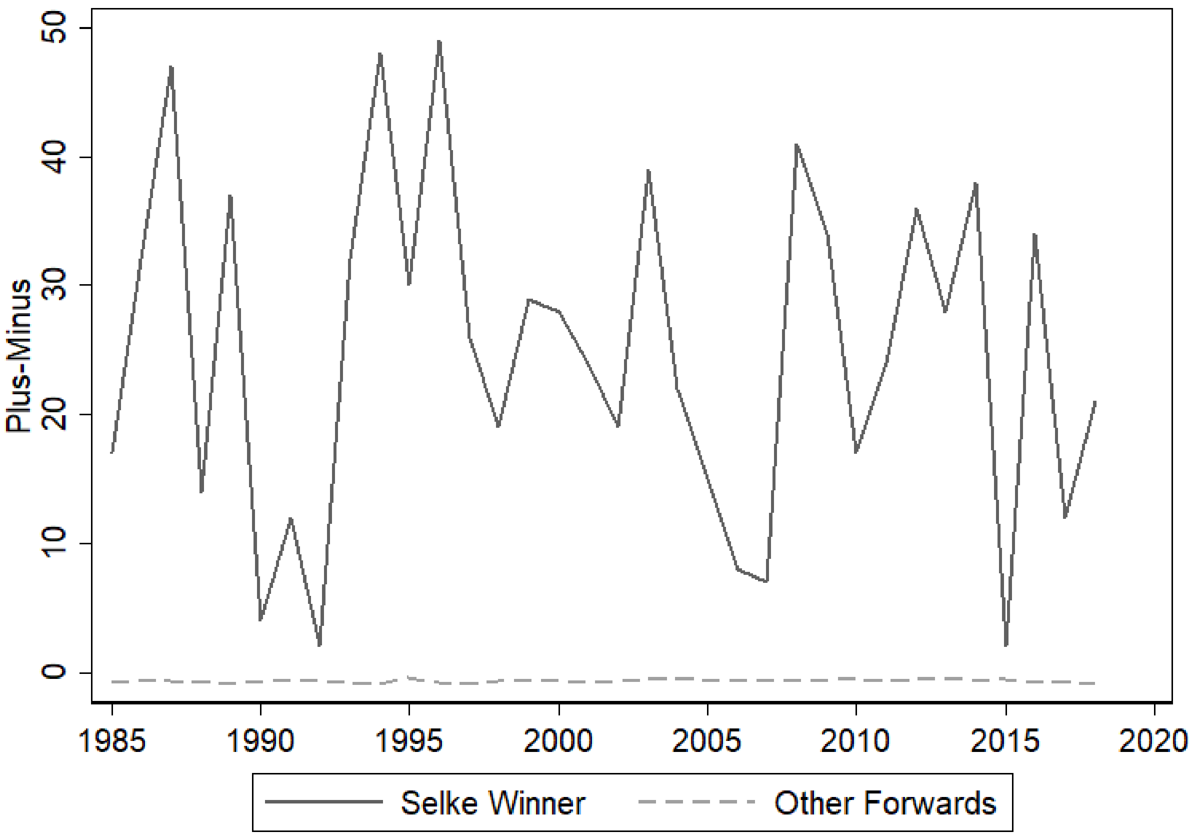
Note: This table displays regression coefficients from the lasso specification which is used to predict player salary and determine whether each player is 'overpaid' or 'underpaid' using salary data from www.markerzone.com and statistical data from www.hockey-reference.com. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.3: Underpaid/No Raise Player Statistics by Season

	1990	1991	1992	p-value
Log Salary	11.96	12.07	12.21	.00
Goals	13.40	12.05	13.80	.56
Assists	23.64	21.65	19.68	.24
Points	37.03	33.70	33.48	.58
+/-	.59	2.23	-.093	.45
PIMs	74.69	85.77	85.65	.44
Shots	109.07	103.96	110.34	.79
Shooting %	11.95	11.30	10.98	.62
Age	26.43	26.61	24.85	.00
Observations	86	85	98	.

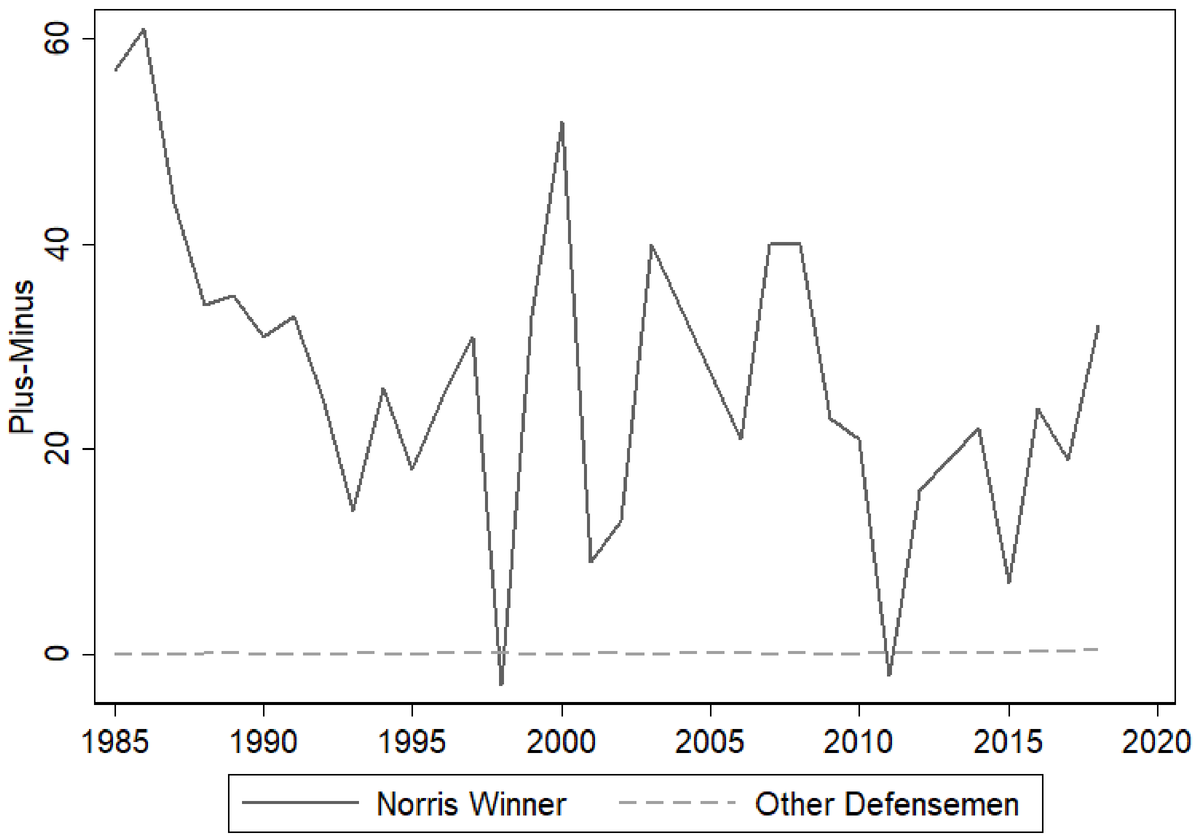
Note: This table compares statistical means for Underpaid/No Raise players in 1990 with the same group in the two subsequent seasons.

