Essays in Health and Applied Microeconomics

By

Matthew Tomback Knowles

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

May 13, 2022
Nashville, Tennessee

Approved:
Christopher S. Carpenter, Ph.D.
Carolyn J. Heinrich, Ph.D.
Analisa Packham, Ph.D.
Michelle M. Marcus, Ph.D.

To Victoria, my parents, and all those who helped me along the way.

ACKNOWLEDGEMENTS

Doing a PhD is sometimes a lonely task. As such, it would mean a great deal to me when people go above and beyond to support me in my studies. I want to take some space to acknowledge those who made my journey a little (and sometime a lot) easier. Sometimes this was because of an insightful piece of feedback that I was able to apply to my work. However, more often than not, it was encouragement and camaraderie that helped me the most. First, I would like to thank all of the graduate students at Vanderbilt University for their intellectual and moral support over the past six years. In particular, I would like to thank Oscar O'Flaherty, Craig Sylvera, Martin Schmidt, Thu Tran, Elijah Coleman, and Bing Yang Tan for their kindness and friendship throughout the program.

Furthermore, I would also like to thank the professors in and outside of Vanderbilt University for their time, patience, and insightful feedback. First, I would first like to thank my dissertation committee members, Carolyn Heinrich, Michelle Marcus, and Analisa Packham, for guiding me through this rewarding, and often stressful, process. I would also like to thank Andrew Dustan, Lesley Turner, Matthew Zaragoza-Watkins, Pedro Sant'anna, and Sebastian Tello-Trillo for their continuous support. Furthermore, I want to extend thanks to Mike Grossman, Daniel Dench, and everyone out of the NBER New York offices for taking me in and giving me an academic home-away-from-home. I also want to give special appreciation to my advisor, Kitt Carpenter, who never hesitated to provide me with every opportunity for success and professional development that a graduate student could ask for.

I would also like to acknowledge the financial support I received via the Kirk Dornbush Summer Research Grant from the Vanderbilt Department of Economics, the Dissertation

Enhancement Grant from the Russell G. Hamilton Graduate Leadership Institute, and the Dissertation Fellowship from the Center for Retirement Research at Boston College.

Finally, I want to thank my mother, father, and brother for believing in me and showing me nothing but support and love over the last six years. Last but not least, I want to thank my dear fiancée, Victoria, whose contributions to my success and happiness are too numerous to list here.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The three essays of this dissertation are united under the theme of providing credible empirical evidence on novel, public policy-relevant research questions. The first two essays address topics related to the provision and consumption of healthcare in the United States. The third essay studies the tradeoffs between alternate designs of a cash transfer program implemented in Kenya.

The first chapter is titled "Social Security Eligibility and Healthcare Utilization: Evidence from Administrative Data". I estimate the impact of Social Security receipt and retirement on healthcare utilization by exploiting the discontinuous increase in claiming and labor market exist at the Early Eligibility Age of 62. Using administrative data on several types of healthcare encounters from New York and California, I find a discontinuous increase in emergency department visits that do not result in hospitalization by 1-2% at this age. Further analysis demonstrates that this effect is driven by both emergent and nonemergent conditions and is not completely explainable by changes in health insurance status.

The second chapter is titled "How Access to Addictive Drugs Affects the Supply of Substance Abuse Treatment: Evidence from Medicare Part D". This chapter documents that substance abuse treatment (SAT) providers and services respond to increases in population-level opioid addiction. I do this by exploiting the implementation of Medicare Part D as an exogenous increase in the availability of prescription opioids. Starting in 2006, states with higher shares of the population eligible for Medicare Part D experienced increases in residential and hospital inpatient SAT facilities, beds dedicated to SAT, and SAT facilities offering medication-assisted treatment, relative to states with lower shares. These results suggest that the supply of SAT in the United States is capable of responding significantly to changes in demand.

The last chapter is titled “A Fine Predicament: Conditioning, Compliance and Consequence in a Labeled Cash Transfer Program”. The Kenya Cash Transfer Programme for Orphans and Vulnerable Children (CT-OVC) presents a valuable opportunity to examine the effects of imposing monetary penalties for noncompliance with conditions in cash transfer programs, in contrast to providing only guidance (or “labeling”) for cash transfer use. We take advantage of random assignment to a conditional arm within the CT-OVC treatment locations to understand the impact of imposing conditions with penalties on program beneficiaries, as well as how this effect varies by household wealth. Program beneficiaries (orphans and vulnerable children) were expected to visit health facilities for immunizations, growth monitoring and nutrition supplements and to enroll in and attend school. We find little difference in program outcomes between households in the conditional treatment arm compared to those in the treatment arm with labeling only (in which information was provided about these expectations but compliance was not monitored). However, among the poorest CT-OVC beneficiaries, assignment to the conditional arm was associated with penalty fines and a significant decrease in non-food consumption. This suggests that in comparison to labeled cash transfers, conditional cash transfers may produce unintended, regressive policy effects for the most vulnerable participants

CHAPTER 1

SOCIAL SECURITY ELIGIBILITY AND HEALTHCARE UTILIZATION: EVIDENCE FROM ADMINISTRATIVE DATA

1.1. INTRODUCTION

The depletion of the Social Security trust fund is a major policy concern in the United States. Many of the current and proposed solutions to this issue, such as increasing the Full Retirement Age, may result in people postponing Social Security benefits and/or retirement. While these efforts may help preserve the trust fund, it’s possible that disincentivizing retirement could unintentionally affect social welfare through alternate channels. For example, high quality, quasi-experimental work has demonstrated a causal link between retirement and mortality in the United States (Fitzpatrick and Moore, 2018). However, this effect on mortality may only represent one dimension of the total effect of retirement on health in the U.S. In order to more fully understand this relationship, it is also imperative to examine how retirement impacts other important health outcomes such as healthcare utilization. This is an important gap in the literature on Social Security policy, retirement, and health for several reasons. First, mortality is a particularly extreme measure of health status and it is possible that retirement may also affect people's morbidity. This could be reflected by changes in the utilization of acute care services upon retirement, if healthcare access is kept constant. Second, it is also possible that retirement could improve health by reducing the opportunity cost of time for pursuing treatment for non-emergent conditions. Lastly, unlike mortality, the bulk of any healthcare costs incurred by retirement are placed on other people in the same insurance pools or on the taxpayers, thereby creating an externality. Due to these reasons, studying the relationship between retirement and healthcare utilization in the U.S. is important for determining optimal Social Security policy and for the health and retirement literature as a whole.

While there have been a few studies in European countries and China using quasi-experimental designs to study the effect of retirement on healthcare utilization, it is difficult to extrapolate their results to the U.S. context due to cultural and institutional differences (Frimmel and Pruckner, 2020; Nielson, 2019; Zhang et al., 2018; Zhou et al., 2021; Rose, 2020). Indeed, I am aware of only one other study that has attempted to study this reseach question in the U.S. Gorry et al., (2018) use an instrumental variable design based on year-of-age relative to various Social Security policy thresholds, along with data from the Health and Retirement Study, to estimate a decrease in hospitalizations following retirement. My study improves upon the limitations of prior work in several ways. First, I take advantage of the fact that a significant share of the population begin to receive Social Security payments and leave work immediately upon reaching the Early Eligibility Age (EEA) of 62 in a natural experiment to analyze the causal effects of retirement (Fitzpatrick and Moore, 2018). I do this using a regression discontinuity design (RDD) that examines age-related trends in healthcare outcomes for people on each side of the age 62 threshold. The principal advantage of the RDD, compared to alternative research designs, is its high internal validity and transparent identifying assumptions. To the best of my knowledge, mine is the first study to use this approach to study retirement and healthcare outcomes in the United States.