Figure B.1: Plus-Minus Comparison - Selke Winner vs. Other Forwards



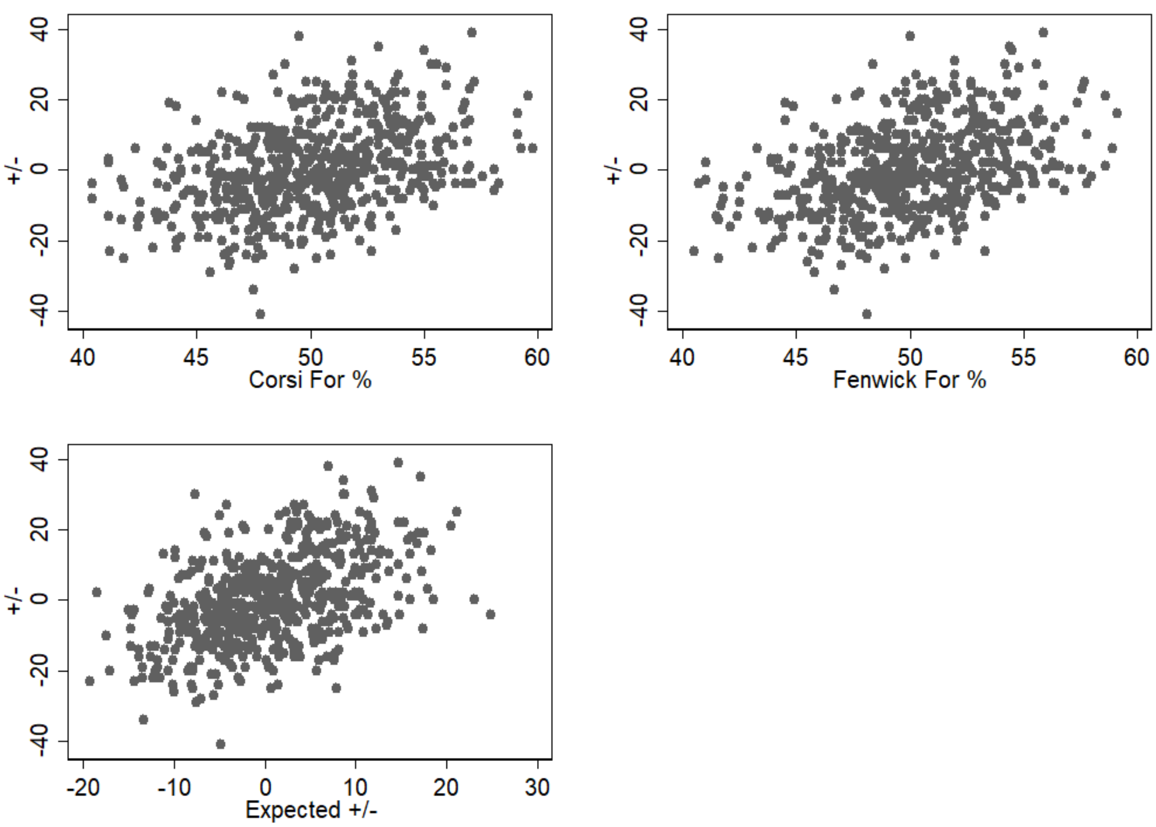
Note: This figure displays the plus-minus of the Selke Trophy winner compared to the mean for forwards in each season, using statistical data from hockey-reference.com

Figure B.2: Plus-Minus Comparison - Norris Winner vs. Other Defensemen



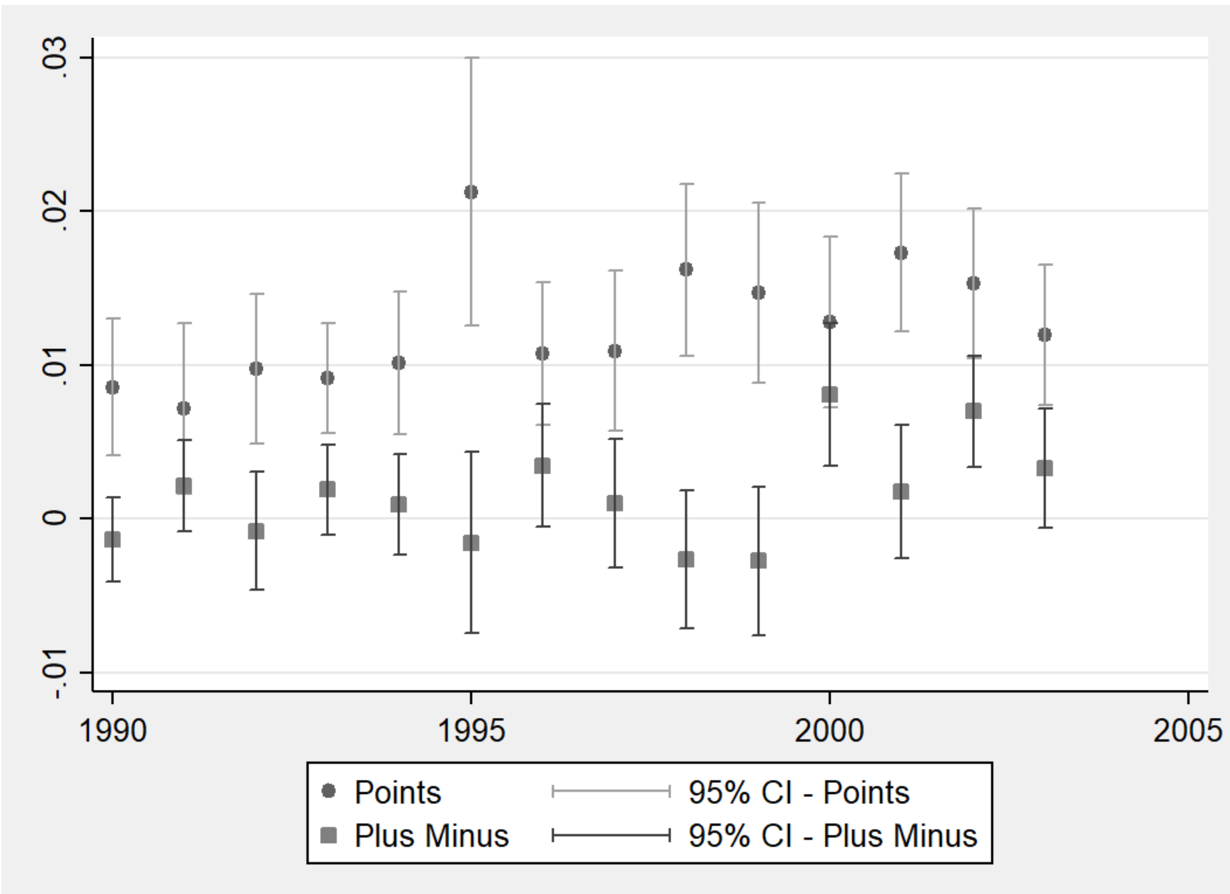
Note: This figure displays the plus-minus of the Norris Trophy winner compared to the mean for defensemen in each season, using statistical data from hockey-reference.com