Second, I am able to overcome the small sample sizes inherent to survey data by using administrative data on healthcare encounters from New York and California between 2006-2017. These data, which contain observations on tens of thousands of healthcare encounters per month of age, allow me to examine relatively rare but costly healthcare episodes (i.e., ED visits, inpatient hospitalizations, and ambulatory surgery encounters) with statistical precision. Third, these administrative data are not susceptible to the self-report error inherent to survey data since they are based on hospital billing records. This reduces the possibility of measurement error contributing to misleading estimates. Fourth, ED visits and ambulatory surgery encounters have not previously been examined by prior work on health and retirement in the United States. The examination of ED visits is particularly important since the ED is used as a source of both acute and primary care. Therefore, the frequency of ED visits may change in response to retirement not only due to changes in health status, but also due to increases in free time to consume more discretionary forms of healthcare. Fifth, the administrative data contain detailed information on each healthcare encounter, such as primary diagnosis, that is unavailable in common survey data sets. I am able to use this information to determine whether any effects on ED visits are being driven by emergent versus nonemergent conditions.

I estimate that, across the entire U.S., at least 17% of people claim Social Security retirement benefits and approximately 8% of people retire immediately upon reaching the EEA of 62. I also show that turning 62 decreases the amount of time people spend working by around 2 hours per week on average. Furthermore, reaching the EEA is associated with a discontinuous increase of 1-2% in all-cause ED visits that do not result in hospitalization in New York and California. Estimates for discontinuities in elective inpatient admissions, admissions from the ED, and ambulatory surgery encounters are statistically insignificant. Furthermore, the increase in ED visits occurs across a wide variety of both emergent and nonemergent conditions. I am also able to show that changes in health insurance status and income at retirement do not play a significant role in the effect on ED visits, despite estimating a decrease in employer-sponsored private insurance coverage at age 62 of approximately 4 percentage points. Taken together, these results suggest that the increase in ED visits is potentially driven by both a decrease in health status and a decrease in the opportunity cost of receiving care upon retirement.

My study's methodological improvements also contribute to the broader quasi-experimental literature on retirement and health. The earliest of these studies in the U.S. also rely mostly on survey data, along with a variety of IV approaches, to study this topic (Charles, 2004; Dave et al., 2008; Neumann, 2008; Coe and Lindeboom, 2008; Coe et al., 2012; Insler, 2014). In contrast, Fitzpatrick and Moore (2018) use U.S. administrative data and an RDD to show that mortality increases by 2% at the EEA. Fitzpatrick (2020) deems this integration of U.S. administrative data into the literature on retirement on health as “the next generation of US studies” due to the ability to answer new questions with rigorous methods that require these data, such as RDDs. My study seeks to be the next step in this literature by bringing to bear administrative data on healthcare utilization.

1.2. BACKGROUND: SOCIAL SECURITY BENEFITS AND THE EARLY ELIGIBILITY AGE

1.2.1. Eligibility Requirements

The Social Security program is the primary public retirement benefit for aged individuals in the United States, with approximately 50% of the population aged 65+ relying on Social Security for at least half of their household income as of 2015 (Dushi et al., 2017). Furthermore, about 25% of this population receive a full 90% or more of their income through these monthly payments. Most people first become eligible to receive Social Security retirement benefits after working for at least 10 years in qualified employment and reaching the Early Eligibility Age of 62

(SSA, 2021a). The EEA was established in 1959 for women and 1961 for men as an alternative to claiming at the age of 65 (McSteen, 1985). However, this policy also made it so that claiming at 62 reduced an individual's monthly payments by 5/9 of 1% for each month between the claiming month and when the individual turns 65, which was rebranded as the Full Retirement Age (FRA) (SSA, 2021b). The FRA was eventually increased by two months for each yearly birth cohort after 1937 until the 1943-54 cohorts, where it remained at 66. Starting with the 1955 cohort, the FRA continued its increase by two months by year of birth until it stopped at 67 for cohorts born in 1960 or later (Li, 2021). This implies that someone reaching the EEA and claiming Social Security benefits in 2006 would receive approximately 73.33% of the monthly amount they would have earned if they had waited for their FRA. This percent decreased to 72.77% for someone claiming at the EEA in 2017.

People are allowed to apply for Social Security shortly before they turn 62, which means that they can begin to receive benefits within a couple of days of their birthday (SSA, 2021b). Applications must be made at least four months in advance, which means claiming decisions cannot be made entirely spontaneously. People who claim before their FRA and continue work are subject to the Retirement Earnings Test, whereby one dollar of benefits is withheld for every two dollars earned from work above a certain maximum (SSA, 2021a). Social Security is also available to the dependents and survivors of those who qualified based on work history, as decribed above. While spouses and ex-spouses are not beholden to the same work history requirements as the primary claimant, they must still reach age 62 before they are allowed to claim with few exceptions (SSA, 2021b). A subset of other dependents may claim under the primary claimant as well, including children under 18 (or 19 if still in high school) or children of all ages disabled before 22 (SSA, 2021b).

1.2.2. What Kind of Person Claims and Retires at 62?

Approximately 17% of people first claim Social Security upon turning 62 and, on average, these individuals are more likely to have only a high school education, be of self-reported poor health, to be black, and to have lower personal and household income than those who fully retire at later ages (Rutledge and Wettstein, 2020). Additionally, those claiming at 62 are more likely to have developed work-limiting health conditions by the time they claimed benefits and are more likely to have worked in physically laborious and/or blue-collar jobs (Li et al., 2008; Munnell et al., 2016). It is also the case that individuals who claim at age 62 are more likely to have health insurance policies that are not dependent on employment status (e.g., uninsured, private insurance with retiree coverage, and/or Medicaid) than those who retire at other ages (Rust and Phelan, 1997). However, these are just differences in averages. In fact, those who retire at 62 can be further separated into two groups of roughly equal size: the advantaged and the disadvantaged (Munnell, et al, 2016). In the advantaged group, approximately 87% have at least some college education, only 10% are blue collar workers, and 59% are in the top quartile of wealth. On the other hand, only 17% of the disadvantaged group have attended at least some college, 77% are blue collar workers, and only 8.3% reside in the top quartile of wealth. This suggests that many people who claim at 62 are not forced into by life circumstances, but because of a desire to retire earlier.

Particularly important for my study are the people who claim within only a couple of months of turning 62 and their retirement decisions. In fact, as of 2006, the majority of those who claim within two months of turning 62 have actually previously retired: approximately 60% of male claimants and 69% of female claimants (Waldron, 2020). Furthermore, approximately 28% of male claimants and 21% of female claimants who claim benefits upon turning 62 also retire concurrently and the remaining 12% of men and 10% of women claimants continue to work (Waldron, 2020a). Selection into these subcategories of age 62 claimants is highly correlated with lifetime earnings. Specifically, people in the lower deciles of lifetime earnings are more likely to have retired prior to claiming at 62, while people in the middle of the distribution are most likely to retire when claiming (Waldron, 2020b).

1.3. CONCEPTUAL FRAMEWORK

Initiation of Social Security payments and retirement constitute major life changes that have the potential to impact healthcare utilization in a myriad of ways. Figure 1 is a diagram suggesting several potential channels through which this could occur. First, receiving Social Security benefits, all else equal, constitutes an increase in monthly income of about $1,438 on average as of 2019 (Van de Water and Romig, 2020). Increased lifetime Social Security income has been shown to impact health and healthcare utilization both beneficially (Berman, 2020) and detrimentally (Snyder and Evans, 2006), depending on whether the increased income is accompanied with earlier retirement. Additionally, short-run liquidity shocks induced by the receipt of monthly Social Security checks reduce liquidity and allow people to purchase medical care (Gross et al, forthcoming).

Second, many people decide to retire once they claim Social Security (Fitzpatrick and Moore, 2018). The impact of retirement on healthcare utilization is unclear *a priori* given the large number of potential channels involved. While receiving Social Security increases income, losing wages from exiting the labor force may provide a countervailing effect. Furthermore, some employers do not provide employer-sponsored health insurance to employees after retirement which may reduce rates of health insurance coverage prior to Medicare eligibility (McArdle et al., 2014). However, one of the biggest changes accompanying retirement is an increase in newly freed time that was formerly spent on work. This time could be spent on a wide variety of behaviors that could either positively or negatively affect healthcare utilization in the short run. For example, spending this time on sedentary activities could decrease health (Koster et al., 2012; Matthew et al., 2012), thereby increasing healthcare utilization, while time spent exercising or eating well could do the opposite. Furthermore, retirement is known to be a time during which people increase their consumption of unhealthy substances such as tobacco or alcohol (Ayyagari, 2016; Chuard-Keller, 2021), again decreasing health. Additionally, free time spent outside of work could be directed to more injury-prone activities such as home maintenance or driving. Finally, retirees may simply choose to spend their additional time consuming additional healthcare that is not prompted by a change in underlying health status. In my discussion of mechanisms in Section 1.6, I evaluate many of these possible channels to determine which are most likely to be driving the effects on ED visits.

1.4. DATA AND EMPIRICAL STRATEGY

1.4.1. Administrative Data on Healthcare Utilization

This study's measures of healthcare utilization are based on various sources of administrative data from California and New York. The California data were provided by the Department of Healthcare Access and Information (HCAI)[1] and the New York data were provided by the Healthcare Cost and Utilization Project (HCUP)[2]. The primary outcome I study is

[1] Formerly called the Office of Statewide Health Planning and Development (OSHPD).

[2] The New York data are from HCUP's State Emergency Department Database (SEDD) and State Inpatient Database (SID). The data exclude stays in long-term care units of short-term hospitals, Federal hospitals, and free-standing psychiatric hospitals.

emergency department (ED) visits without hospitalization that occurred between 2006-2017 in either state. ED visits are well as suited as an outcome in this study because people receive care in the ED for a wide variety of reasons. While the prototypical ED visit is for an emergent condition that could not be appropriately treated in an alternative setting, a large share of patients are seen in the ED for nonemergent conditions or for primary care treatable conditions (Johnston et al., 2017). Furthermore, as laid out in Section 1.3, retirement could affect healthcare utilization through several channels such as changes in health status and/or time use. Given the wide variety of conditions and reasons for which people receive care in the ED, and the range of plausible mechanisms through which Social Security eligibility could affect healthcare utilization, it is reasonable that the likelihood of visiting the ED might change discontinuously when reach the EEA. In addition to ED visits, I also examine changes in the rate of inpatient hospitalizations in both states between 2006-2017. Unlike most survey data sets, I am able to separate out inpatient hospitalization as elective or admitted via the ED. Lastly, I examine ambulatory surgery (AS) encounters between 2006-2017 from California only due to to data limitations.

All of these data are based on the billing information associated with each visit and include a substantial amount of detail about the patient encounter including patient demographics, diagnoses, and patient ZIP codes. Importantly, all of these data contain both the date of encounter and the birthdate of each patient at the month-year level. Combining these pieces of information allows me to calculate a person's age at encounter. However, since I do not have people's exact day of birth, I am only able to determine people's age by calendar month. Therefore, in the month that people turn 62 I cannot distinguish those who have passed their birthday from those that have not yet. I discuss my method for accounting for this measurement error in my empirical strategy (Section 1.4.4).

1.4.2. Survey Data for Evaluating Mechanisms

Becoming eligible for Social Security at the EEA has the potential to trigger a wide variety of behavioral responses and changes in life circumstances, many of which could foreseeably affect healthcare utilization. The principal source of survey data that I use is the Health and Retirement Study (HRS), which is a longitudinal survey of a representative sample of the U.S. population aged 50+. Since this survey is only conducted in even years, I use the 2006-2018 waves of the HRS to approximate the 2006-2017 time period of my healthcare utilization data. It is from these data that I draw my primary outcome measures for Social Security uptake, retirement, labor force participation, and health insurance coverage. Importantly, the HRS also contains information on month/year of birth and month/year of interview during the relevant sample period. This information allows me to calculate age (in months) at interview in a similar fashion as I do for the administrative data on healthcare utilization. All estimates using the HRS are made at the national level unless otherwise stated in order to maximize statistical power. However, I also include robustness checks for certain outcomes in which I narrow analyses only to the census divisions that include NY and CA. This is the most granular level of geography that is present in the publicly-available HRS files.

1.4.3. Other Data Sets

This study uses a few additional data sets for supplemental analysis. I use the 2007-2018 Current Population Survey - Annual Social and Economic Supplement (CPS-ASEC) for data on Social Security and state pension receipt for the previous year, as well as for data on the national distribution of household income (Flood et al., 2020). I also make use of natality data collected from historical vital statistics records in combination with population data from the 2010 census 10% sample to construct population counts by age cohort[3] (Ruggles et al., 2020). Lastly, I make use of the National Health Interview Survey (NHIS), a representative annual survey on health and economic outcomes in the U.S., as a supplement to the HRS. I use the publicly available version from IPUMS, which contains information on date of interview and birth at the month level between 2006-2014 (Blewett et al., 2019).

1.4.4. Empirical Strategy

I estimate the causal effect of reaching the EEA of 62 on healthcare utilization through the use of an age-based regression discontinuity design (RDD). I rely upon the "continuity framework" for identifying the unbiased effect of gaining eligibility at the EEA via RDD (Hahn, Todd, and van de Klaauw, 2001; Cattaneo and Titiunik, 2021). The identifying assumption of this approach is that the potential healthcare outcomes are continuous functions of age at the age 62 cutoff in the absence of treatment. Another way of viewing this is that, in the absence of treatment, outcomes would have evolved continuously with respect to age rather than jump discontinuously. This is a good approach for establishing the causal effect of Social Security eligibility as long as people on either side of the age threshold are very similar in almost every way, on average, other than their eligibility status for Social Security. This research design can be viewed as a "fuzzy" RDD, rather than "sharp", for several reasons. First, not everyone who turns 62 is eligible to receive retirement benefits. In fact, the SSA estimated that in 2010, about 4% of people aged 62-84 that year would never go on to claim retirement benefits, principally due to a lack of qualifying work history (Whitman et al., 2011). Furthermore, eligible individuals are neither forced to claim benefits nor

[3] See Appendix A for further discussion and construction of the population denominator.

retire when they turn 62. Therefore, any causal effects on healthcare utilization may attributed to the "compliers", or those who claim benefits and/or retire upon their 62nd birthday.