Figure B.3: Plus-Minus Compared with Other Advanced Statistics - 2019



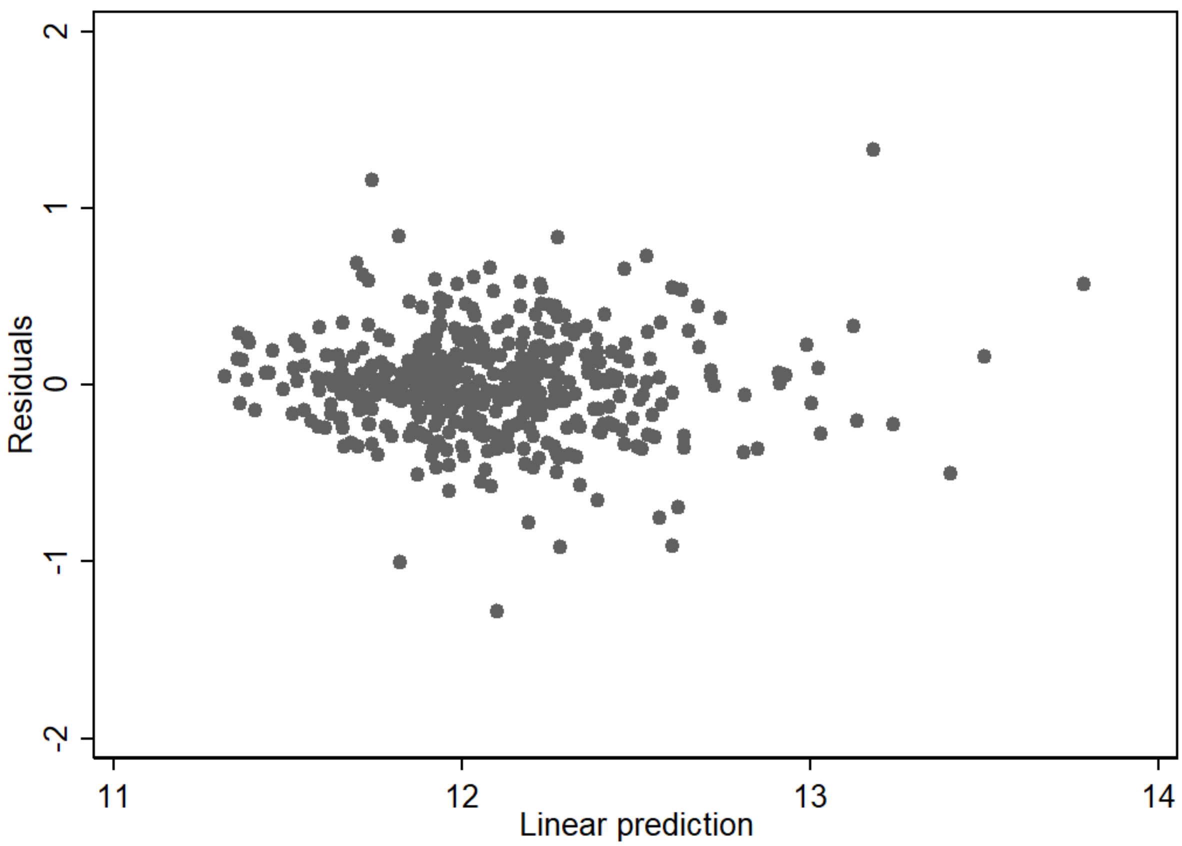
Note: This figure displays the correlation between plus-minus and the Corsi for %, the Fenwick for % and expected plus-minus, using statistical data from hockey-reference.com

Figure B.4: Coefficients on Points vs. Plus-Minus



Note: This figure displays the coefficient estimates for points and plus minus from hedonic regressions of log salary on statistics for each year from 1989-90 to 2003-04, using salary data from markerzone.com and statistical data from hockey-reference.com

Figure B.5: Residual Plot from Lasso Model - 1990



Note: This figure displays the residual plot from the hedonic lasso model from the 1989-90 season, using salary data from markerzone.com and statistical data from hockey-reference.com

B.1 Labor Relations in the NHL

Since June of 1967, the NHLPA has represented all NHL players. In 1970, the NHL became the first professional sports league in North America to use salary arbitration to settle disputes between players and teams. Previously, player contracts all included an *option clause*, allowing teams to unilaterally extend player contracts when they expired, at whatever salary they determined appropriate. Following this major change. though there were tensions between the NHLPA and the team owners, labor relations were relatively smooth through the rest of 70's and 80's.

In the early 90's, free agency as we now understand it did not exist within the NHL. If a player changed teams in free agency, the team which signed him was often required to send a player of similar calibre back to the original team. The most famous example of this occurred in 1991, when Brendan Shanahan left the New Jersey Devils to sign with the St. Louis Blues ([276]). In response, a league arbitrator ruled that the Blues would have to send Scott Stevens, their captain, to New Jersey in return. Both players were superstars in the league and both would eventually be inducted into the Hockey Hall of Fame. Because teams were afraid of having to give up valuable assets, they almost never signed marquis players in free agency. In 1992, when the league's Collective Bargaining Agreement (CBA) expired, the NHL had it's first ever player's strike. At issue were pensions, player bonuses, the arbitration process, revenue sharing and free agency. Among many changes to the new CBA were some liberalizations to the free agency process, making it somewhat easier for players to switch teams. Complete details of these liberalizations are included below. For the most part, players during this era that moved in free agency were fringe-level players whose current team was not interested in resigning them. Even with these liberalizations, [316] still write that the NHL had the most restrictive free agency policies of any of the four major North American sports leagues at that time.

Two years after this agreement was reached, the NHL had its first lockout, which ended up canceling 468 games in the 1994-95 season. Unlike the 1992 strike, this dispute was largely brought on by the owners, who were seeking to curb the rapid increase in salaries in the wake of salary disclosure. In the four years since disclosure, the inflation-adjusted league-wide mean salary had gone up 113% from $202,000 to $430,000 (in 1990 US Dollars). Owners also wanted to institute a salary cap (a maximum total salary outlay a team can pay in any given season) in order to keep teams from weaker markets competitive. In the 1990's, four teams went bankrupt and had to move move their team to bigger markets, even though this had only happened to five other teams since the end of World War II[1] Each time, owners cited a search for better financial conditions as a motivating factor behind the move. The owners were unable to get the NHLPA to agree to a salary cap and the issue was tabled until the next CBA expired.

When this happened in 2004, another lockout occurred which caused the entire 2004-05 season to be canceled. Eventually, the players' union agreed to a salary cap under the condition that they also enter into a revenue sharing agreement, where the salary cap is determined by league revenues in order to make sure the players continued to get a large percentage of the leagues total revenues.

B.1.1 Free Agency Changes in 1992

Before the 1992 agreement, the existing free agency groups were:

Group I - Players under the age of 24 who have not played five years of professional hockey

Group II - Players 30 years of age or over

Group III - All other players.

Under the eventual 1992 agreement, three new categories of free agency were created, in addition to the three that were already in existence:

(i) - A player who has completed ten professional seasons and did not earn more than the average league salary in his last contract year. The player can elect once in his career to become and unrestricted free agent at the end of his contract;

(ii) - A player who is 25 years of age or older and has completed three professional seasons and has not played in more than 80 NHL games can become an unrestricted free agent. The number

[1] In 1993, the Minnesota North Stars moved to Texas and become the Dallas Stars. In 1995, the Quebec Nordiques moved to Denver and became the Colorado Avalanche. In 1996, the Winnipeg Jets became the Arizona Coyotes, while in 1997, the Hartford Whalers became the Carolina Hurricanes.

of games played by a goaltender to become eligible under this category is to be determined by the mediation committee;

(iii) - A defected player, defined in the CBA as "a player not unconditionally released released by his NHL club, (a) whose contract with the club has not been completely fulfilled and who contracts with a club in a league not affiliated with the NHL or such a league itself or with any other pro hockey club for a period including any part of the unfulfilled portion of his NHL contract; or (b) as to whom a club holds negotiations rights though the player was never under contract to any NHL club, where the player has contracted on contracts with such an unaffiliated club."