One potential issue in this approach is if other, non-federal retirement plans also use 62 as a policy-relevant age. While "defined contribution plans" with no age-based eligibility have become the standard for private-sector employees (Dushi et al. 2011), many state and local pensions frequently use age to determine eligibility. Indeed, New York and California both have pension programs for state and local government employees that use age 62 as a relevant threshold for payments[4]. As long as there is not a spike in the claiming of state and local government pensions in these two states at 62 that is comparable to the spike in Social Security claiming, these policies should not significantly influence my results for healthcare utilization. Appendix Figure B.1 compares fractions receiving Social Security and state/local pension income in New York and California by year and by age from 2006-2017 according to the CPS-ASEC. Although Social Security receipt increases dramatically for people aged 62, state and local pension claiming does not increase meaningfully.

In equation (1), I display the primary RDD specification used with the healthcare utilization data. The data are organized into month of age (a), state (s), and encounter year (y) cells. y asy is the outcome of interest, such as the rate of healthcare encounters per 1,000 population[5]. Visits are aggregated in age bins that correspond to the calendar month when people turn a certain age (e.g., the calendar month people turn 62). EEA_a is a dummy that equals 1 for encounters occurring in the month individuals turn 62 or later. $f(admitAge_a, EEA_a)$ is a polynomial function of the

[4] New York's NYSLRS uses 62 as its internal "full retirement age" and California's CALPERS uses 62 as its maximum "normal retirement age".
[5] Specifying the outcome in terms of rates instead of counts controls for sudden changes in population size across age bins. This is particularly important when disaggregating analysis by year.

running variable that is allowed to vary on either side of the cutoff. d_a is a dummy that equals 1 in the month that people turn 62 and 0 otherwise to account for attenuation bias (Dong, 2015). Lastly, $\delta_s \times \sigma_y$ are state-by-year fixed effects to increase the model's explanatory power. Main specifications use triangular kernels and standard errors are heteroskedasticity-robust. In specifications estimated using the Calonico, Cattaneo, and Titiunik (CCT) Mean-Squared Error (MSE) optimal bandwidth, I also display p-values derived from their robust bias-correct confidence intervals (Calonico, Cattaneo, and Titiunik, 2014).

$$y_{asy} = \alpha + \beta EEA_a + f(admitAge_a, EEA_a) + d_a + \delta_s \times \sigma_y + \epsilon_{asy} \tag{1}$$

In models using outcomes derived from survey data, I aggregate data only by month of age in order preserve the role of sample weights and estimate equation 2.

$$y_a = \alpha + \beta EEA_a + f(admitAge_a, EEA_a) + d_a + \epsilon_a \tag{2}$$

A notable threat to identification of causal effects on healthcare outcomes using this research design is if people systematically migrate away from NY and/or CA when the turn 62. I can test this question directly by plotting in Appendix Figure B.2 the estimated population of people living in these states by quarter of age using the 10% sample of the 2010 Decennial Census. I also fit linear polynomials on either side of the cutoff in order to estimate any change in population. The RD estimate, as displayed in Figure B.2, is statistically insignificant which suggests that aggregate population does not change at the cutoff.

1.5. MAIN RESULTS

1.5.1. Effects on Social Security Uptake and Labor Market Outcomes

As discussed in the previous section, there are two primary channels through which eligibility for Social Security at the EEA are likely to affect healthcare utilization outcomes: retirement and benefit receipt. Many of the other potential mechanisms that could affect healthcare utilization, such as changes in health insurance or changes in income, would likely flow from at least one of these channels. Below, I use the HRS to estimate the share of people who claim Social Security and retirement upon turning 62 during my sample period. I also use information from the HRS to quantify the substitution in time use away from working. I show that people do indeed claim Social Security, retire, and exit the labor force discontinuously at 62 between 2006-2018. Figure 2 shows the RD plot for the fraction of people, by age, who self-report receiving some type of Social Security income using the HRS full national sample. The figure displays a visible positive discontinuity in Social Security receipt at 62 as well as a point estimate of about 16.6 percentage points that is statistically significant at $p < .01$[6]. Figure 3 shows RD plots and point estimates of discontinuities in various measures of retirement and labor force participation at 62, also for the entire U.S. These measures include the average number of hours people report working per week in order to quantify the substitution in time use that occurs upon retirement. All four outcomes display discontinuous changes at 62 of large magnitudes. The estimates in Panels (a) - (c) display an 8.2 percentage point increase in share retired, a 6.3 percentage point increase in labor force non-participation, and a 7.6 percentage point decrease in the share working for pay[7]. Panel D displays

[6] The outcome is not limited to individuals receiving Social Security Retirement Insurance, and includes individuals receiving Disability and/or Survivors Insurance. This explains the non-zero levels of Social Security uptake prior to 62.

[7] Figure 3 appears to suggest that these effects on labor market outcomes might not show up immediately and could possibly take up to two months to do so. In order to verify if this is truly the case, or just due to sampling variability, I compare the labor force results in the HRS to those in the NHIS between 2006-2014 in Appendix Figure B.3. While both RD plots demonstrate discontinuities at 62, HRS results display a delayed effect, which suggests that it is likely due to sampling variation.

that people report working almost two hours less per week, on average, when they turn 62[8]. This discontinuity can be viewed as an "intent-to-treat" effect. since only about 8 percent of people retire discontinuously at this age. If one assumes that the entire effect on work-related time use is driven by the effect on retirement, then a back-of-the-envelope calculation suggests that people who retire at 62 work approximately 23.9 hours less per week on average[9]. The effects on Social Security uptake and labor market outcomes have been fairly stable over time, as demonstrated in Appendix Figure B.4. Despite some apparent cyclicality, the effects on Social Security claiming and retirement at 62 have hovered around 20% and 10%, respectively, over the two decades prior to 2018.

Previous research on Social Security eligibility at the EEA has shown that the discontinuities in claiming and labor force outcomes vary considerably by demographic subgroup. In Table 1, I show results from estimating equation (2) with these outcomes across sexes and racial/ethnic categories. Column (1) demonstrates that each demographic subgroup experiences an increase in Social Security receipt at 62 that is significant at at least the 10% level. On the other hand, columns (2) - (4) show that the effects on labor market outcomes are not as widespread across demographic subgroups. First, unlike in Fitzpatrick and Moore (2018), I find significant increases in retirement, labor force non-participation, and decreases in working for pay among females as well as males. This is likely because labor force attachment is higher in my sample period than their earlier one[10]. Second, I find that changes in labor market outcomes are concentrated among non-Hispanic white people and, less robustly, Hispanic people.

---

[8] Respondents who do not report working any job or spending time on any businesses are coded as working zero hours per week.

[9] $\frac{Reduced\ Form}{First\ Stage} = \frac{-1.959}{0.08} = -24.49$

[10] Appendix Figure B.5 demonstrates that this finding is not particular to the HRS by estimating the discontinuity in labor force non-participation by sex using the NHIS. In this data set, the coefficient estimate is actually larger (though not significantly so) among females than males.

As I discuss in Section 1.4.2, the limited sample size of the HRS necessitates the use of the national sample to maximize statistical precision when estimating treatment effects. However, since the healthcare data I use in this study are from NY and CA, it is useful to provide additional evidence that the social security and labor market effects observed in the national sample are also present for these states specifically. Using the publicly-available HRS files, I can limit the sample to the Pacific and Middle Atlantic Census Divisions. According the US Census Bureau, people living in NY and CA comprise approximately 60% of the population aged 60-64 in these divisions (U.S. Census Bureau, Population Division, 2020)[11]. Appendix Figure B.6 shows the results for Social Security uptake and labor market outcomes in these two Census Divisions alone. The effects are even larger in magnitude and significance than those for the entire sample for each outcome, which suggests that these effects are likely to hold in NY and CA independently.

1.5.2. Effects on Healthcare Utilization

ED Visits

Figure 4 plots rates of total ED visits per 1,000 population by age in months with linear and quadratic fits by estimating equation (1) with 8-month bandwidths. Both panels (a) and (b) show clear jumps in visits between the month before people turn 62 and the month after people turn 62 that are robust to choice of polynomial. I evaluate these visual assessments by estimating the RD coefficients for each discontinuity under varying polynomial and bandwidth assumptions using equation (1) and display the results in Table 2. The top panel, which displays results for the aggregate sample of both states combined, is consistent with the respective plot in Figure 5.

[11] The Pacific Census Division consists of Alaska, California, Hawaii, Oregon, and Washington. The Middle Atlantic Census Division consists of New Jersey, New York, and Pennsylvania.

Column (1) indicates that becoming eligible for Social Security at the EEA of 62 is associated with an increase of 0.204 visits per 1,000 people, which is equivalent to an increase of about 1.1% compared to the month before turning 62. This effect is statistically significant at less than the 1% level and is robust to the use of a quadratic polynomial, as shown in column (2). Columns (3) and (4) re-estimate the models using the mean-squared error (MSE) optimal bandwidth from Calonico, Cattaneo and Titiunik (2014) with 24 months of data on either side of the cutoff. These point estimates are statistically significant at the 1% level, regardless of polynomial choice and inference method. Lastly, columns (5) and (6) estimate alternative models using uniform kernels with CCT-selected bandwidths. These point estimates are of similar magnitudes and statistical significance to those in columns (3) and (4), although the bandwidth selection is notably smaller for the linear model in column (5). The bottom two panels separately re-estimate these specifications for each state. Although the estimates for New York are generally larger and more statistically significant than those for California, the results indicate that the effect on ED visits is present in both states. These findings suggest that the discontinuity at 62 may be generalizable to other states as well since they do not appear to be driven by a single state's policies.

Column (1) of Table 3 displays how the effect on ED visits varies across demographic subgroups. Outcomes are specified as the natural log of visits by sex and race/ethnicity since the population denominator cannot be cut by all of these categories. Similar to the effects on Social Security claiming and labor market outcomes, there are significant discontinuities for both males and females at age 62. On the other hand, the positive effect on ED visits appears to be mostly concentrated among non-Hispanic white individuals, which is consistent with the effects on labor market outcomes as well. Lastly, I examine how the effect on ED visits has changed over the time period of study in Figure 5. The coefficient estimates are mostly positive throughout the entire time period with no visible increases or decreases over time. This is also consistent with the results for Social Security receipt and labor market outcomes which are mostly positive across the relevant time period.

Inpatient Hospitalizations

I turn my attention next to the effect of reaching the EEA on inpatient hospitalizations. I focus on two types of admissions: those originating in the ED and those for elective procedures[12]. Figure 6 shows how these two types of admissions change through the age 62 cutoff. Neither panel (a) nor panel (b) display large discontinuous changes in admissions at 62. Table 4 confirms these null effects across polynomials and bandwidths. However, it is possible that these aggregate results belie significant heterogeneity across demographic groups. Columns (2) and (3) of Table 3 display RD estimates for each type of inpatient admission by subgroup. While coefficients are statistically insignificant in most categories, I estimate decreases in admissions from the ED for black people and people of non-white/Hispanic/black races and ethnicities that are significant at the 1% and 10% levels, respectively.

Ambulatory Surgery Encounters

The last form of health care utilization I analyze is ambulatory surgery encounters. Figure 7 shows the RD plot for AS encounters in California and does not suggest an exceptionally large discontinuity at 62. Regression estimates in Table 5 confirm this assessment, displaying

[12] As of 2017, hospitalizations from the ED account for approximately 65% of admissions in New York and California for people aged 62 and 11 months. Elective hospitalizations account for approximately 25% of admissions in the same sample.

insignificant coefficients across all specifications. Furthermore, I do not estimate significant effects within any of the demographic subgroups as displayed in column (4) of Table 3.

1.5.3. Robustness Checks

I now conduct two evaluations of the robustness of the estimated effect on ED visits. First, I systematically test the robustness of the aggregate ED estimates with triangular kernel to various alternate bandwidths by choice of polynomial and plot the results in Appendix Figure B.7. The coefficient estimates are significant at the 5% level for bandwidths of 4 through 12 months when using a linear fit. Since smaller bandwidths tend to suffer less from a variety of undesirable characteristics (e.g., erratic behavior near boundary points, counterintuitive weighting, and overfitting), these findings are reassuring (Calonico and Titiunik, 2021). When using a quadratic fit, the coefficients estimates are significant at the 5% level for every bandwidth except for 4 months where the confidence interval is relatively large due to the lack of observations.

Next, I present the results from re-estimating the effect on ED visits at a variety of placebo cutoffs. I do this in two ways. First, I re-estimate equation (1) with the natural log of aggregate ED visits as the outcome at nearly every month within a 10-year radius of the age 62 cutoff[13]. In total this results in 204 placebo cutoffs. Then, I plot the distribution of these RD coefficient estimates in panel (a) of Appendix Figure B.8. The dashed lines indicate the boundaries containing 95% of estimated coefficients and the red line indicates the treatment effect estimate. The second way I conduct this placebo test is by using the t-statistics generated via the CCT robust inference methods

[13] I use the natural log of ED visits instead of the rate per 1,000 population because the vital statistics data do not go back far enough to estimates effects for older individuals. Additionally, for the months to truly be placebo cutoffs, they must not overlap with the effect of the EEA or any other relevant age-based policy. That is to say, I want to avoid "treatment effect contamination" (Calonico and Titiunik, 2021). For this reason, I exclude months within an 8-month radius of age 62 and age 65, as well as the month people turn 65 (and become eligible for Medicare).

instead of just the coefficient estimates. I plot the distribution of t-statistics in panel (b) of Appendix Figure B.8. In both cases, the estimate at age 62 is larger than 95% of estimates. This suggests that the estimated effect of interest is exceptionally large when compared to placebo estimates and is therefore likely to be a true policy effect.

1.6. UNDERSTANDING THE HEALTHCARE EFFECTS AT 62

In Section 1.3, I articulate the various channels through which Social Security eligibility might impact healthcare utilization. As shown in Figure 1, uptake of Social Security payments and labor force departure are the two primary behaviors that are likely to change at age 62. However, each of these behaviors has the potential to cause a variety of cascading effects that may ultimately affect utilization patterns. Understanding the relative contributions of these mechanisms in important for policy making and is worth further analysis. In this section, I first provide some evidence about types of ED visits that change discontinuously at 62. Then, I show that subsequent changes in health insurance coverage and household income at 62 are not likely to be the main mechanisms at play.