According to [314], the vast majority of players remained in the original three categories and were thus not immediately impacted by the change.

B.2 Was There an Endogenous Midseason Response to Salary Disclosure?

To address whether the group sorting in the paper was affected by an endogenous midseason response by some players, I have gathered statistics from NHL.com, which allows users to view statistics over a given date range. This allowed me to gather performance statistics for the 1989-90 season, both before and after January 30. I was able to to do this for each player who appears in the main regressions who played in both the pre and post disclosure portion of the 1989-1990 season (this includes 351 of the 362 players from the main regressions). For each player, I calculate the change in each statistic (goals, assists, shots, plus-minus) on a per-game basis from the portion of the season played under pay secrecy to the games played after disclosure. For example, a player who scored 20 goals in 40 games before January 30, but increased their per-game production after disclosure to 15 goals in 20 games after January 30 would have a change in goals per game of .25 ($\frac{15}{20} - \frac{20}{40} = .75 - .5 = .25$). I then test whether receiving a raise in the summer of 1990 is correlated with a change in performance before and after disclosure in Table C.1. As in the main text, age and position fixed effects are included in each regression and standard errors are clustered at the team level.

The top panel of Table C.1 estimates whether receiving a raise at all (regardless of whether

Table B.1: Was There and Endogenous Midseason Response to Disclosure?

Panel 1 - Any Raise	(1) Big Raise	(2) Big Raise	(3) Big Raise	(4) Big Raise	(5) Big Raise
Goals per Game Change	0.152				0.0725
	(0.231)				(0.205)
Assists per Game Change		-0.224			-0.252
		(0.138)			(0.155)
Shots per Game Change			0.0306		0.0438
			(0.0698)		(0.0680)
+/- per Game Change				-0.0207	0.0210
				(0.0657)	(0.0797)
Panel 2 - Underpaid/Raise	**(1) UR**	**(2) UR**	**(3) UR**	**(4) UR**	**(5) UR**
Goals per Game Change	0.0831				-0.0960
	(0.225)				(0.184)
Assists per Game Change		-0.217			-0.304*
		(0.117)			(0.140)
Shots per Game Change			0.0671		0.102
			(0.0600)		(0.0571)
+/- per Game Change				0.0154	0.0745
				(0.0751)	(0.0909)
Panel 3 - Underpaid/No Raise	**(1) UN**	**(2) UN**	**(3) UN**	**(4) UN**	**(5) UN**
Goals per Game Change	0.189				0.299
	(0.183)				(0.217)
Assists per Game Change		0.165			0.218
		(0.111)			(0.118)
Shots per Game Change			-0.0166		-0.0647
			(0.0351)		(0.0424)
+/- per Game Change				0.0185	-0.0344
				(0.0464)	(0.0574)
Observations	352	352	352	352	352

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a player was eventually deemed to be over or underpaid) is correlated with a shift in performance post-disclosure. In one sense, this can be thought of as a test for recency bias, or whether players who happen to play better than average right at the end of the season are rewarded because their recent good play is more salient to the organization when contracts are negotiated. The five columns of Table 2 test each of the four statistics individually before including them all in the specification in column 5. None of the coefficients are significant, and only the change in assists has a p-value of less than .5.

The real threat to identification here is that maybe some players were more savvy in their response to disclosure and changed their behavior immediately and were subsequently rewarded for it. That would cause selection bias into the Underpaid/Raise group for players who might then revert to the mean. Panels 2 and 3 of Table C.1 test for this possibility. In panel 2, whether a player received a raise is replaced on the left-hand side of the regression with whether they were eventually sorted into the UR (underpaid/raise) group. If the coefficient on goals is greater and the coefficient on plus-minus is smaller, then there would be concerns over selection in my main specifications. This does not appear to be the case, however, as the goals coefficient is actually smaller than in Panel 1, the assists and shots coefficients are relatively similar, and plus-minus is greater than in panel 1. If anything, the direction of the coefficients is the opposite of what we would expect to see from the identification threat above. When all of the coefficients are included in column 5, goals is weakly negative and plus-minus is weakly positive. This shows that, among players who got a raise, the underpaid players did not have more of a response in the last part of the season, but perhaps they showed more of a response than the underpaid players who did not get a raise?

Panel 3 tests for this possibility by investigating whether being sorted into the UN (underpaid/no raise group) is correlated with a performance change. If the goals coefficient were smaller and the plus-minus coefficient larger than in Panel 2, we would be concerned. If anything, however, the opposite is true. The goals change coefficient is larger in Panel 3 than in Panel 2 and the plus-minus coefficient is similar when estimated by itself and smaller when all four statistics are included. This means that if any group began making the behavior changes we see in the main text towards the end of the 1989-90 season, it is the underpaid/no raise group, which would actually bias the regressions in the main specifications of my paper towards zero. It's worth pointing out, however, that of the 24 coefficients in Table C.1, only the estimate on assists in column five of Panel 2 is significant, suggesting that the most likely explanation is that there was no substantial midseason response to disclosure.

B.3 Robustness Checks

In this section I demonstrate that the results in Table 2 are robust to various concerns about model specification, the threshold for determining whether someone received a large raise, and influential observations. Earlier, I intentionally left out team fixed effects from the specification because I am more concerned with how underpaid someone is compared to the league overall as opposed to within their own team. However, if all of the underpaid players are clustered onto a few teams, and we know that the performance of these lower paying teams suffered, that could be what is driving the main results. To address this, Table D.1 re-estimates Equation 1 but with team fixed effects included, and the estimates are virtually unchanged. It appears that underpaid players who did not receive a raise in the summer of 1990 reallocated their effort from defense to offense both within teams and across the entire league. Table D.2 reestimates the specification testing reference-dependent utility with team fixed effects, and if anything, the results are slightly larger and more precise.

Another possible concern is that maybe the results hinge upon the inclusion of a variable that the lasso model dropped. maybe something that wasn't important for predicting salary in 1989-90 was important for understanding how the players effort changed from one year to the next. To address this, Table D.3 re-estimates the entire model but using the fully specified version of the wage regression, including every statistical variable I have access to. Again, the results are virtually unchanged. If anything, they are slightly more precise.

Additionally, the choice to count everything over 20% as a big raise was somewhat arbitrary.

Table B.2: Team Fixed Effects - 1990-91

	(1) Goals	(2) Shots	(3) Assists	(4) +/-	(5) Adj. +/-
Underpaid/No Raise	10.01	26.73	4.589	-9.154	-23.75
	(2.960)	(16.03)	(4.982)	(7.165)	(9.283)
Underpaid/Big Raise	-1.791	-25.71	-8.529	5.171	15.49
	(3.631)	(10.30)	(3.345)	(4.535)	(5.993)
Overpaid/No Raise	-1.078	-5.925	-4.080	12.31	17.47
	(4.753)	(26.66)	(8.368)	(3.830)	(15.41)
Overpaid/Big Raise	1.283	31.08	7.023	7.125	-1.181
	(3.667)	(31.66)	(8.204)	(5.026)	(11.54)
Observations	362	362	362	362	362

Note: This table displays regression coefficients comparing changes in statistical output for the four different groups from 1989-90 to 1990-91 including team fixed-effects, using salary data from www.markerzone.com and statistical data from www.hockey-reference.com

Table B.3: Reference-Dependece - Team Fixed Effects - 1990-91

	(1) Goals	(2) Shots	(3) Assists	(4) +/-	(5) Adj. +/-
UN - Below Avg	5.697	7.379	-3.274	-8.040	-10.46
	(3.083)	(22.23)	(5.868)	(7.555)	(10.36)
UN - Above Avg	14.22	45.63	12.27	-10.24	-36.74
	(2.703)	(15.03)	(7.053)	(10.89)	(9.593)
Underpaid/Big Raise	-1.772	-25.63	-8.493	5.166	15.43
	(3.653)	(10.73)	(3.515)	(4.536)	(6.108)
Overpaid/No Raise	-0.872	-4.998	-3.703	12.26	16.84
	(4.721)	(26.44)	(8.175)	(3.947)	(15.32)
Overpaid/Big Raise	1.287	31.09	7.029	7.124	-1.192
	(3.635)	(31.46)	(8.186)	(5.011)	(11.36)
Observations	362	362	362	362	362

Note: This table displays regression coefficients comparing changes in statistical output for five different groups (underpaid/no raise players on below average payroll teams, underpaid/no raise players on above average payroll teams, underpaid/big raise players, overpaid/no raise players, and overpaid/big raise players) from 1989-90 to 1990-91 using salary data from www.markerzone.com and statistical data from www.hockey-reference.com

Table B.4: Fully Specified Model

	(1) Goals	(2) Shots	(3) Assists	(4) +/-	(5) Adj. +/-
Underpaid/No Raise	10.35	31.08	5.835	-9.915	-26.10
	(2.719)	(13.47)	(4.203)	(7.878)	(8.238)
Underpaid/Big Raise	-0.626	-18.95	-5.504	9.151	15.28
	(3.046)	(9.495)	(3.317)	(4.898)	(5.383)
Overpaid/No Raise	-0.0997	1.246	-2.205	16.40	18.71
	(3.950)	(23.35)	(7.660)	(4.588)	(14.83)
Overpaid/Big Raise	3.254	44.52	11.92	6.575	-8.600
	(3.386)	(30.26)	(7.924)	(6.475)	(15.23)
Observations	362	362	362	362	362

Note: This table displays regression coefficients comparing changes in statistical output for the four different groups from 1989-90 to 1990-91 after using the fully specified predictive model, using salary data from www.markerzone.com and statistical data from www.hockey-reference.com

Tables D.4 and D.5 re-estimate the baseline model from Equation 1, but with the threshold changed to 15% and 25%, respectively. Changing the threshold to 15% causes seven players to move from UN to UR and 11 players to move from ON to OR. In Table D.5, the estimate on the UN group's adjusted plus-minus is slightly smaller, but it remains large and significant at .01. Changing the threshold to 25%, on the other hand, causes 15 players to move from UR to UN and 10 players to move from OR to ON. With a larger shakeup happening with this change, it's unsurprising that the coefficients changed more in this specification, though the important relationships still hold. UN players had a decline in adjusted plus-minus and increases in both goals and shots, though now the goals coefficient is slightly smaller and has a p-value of .07. Players in the UN group still clearly had increasing offense and declining defense, as in the baseline model.

Table B.5: Robustness Check - 15% Threshold

	(1) Goals	(2) Shots	(3) Assists	(4) +/-	(5) Adj. +/-
Underpaid/No Raise	10.27	21.78	4.655	-5.038	-19.97
	(2.821)	(13.95)	(4.717)	(7.229)	(6.716)
Underpaid/Big Raise	-0.0170	-13.97	-4.419	6.480	10.92
	(2.998)	(9.194)	(3.136)	(4.751)	(5.956)
Overpaid/No Raise	-1.194	-4.223	-3.552	13.89	18.63
	(3.631)	(22.06)	(7.464)	(6.120)	(15.46)
Overpaid/Big Raise	4.583	44.98	11.78	13.56	-2.798
	(3.115)	(27.36)	(7.302)	(8.903)	(15.19)
Observations	362	362	362	362	362

Note: This table displays regression coefficients comparing changes in statistical output for the four different groups from 1989-90 to 1990-91 after using a 15% threshold for whether or not a player received a raise, using salary data from www.markerzone.com and statistical data from www.hockey-reference.com

Another concern is that because Wayne Gretzky and Mario Lemieux are both outliers in the 1990 salary distribution, they may be influential in predicting the parameters used in the lasso model. Because they are both famously talented attacking players, this could place too much weight on the offensive categories. To deal with this, I drop these two before running the lasso model and re-estimate Equation 1, with results displayed in Table D.6.

Table B.6: Robustness Check - 25% Threshold

	(1) Goals	(2) Shots	(3) Assists	(4) +/-	(5) Adj. +/-
Underpaid/No Raise	7.544	30.66	2.967	-7.089	-17.60
	(3.801)	(11.55)	(4.656)	(8.780)	(8.270)
Underpaid/Big Raise	-0.0265	-25.68	-4.962	9.747	14.74
	(2.963)	(9.536)	(3.448)	(5.470)	(5.474)
Overpaid/No Raise	0.968	6.028	-1.866	15.55	16.45
	(4.141)	(25.52)	(7.404)	(4.505)	(14.74)
Overpaid/Big Raise	-0.433	32.52	11.53	6.404	-4.695
	(2.541)	(17.93)	(7.237)	(6.986)	(12.30)
Observations	362	362	362	362	362

Note: This table displays regression coefficients comparing changes in statistical output for the four different groups from 1989-90 to 1990-91 after using a 25% threshold for whether or not a player received a raise, using salary data from www.markerzone.com and statistical data from www.hockey-reference.com

Table B.7: Robustness Check - Dropping Gretzky and Lemiuex

	(1) Goals	(2) Shots	(3) Assists	(4) +/-	(5) Adj. +/-
Underpaid/No Raise	11.34	39.10	8.638	-11.20	-31.17
	(3.236)	(15.58)	(5.431)	(7.764)	(8.393)
Underpaid/Big Raise	0.248	-11.13	-3.238	7.839	10.83
	(3.252)	(10.28)	(2.726)	(5.102)	(4.686)
Overpaid/No Raise	2.344	20.98	4.107	14.47	8.016
	(2.918)	(15.41)	(4.195)	(4.820)	(7.503)
Overpaid/Big Raise	2.928	52.96	9.482	3.707	-8.703
	(3.799)	(30.43)	(9.318)	(6.093)	(16.29)
Observations	360	360	360	360	360

Note: This table displays regression coefficients comparing changes in statistical output for the four different groups from 1989-90 to 1990-91 after dropping Wayne Gretzky and Mario Lemiuex from the predictive regression, using salary data from www.markerzone.com and statistical data from www.hockey-reference.com

If anything, the main results are actually strengthened by this change. The lasso model still finds plus-minus to be an insignificant predictor of salary and shots and the goals-assists interaction are both positive, though now the interaction term is only borderline significant, with a p-value of .112. The shots term, however, is positive and significant at .0001, while the p-value on plus-minus is .835. Looking at Table D.5, none of the key results are changed, and almost all of them are greater in magnitude and significance. The difference in plus-minus between the UN and UR groups is now over 19, with a p-value on the difference between the two estimates of .031. If Gretzky and Lemieux's presence in the lasso model is biasing the estimates from section 5 at all, it is actually biasing them *towards* zero.