1.6.1. Emergent versus Nonemergent Care

Stratifying the effect on ED visits by primary diagnosis can aid in evaluating which mechanisms might underpin the aggregate effect[14]. In Figure 8, I display the point estimates and 95% confidence intervals from re-estimating equation (1) by International Classification of Diseases (ICD) category. There are generally positive or near-zero coefficient estimates across

[14] Primary diagnosis is defined as the principal reason why the patient showed up to the ED that day, as determined by the assigned physician.

most diagnosis categories, with no category demonstrating an increase that is disproportionate with its share of total visits. Figure 8 also displays a significant negative effect on visits for endocrine and metabolic disorders. Further analysis indicates that this negative effect is driven by decreases in visits for diabetes (β =-0.023, $p<0.01$), disorders of non-thyroid glands (β =-0.025, $p<0.001$), and various metabolic disorders (β =-0.011, $p<0.05$) and is driven by black and Hispanic individuals. On the other hand, the effects on the remaining categories of endocrine/metabolic disorder (thyroid problems, nutritional deficiencies) are positive and significant at the 5% level in aggregate.

Given the diffuse nature of the change in ED visits across types of conditions, it may aid interpretation to sort visits according to their level of urgency. In particular, visits for emergent conditions may be more likely to be driven by sudden changes in health status than visits for nonemergent conditions. One method for classifying visits is to use an algorithm developed by researchers at New York University (NYU) (Billings et al., 2000). These researchers used extensive internal records on ED visits from six hospitals in the Bronx, NY and categorized each visit into one of the following categories: (1) nonemergent; (2) emergent, primary care treatable; (3) emergent, ED care needed but preventable; or (4) emergent, ED care needed, not preventable. The authors then separated out visits due to injury, mental health, alcohol use, or substance use and placed them in their own designated, or "carved out", categories. Using these classifications, the researchers assigned to each primary diagnosis code (except for those carved out) a vector of four probability weights, one for each category, corresponding to the share of visits that fell into the given category[15]. All diagnosis codes that were not present among the researcher-categorized

[15] For example, if 50% of visits for ICD-9 code X were emergent, but primary care treatable and 50% of visits were nonemergent, then those two categories would receive weights of .5 in each of those two categories and the other categories would receive weights of 0.

visits are designated as “uncategorized”. In my analysis, I use an updated version of this NYU algorithm that supplemented the algorithm with ICD-9 and ICD-10 codes added since the original study (Johnston et al., 2017).

I estimate the effect of turning 62 on ED visits within each of these aforementioned categories and display the point estimates and 95% confidence intervals in Figure 10. Each outcome is constructed by summing the probability weight for each category within each age-state-year bin and the treating the sum as the “expected number of visits” for that category. These results indicate that the effect on ED visits at 62 is being driven by emergent conditions, including unavoidable ones, as well as non-emergent conditions. Estimates for each of the carved-out categories, as well as for the remaining unclassifiable visits, are insignificantly different from zero. Further analysis also reveals the specific diagnosis categories driving the effects for each of these categories. The increase in emergent, unavoidable visits is mostly driven by diagnoses of signs, symptoms, and ill-defined conditions ($\beta = 0.042$, $p<0.01$), though also somewhat by genitourinary issues ($\beta = 0.014$, $p<0.05$) and digestive issues ($\beta = 0.010$, $p<0.10$). Although there is no aggregate increase in avoidable emergent visits, this belies a substantial decrease in endocrine-related visits ($\beta = -0.035$, $p<0.001$) that is counterbalanced by a statistically insignificant by large increase in circulatory and respiratory visits. The aggregate increase in primary care treatable emergent visits is driven by a wide variety of diagnoses, largest of which is an increase in respiratory issues ($\beta = 0.039$, $p<0.001$) mostly related to COPD. COPD-related conditions are important drivers of mortality at 62 as well (Fitzpatrick and Moore, 2018). There is also a statistically significant decrease in endocrine-related visits in this category ($\beta = -0.008$, $p<0.01$). Lastly, the increase in nonemergent visits is nearly entirely driven by musculoskeletal issues ($\beta = 0.046$, $p<0.01$).

However, it is possible that these estimated effects on both emergent and nonemergent visits are just artefacts from the aggregation scheme of the probability weights. In Appendix Figure B.9, I display two alternate approaches if assigning diagnoses to categories. The first approach involves assigning each non-carved out visit as either “emergent” or “nonemergent” depending on whether the sum of emergent categories' weights summed to more than .5 or whether the nonemergent weight is larger the .5. If both the sum of the emergent weights equal .5, then the visit is categories as ambiguous. The second method is similar to this, but classifies visits as “high” or “low” probability emergent versus non-emergent depending on if the weights summed between .75 and 1 or .5 and .75, respectively. Reassuringly, both methods produce similar results as the main method.

Taken together, the findings displayed in Figures 8 and 9 suggest that the change in ED visits is driven both by changes in health status and, potentially, the opportunity cost of time. Assuming the cost of care stays relatively constant after turning 62, an increase in unavoidable emergent visits where ED care is needed is a strong indicator that health status is discontinuously changing. On the other hand, an increase in nonemergent visits upon turning 62 suggests that something about receiving Social Security payments and/or retirement increases the propensity to seek out care for existing issues. Given the discontinuous increase in time spent not working at this age, a possible explanation for the effect on nonemergent visits is an increase in the opportunity cost of time spent receiving healthcare. A piece of evidence in favor of this explanation is the discontinuous drop in ED visits for diabetes at 62. Diabetes is a challenging condition to manage properly while at work and workers with diabetes often sustain higher-than-optimal glucose counts as a result (Ruston et al., 2013). It is entirely possible that an increase in time to administer self-care caused a decrease in ED visits for diabetes-related episodes, particularly since a large share of the drop is among emergent by preventable conditions. Thus, if retiring at 62 can increase people's time to take care of their diabetes at home, it is possible that they also have more time to pursue other forms of treatment outside of the house.

1.6.2. Health Insurance

The interpretations for the results discussed in Section 1.6.1 hinge upon there being no concurrent change in the financial cost of receiving care. Many people lose access to employer-sponsored health insurance when they retire. Furthermore, accessing Social Security income provides additional liquidity that may be used to purchase private health insurance, all else being equal. These changes in health insurance status could, in turn, potentially affect healthcare utilization patterns. I assess whether health insurance status changes discontinuously at age 62 using the HRS in Figure 10. Panel (a) shows that the overall share of respondents reporting having at least one form of health insurance does not change discontinuously at 62, despite the 4.2 percentage point decrease in employer-sponsored private insurance displayed in panel (b)[16]. This decrease is compensated for by a concurrent increase in the share of people reporting other categories of insurance. Appendix Figure B.10 shows how these other categories change independently at 62. Additionally, Appendix Figure B.11 displays statistically insignificant effects on the likelihood of having two or more plans and the total number of plans held.

The statistically insignificant effect on the likelihood of having health insurance suggests that insurance coverage is not driving the effect on ED visits. However, the relatively small sample size in the HRS may underpower these models to detect small but meaningful effects on health insurance status. In order to provide further evidence that loss of employer-sponsored private

[16] This excludes specialty plans such as dental, vision, etc.

insurance cannot explain the effect, Appendix Figure B.12 displays RD plots for both the share and rate of privately insured ED visitors per 1,000 population. If the effect on ED visits is driven entirely by formerly insured individuals seeking care at the ED, one would expect to see a net decrease in privately insured visits at 62. Instead, my findings suggest the opposite. While the relative share of patients with private insurance stays constant, I estimate a significant increase in the total number of patients with private insurance at age 62. This implies that the ED results cannot be entirely driven by reductions in private insurance coverage since I estimate a significant increase in ED visits for this population at age 62.

1.6.3. Income

It is also possible that substantial chages in household income upon individuals reaching 62 could have affected their healthcare use. I evaluate this by plotting household income by year of age from the CPS-ASEC in Figure 11[17]. Given the coarse granularity of this variable it is difficult to draw firm conclusions, but the income level at age 62 does not appear to lie significantly outside of the preceding trend.

1.7. CONCLUSION

This study presents evidence that claiming Social Security benefits and retiring at the Early Eligibility Age of 62 increases healthcare utilization for both emergent and nonemergent conditions. I am able to provide more comprehensive and credible evidence on this question than previous studies due to my use of administrative data on healthcare encounters and an internally

[17] I use the CPS-ASEC instead of the HRS since both surveys ask about income in the past 12 months, instead of current income, but the CPS-ASEC has a larger sample size.

valid regression discontinuity design. I also show that these healthcare effects cannot be entirely accounted for by concurrent losses of employer-sponsored health insurance. This paper's findings that both nonemergent and emergent healthcare utilization increase upon retirement suggests a complementary pair of policy recommendations. First, since early claiming of benefits and retirement appear to decrease health status and burden the Social Security trust fund, it may be beneficial to both population health and public finances to reduce policy incentives to do so. On the other hand, retirement also appears to increase episodes of nonemergent healthcare utilization that could protect the elderly's future health status. The available evidence suggests that this is most likely due to a reduction in the opportunity cost of time that occurs when people are no longer working. Therefore, encouraging public and firm policies to protect workers' ability to seek out primary care during the workday may serve to both increase long-run health and disincentivize early retirements.

1.8. REFERENCES

Arenberg, Sam, Seth Neller, and Sam Stripling (2020). The Impact of Youth Medicaid Eligibility on Adult Incarceration. Working Paper.

Ayyagari, Padmaja (2016). The impact of retirement on smoking behavior. *Eastern Economic Journal* 43: 270-287.

Berman, Jacob (2020). Can Income Buy Health? Evidence from Social Security Benefit Discontinuities and Medicare Claims. Working Paper.

Billings, John., Nina Parikh, and Tod Mijanovich. 2000. Emergency Department Use in New York City: A Substitute for Primary Care? *Issue Brief Commonwealth Fund.*

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica 82 (6)*, 2295-2326.

Cattaneo, Matias D., Rocio Titiunik, (2021). Regression Discontinuity Designs. Working Paper.

Charles, Kerwin K. (2004). Is Retirement Depressing?: Labor Force Inactivity and Psychological Well Being in Later Life. *Research in Labor Economics* 23: 269-299.

Chuard-Keller, Patrick (2021). With Booze, you Lose: The Mortality Effects of Early Retirement. Working Paper.

Coe, Norma B. & Maarten Lindeboom (2008). Does Retirement Kill You? The Evidence from Early Retirement Windows. CentER Discussion Paper Series No. 2008-93. Tilburg, ND: Tilburg University. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1295315

Coe, Norma B., Hans-Martin von Gaudecker, Maarten Lindeboom, and Jürgen Maurer (2012). The Effect of Retirement on Cognitive Functioning. *Health Economics*, 21 (8): 913-927.

Dave, D., Rashad, I., & Spasojevic, J. (2008). The Effect of Retirement on Physical and Mental Health Outcomes. *Southern Economic Journal*, 75 (2): 497-523.

Dong, Yingying (2014). Regression Discontinuity Applications with Rounding Errors in the Running Variable. *Journal of Applied Econometrics*, 30 (3): 422-446.

Dushi, Irena, Howard M. Iams, and Jules Lichtenstein (2011). Assessment of retirement plan coverage by firm size using W-2 tax records. Social Security Bulletin, 71 (2) (2011), pp. 53-65.

Dushi, Irena, Howard M. Iams, and Brad Trenkamp (2017). The Importance of Social Security Benefits to the Income of the Aged Population. Social Security Bulletin, 77 (2) (2017).

Fitzpatrick, Maria D. & Timothy J. Moore (2018). The mortality effects of retirement: Evidence from Social Security eligibility at age 62. *Journal of Public Economics*, 157: 121-137.

Fitpatrick, Maria D. (2020). Does Working Longer Enhance Old Age? Wharton Pension Research Council Working Paper No 2020-20.

Frimmel, Wolfgang and Gerald J. Pruckner (2020). Retirement and healthcare utilization. *Journal of Public Economics*, 184: 104146.

Gorry, Aspen, Devon Gorry, and Sita Nataraj Slavov (2018). Does retirement improve health and life satisfaction? *Health Economics*, 27 (13): 2067-2086.

Gross, Tal, Timothy J. Layton, and Daniel Prinz (forthcoming). The Liquidity Sensitivity of Healthcare Consumption. *American Economic Review: Insights*.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69 (1): 201-209.

Insler, Michael (2014). The Health Consequences of Retirement. *Journal of Human Resources*, 49 (1): 195-233.

Johnston, Kenton J., Lindsay Allen, Taylor A. Melanson, and Stephen R. Pitts (2017). A "Patch" to the NYU Emergency Department Visit Algorithm. *Health Services Research*, 52 (4): 1264-1276.

Koster, Annemarie., Paolo Caserotti, *et al.* (2012). Association of sedentary time with mortality independent of moderate to vigorous physical activity. *PLoS One*, 7 (6): Article e37696.

Lucifora, Claudio and Daria Vigani (2018). Health care utilization at retirement: The role of the opportunity cost of time. *Health Economics*, 27 (12): 2030-2050.

Matthews, Charles E., Stephanie M. George, *et al.* (2012). Amount of time spent in sedentary behaviors and cause-specific mortality in US adults. *American Journal of Clinical Nutrition*, 95: 437-445.

McArdle, Frank, Tricia Neuman, and Jennifer Huang (2014). Retiree health benefits at the crossroads. Menlo Park, CA: Kaiser Family Foundation.

McSteen, Martha A. (1985). Fifty Years of Social Security. Social Security Bulletin, 48 (8): 36-44.

Munnell, Alicia H., Geoffrey T. Sanzenbacher, Anthony Webb, and Christopher M. Gillis (2016). Are Early Claimers Making a Mistake? CRR WP 2016-5.

Li, Xiaoyan, Michael Hurd, and David S. Loughran (2008). The Characteristics of Social Security Benefits Who Claim Benefits at the Early Eligibility Age. AARP Public Policy Institute Research Report No. 2008-19. AARP, Washington DC.

Lynn A. Blewett, Julia A. Rivera Drew, Miriam L. King and Kari C.W. Williams. IPUMS Health Surveys: National Health Interview Survey, Version 6.4 [dataset]. Minneapolis, MN: IPUMS, 2019.

Neuman, Kevin (2008). Quit Your Job and Get Healthier? The Effect of Retirement on Health. *Journal of Labor Research*, 29 (2): 177-201.