B.3.1 Influential Observations - Manual Row Deletion Analysis

Another serious concern is that these results may be sensitive to influential observations. [58] point out that the main findings of many important studies in recent years can lose significance by dropping less than 1% of observations. Because the main findings of this paper depend on the 86 players in the UN group, it could be that a few influential observations are driving the results. In order to address this, I implement the Manual Row Deletion Analysis strategy introduced in [189].

To test a model's sensitivity to j influential observations, MRDA runs iterative simulations which randomly drop different combinations of j obervations and then run the model. With a small enough sample size, like in this model with only 86 observations in the UN group, MRDA can iterate through all possible combinations of j observations when j is sufficiently small. A model that is robust to influential observations should have results of the same sign and significance as the original model in every iteration.

I test whether the main results of the previous section are sensitive to the exclusion of $j = 1, 2, 3$ observations from the UN group. Testing for j influential observations is akin to running $\frac{n!}{k!(n-k)!}$ regressions, where $n = 86$ and $k = n - j$. Therefore, testing $j = 1$ requires 86 iterations, $j = 2$ requires 3,655 iterations, $j = 3$ requires 102,340 iterations. I run the regression with both goals and adjusted plus-minus on the left-hand side to see whether the sign or significance of either changes when certain observations are dropped. When $j = 1$, the coefficient on goals is between 9.6 and 12.3 in each of the 86 iterations, while the smallest t-statistic is 3.28 (p-value = .0011). For adjusted plus-minus, the coefficient is always between -24.6 and -30.7, with the smallest t-statistic (in magnitude) being -2.6 (p-value = .0097). When $j = 2$, the coefficient on goals is always between 7.6 and 12.9, with the smallest t-statistic being 2.34 (p-value = 0.0198), while the coefficient on adjusted plus-minus is always between -22.7 and -34.3, with the smallest t-statistic being -2.47 (p-value = 0.0140). Finally, when $j = 3$, the coefficient on goals is between 6.4 and 13.7 with the smallest t-statistic being 2.12 (p-value = .0347), while the coefficient on adjusted plus-minus is always between -21.1 and -38.1, with the smallest t-statistic being -2.24 (p-value = .0257). This means that even dropping the three most influential observations, which represents 3.5% of the entire UN group, both of the key findings of the paper remain significant at 5%. The distributions of the parameter estimates and t-statistics for both goals and plus-minus when $j = 3$ are displayed in Figure B.1

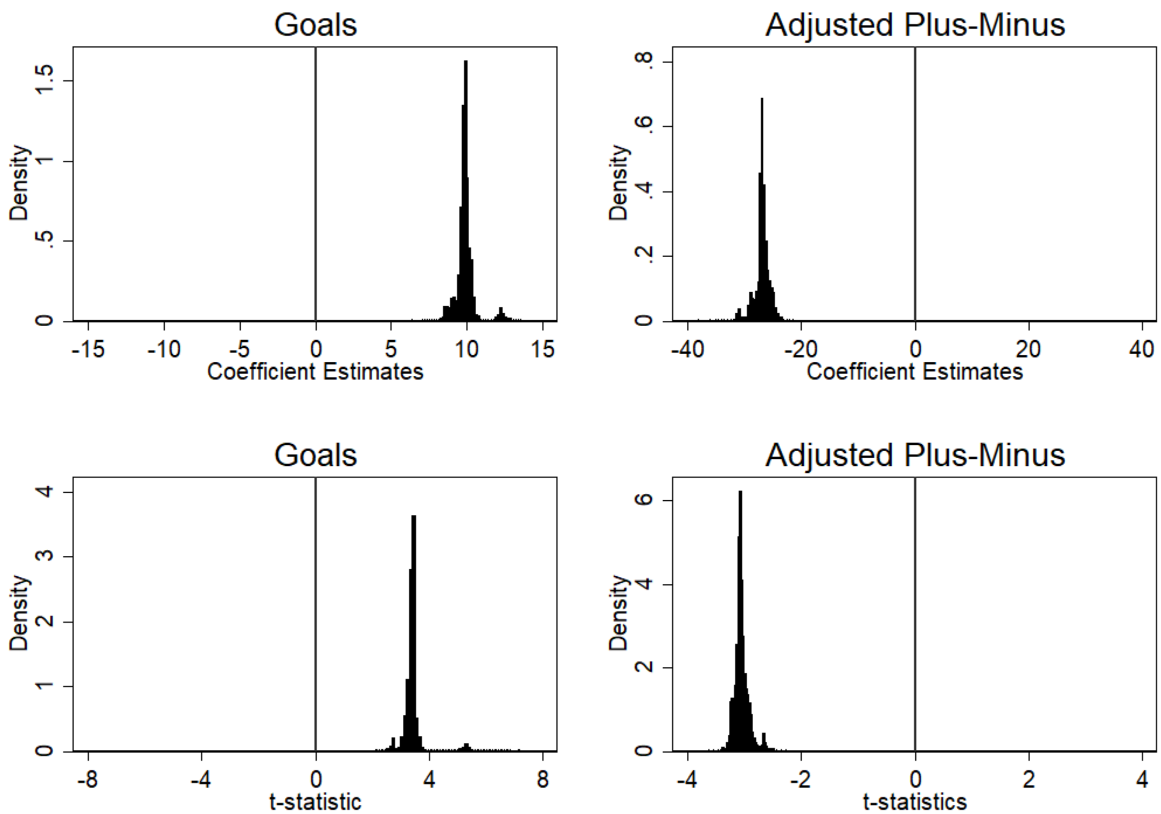
B.4 Individual Outcomes After Disclosure

In order to address whether the behavior shifts outlined in Section 4 did indeed result in increased salaries, Figure B.2 which tracks each group and analyzes whether they returned to play again the following season, whether they received a raise, and what the average raise was, conditional on receiving one. The top left panel of the figure tracks the size of each group over time. For example, the points at 1990 illustrate the number of players who were deemed to have been underpaid in the 1989-90 season and who did not receive a raise in the summer of 1990. For the most part these groups are relatively stable over time, with the exception of what took place surrounding the lockout of 1994-95. Basically, almost no one received a raise going into 1994-95, which is why the underpaid/no raise and overpaid/no raise groups are so large in this season. Then, after the dispute was settled, virtually everyone got a raise in the summer of 1995, so the size of the raise vs. no raise groups flips. Other than this, the relative sizes of the groups is stable over the course of the decade. The top right graph in the figure looks at what percentage of each group returns to play again in the following season after being sorted into each group. For example, the points at 1990 show that just under 80% of the players deemed to be underpaid/no raise or overpaid/no raise going into the 1990-91 season returned to play again in 1991-92. One concern would be that maybe the coaches noticed the behavior change and cut the players who switched to playing more offense, but that doesn't appear to be the case. In general, players are more likely to return if they just received a raise, but that is unsurprising because receiving a raise is evidence that the team values them and wants them to continue. The two groups which do not receive a raise track relatively closely over time. The outlier in 1995 where all of the underpaid/no raise players returned is because there were only two players in that group in 1995 because the vast majority of players received raises post lockout.

The bottom left graph in the figure displays the likelihood of receiving a raise in the following season, conditional on being sorted into each group. This indicates that, among the underpaid/no raise players in the summer of 1990, roughly 60% of these players received a raise after the 1990-91 season. Underpaid/no raise players are the most likely to receive a raise in any given season, which is unsurprising. Finally, the bottom right graph displays the average percentage raise by group, conditional on having received a raise. Here, the underpaid/no raise players in 1990 stand out, as they received an average raise of 186%. This is the largest average raise for any year, excluding 1995 where only two players were in the UN group. There is a large decline in the size of the raises received by UN players in the following season, even though the other groups received roughly similar raises. After this, the trends in the UN group roughly track the rest of the market, with increasing raises in the early 90s which level off and eventually decline. Overall, this figure suggests that being in the underpaid/no raise group in 1990 did not materially affect the likelihood of returning to play again the following season, or the likelihood of receiving a raise, but the underpaid/no raise players do appear to have gotten larger than average raises, conditional on having received one.

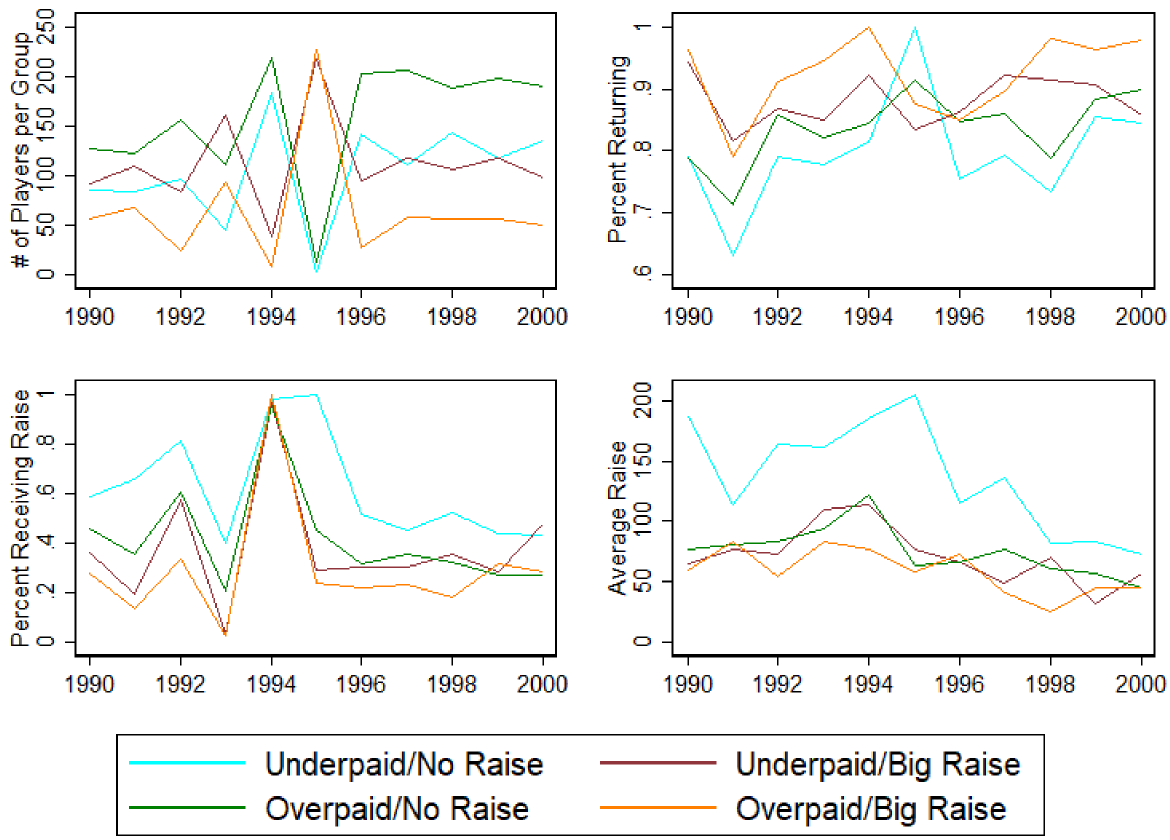
Still, this doesn't fully answer the question of whether it was the behavior shift from defense to offense which directly led to these large raises. To get a better idea of this, I have created a measure of the magnitude of the individual's behavior change, calculated by multiplying their change in goals by the negative of their change in adjusted plus-minus. Players who scored more goals but let many more in will have a large, positive value for this variable. I then regress whether or not a player returned in the following season, whether they got a raise and the size of their raise on this behavior change variable, both for everyone in the league and for the underpaid/no raise players specifically, for each season from 1990-2000. Results are reported in Figure B.3. For the underpaid/no raise players, shifting effort from defense to offense does not appear to have impacted whether they returned to play again the next season, but it is associated with an increased likelihood in getting a raise and with larger raises in the first few seasons after salary disclosure. Results are qualitatively similar across the entire league, though with smaller effect sizes. In the years just after salary disclosure, it does appear that shifting from defense to offense led underpaid players to be more likely to receive any raise, and to increase the size of the raise they received. In both cases, the coefficient on the behavior change variable in the season directly after salary disclosure is the largest in the 11 year window.

Figure B.1: MRDA - Distributions of Coefficient Estimates and Test Statistics



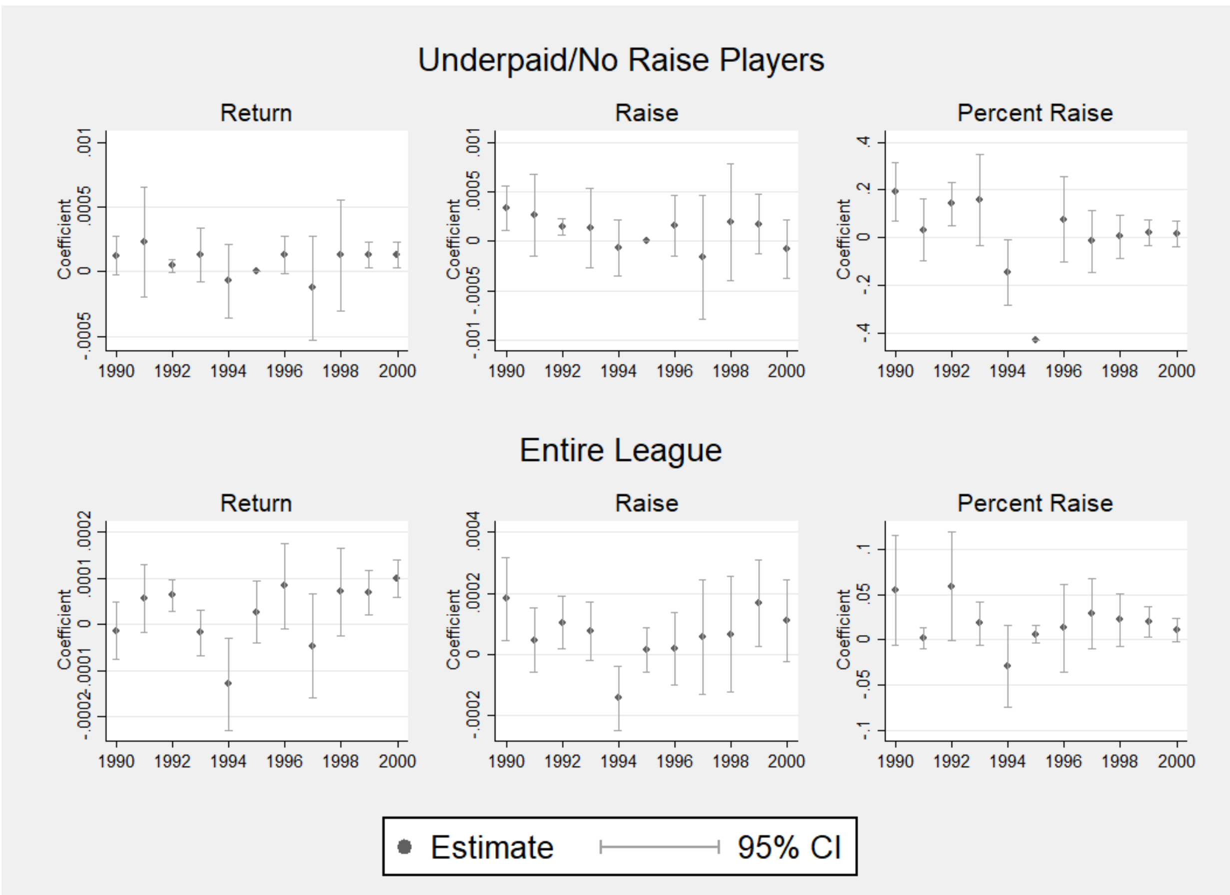
Note: This figure displays distributions of parameter estimates and t-statistics on goals and adjusted plus-minus after dropping every possible combination of three players from the UN group, using salary data from markerzone.com and statistical data from hockey-reference.com

Figure B.2: Tracking Player Outcomes by Group, 1990-2000



Note: This figure tracks several outcomes, broken out by which group players are assigned to in a given year. The top right panel of the figure tracks the size of each group over time. The points at 1990 illustrate the number of players who were deemed to have been underpaid in the 1989-90 season and who did not receive a raise in the summer of 1990. The top right graph in the figure looks at what percentage of each group returned to play again in the following season. The points at 1990 show that just under 80% of the players deemed to be underpaid/no raise or overpaid/no raise going into the 1990-91 season returned to play again in 1991-92. The bottom left graph in the figure displays the likelihood of receiving a raise in the following season, conditional on being sorted into each group. This indicates that, among the underpaid/no raise players in the summer of 1990, roughly 60% of these players received a raise after the 1990-91 season. Underpaid/no raise players are the most likely to receive a raise in any given season, which is perhaps unsurprising. Finally, the bottom right graph displays the average percentage raise by group, conditional on having received a raise. Here, the underpaid/no raise players in 1990 stand out, as they received an average raise of 186%.

Figure B.3: The Effect of Shifting Effort from Defense to Offense on Player Outcomes, 1990-2000



Note: This figure displays coefficient estimates from regressions which measure the association between shifting effort from defense to offense and 1) whether or not the player returned to play again the following season, 2) whether or not they received a raise after shifting their behavior, and 3) the magnitude of the raise they received, conditional on having received one.

Appendix C

Appendix for Chapter 3 - Do Sugar-Sweetened Beverage Taxes Improve Public Health for High School Aged Adolescents

C.1 Choice of Control Groups

The baseline group, C1, includes Brooklyn, Queens, the Bronx, and Los Angeles. Just 98 miles separate New York City from Philadelphia, and New York has tried and failed to implement a soda tax on multiple occasions ([64]), which means residents are likely to be aware of the harmful effects of SSB consumption. Of the five boroughs of New York City, these three are also most similar to Philadelphia in income and demographic makeup. According to the US Census Bureau, Philadelphia is 44.8% white, compared with 43.7% for Brooklyn, 42.7% for Queens, 29.9% for the Bronx. Manhattan and Staten Island both have much larger white populations, at 58.9% and 65.8%, respectively. Additionally, the median income in Philadelphia is $26,200 compared with $22,200 for the Bronx, $31,400 for Brooklyn, and $31,900 for Queens. The other two boroughs, Manhattan and Staten Island, have median incomes of $51,000 and $39,800, respectively.

Similarly, Los Angeles is 52.1% white, compared with San Francisco (52.8%) and Oakland (35.5%), while San Diego has a much larger white population at 65.1%. A statewide soda tax has been under debate in California for several years, though Governor Jerry Brown put a moratorium on new local taxes in 2018[1] , which suggests that residents of California are all likely familiar with these taxes and are aware of the potential harmful effects of SSB consumption. After searching for newspaper articles in each of the Florida cities about proposed soda taxes, I was only able to find

[1] [2]

anything related to the issue in Jacksonville, where a local CBS affiliate discussed the merits of a potential tax[2] , and Orlando, where an op-ed in the Orlando Sentinel written by the CEO of the American Beverage Association claimed that SSB taxes are costly to consumers without making people healthier[3] . For all of these reasons, I include Brooklyn, Queens, the Bronx and Los Angeles in C1, and then add Staten Island, Manhattan and San Diego in C2. Finally, I add the four Florida districts[4] in C3. Summary statistics on a range of outcomes possibly related to SSB consumption are reported in Table 3.1, which confirms that C1 is the most similar group to the treated group. For 11 of the 14 variables, equality of means cannot be rejected between the treated group and C1. Only the % of the sample which is Hispanic, the % which is Asian and the age of the respondent can be rejected at 5% for C1. Importantly, both race/ethnicity and age are controlled for in every specification in this paper. For C2, mean equality for six of the 14 variables can be rejected at 5%, while for C3 eight of the 14 can be rejected at 5%, suggesting that the choice of control groups is reasonable.

C.2 Appendix Tables

Table C.1: Sugar-Sweetened Beverage Taxes in the United States

City	Approval Date	Effective Date	Cents per Oz.	Diet Drinks Taxed?
Berkeley	Nov. 2014	Jan 1, 2015	1	No
Philadelphia	June 2016	Jan 1, 2017	1.5	Yes
Oakland	Nov. 2016	July 1, 2017	1	No
Albany	Nov. 2016	April 1, 2017	1	No
San Francisco	Nov. 2016	Jan. 1, 2018	1	No
Boulder	Nov. 2016	July 1, 2017	2	No
Seattle	June 2017	Jan. 1, 2018	1.75	No

2 [1]
3 [237]
4 Orlando, Jacksonville, Fort Lauderdale and Palm Beach

Table C.2: Census Demographics for Cities Included in YRBSS Sample

City	% White	% Black	% Asian	% Hispanic	% Per Capita Income
Philadelphia	44.8%	43.6%	7.8%	15.2%	\$29,644
Brooklyn	49.8%	33.8%	12.7%	18.9%	\$36,295
Bronx	44.7%	43.6%	4.6%	56.4%	\$22,749
Queens	47.8%	20.7%	26.9%	28.2%	\$33,626
Staten Island	74.5%	11.6%	10.9%	18.6%	\$38,096
Manhattan	64.6%	17.8%	12.8%	25.6%	\$78,771
San Francisco	52.8%	5.6%	36.0%	15.2%	\$72,041
Los Angeles	48.9%	8.8%	11.8%	48.1%	\$37,143
San Diego	62.0%	6.1%	17.3%	30.1%	\$43,090
Oakland	34.4%	22.7%	15.8%	27.0%	\$46,407
Orlando	68.0%	22.8%	5.7%	32.7%	\$31,409
Jacksonville	60.6%	30.8%	5.0%	10.5%	\$32,233
Palm Beach	74.6%	19.8%	2.9%	23.4%	\$40,957
Ft. Lauderdale	63.1%	30.2%	3.9%	31.1%	\$34,063