Nielsen, Nick F. (2019). Sick of retirement? *Journal of Health Economics*, 65: 133-152.

Rose, Liam (2020). Retirement and health: Evidence from England. *Journal of Health Economics*, 73: 102352.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2020.

Rutledge, Matthew S. and Gal Wettstein (2020). Is Nontraditional Work At Older Ages Associated with Better Retirement Security? CRR WP 2020-13.

Rust, John and Christopher Phelan (1997). How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets. *Econometrica, 65 (4)*, pp. 781-831.

Ruston, Annmarie, Alison Smith, and Bernard Fernando (2013). Diabetes in the Workplace - Diabetic's Perceptions and Experiences of Managing Their Disease at Work: A Qualitative Study. *BMC Public* Health, 13: 386.

Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 7.0 [dataset]. Minneapolis, MN: IPUMS, 2020.

Snyder, S. E. and William N. Evans (2006). The effect of income on mortality: Evidence from the Social Security Notch. *Review of Economics and Statistics*, 88 (3): 482-495.

Social Security Administration. (2021a). Retirement Benefits. Publication No. 05-10035.

Social Security Administration (SSA). (2021b). Annual Statistical Supplement to the Social Security Bulletin, 2021. Retrieved from: https://www.ssa.gov/policy/docs/statcomps/supplement/2021/supplement21.pdf

United States Census Bureau, Population Division. (2020). Annual Estimates of the Resident Population for Selected Age Groups by Sex for Alaska, California, Hawaii, New Jersey, New York, Oregon, Pennsylvania, Washington: April 1, 2010 to July 1, 2019.

Uscher-Pines, Lori, Jesse Pines, Arthur Kellermann, Emily Gillen, and Ateev Mehrotra (2013). Deciding to Visit the Emergency Department for Non-Urgent Conditions: A Systematic Review of the Literature. *American Journal of Managed Care*, 19 (1): 47-59.

Van de Water, Paul N. and Kathleen Romig (2020). Social Security Benefits are Modest. Center of Budget and Policy Priorities and Policy: Policy Futures.

Waldron, Hilary (2020a). Trends in Working and Claiming Behavior at Social Security's Early Eligibility Age by Sex. ORES Working Paper Series No. 114. Washington, DC: SSA, Office of Retirement and Disability Policy, Office of Research, Evaluation, and Statistics.

Waldron, Hilary (2020b). Working and Claiming Behavior at Social Security's Early Eligibility Age Among Men by Lifetime Earnings Decile. ORES Working Paper Series No. 115. Washington, DC: SSA, Office of Retirement and Disability Policy, Office of Research, Evaluation, and Statistics.

Whitman, Kevin, Gayle L. Reznik, and Dave Shoffner (2011). Who Never Receives Social Security Benefits? Social Security Bulletin, 71 (2): 17-24.

Zhang, Yi, Martin Salm, and Arthur van Soest (2018). The effect of retirement on healthcare utilization: Evidence from China. *Journal of Health Economics*, 62: 165-177.

Zhou, Q., Eggleston, K., & Liu, G. G. (2021). Healthcare utilization at retirement in China. *Health Economics*, 30 (11): 2618-2636.

Figure 1: Conceptual Channels Connecting Social Security Eligibility to Healthcare Utilization

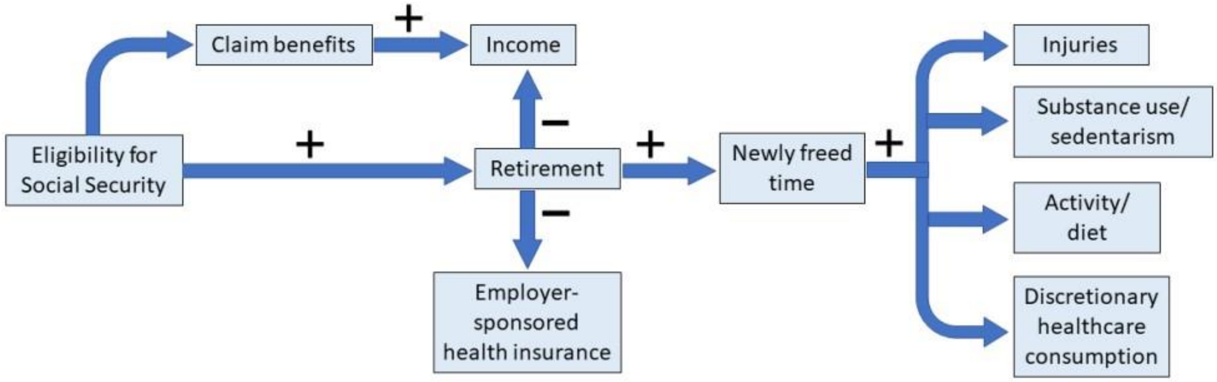
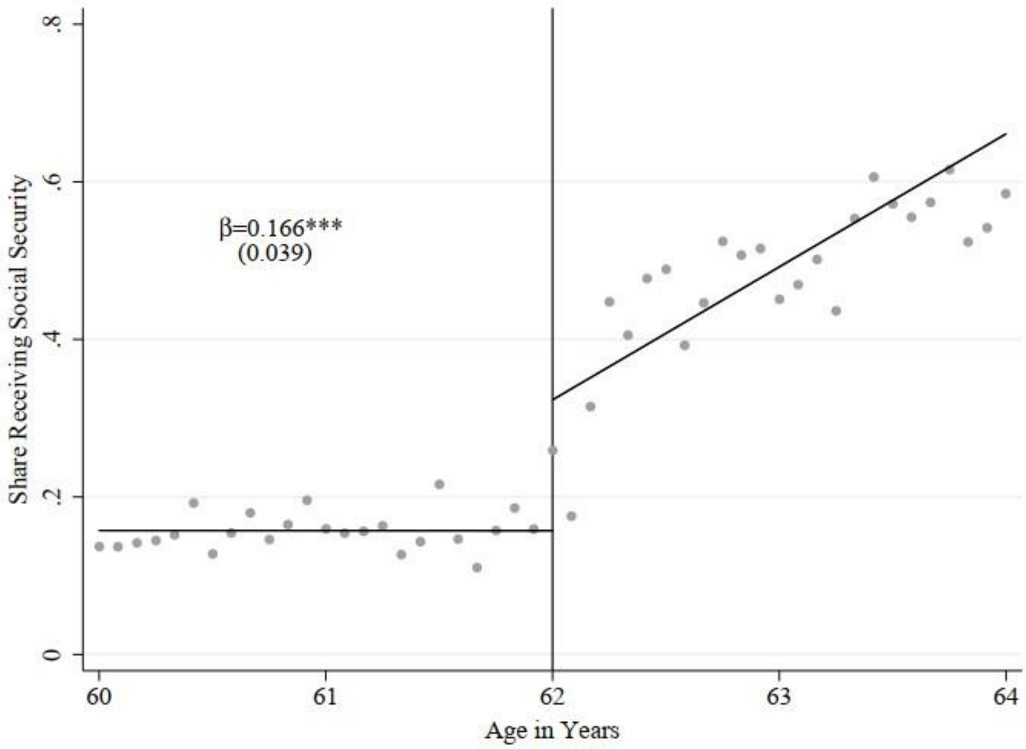
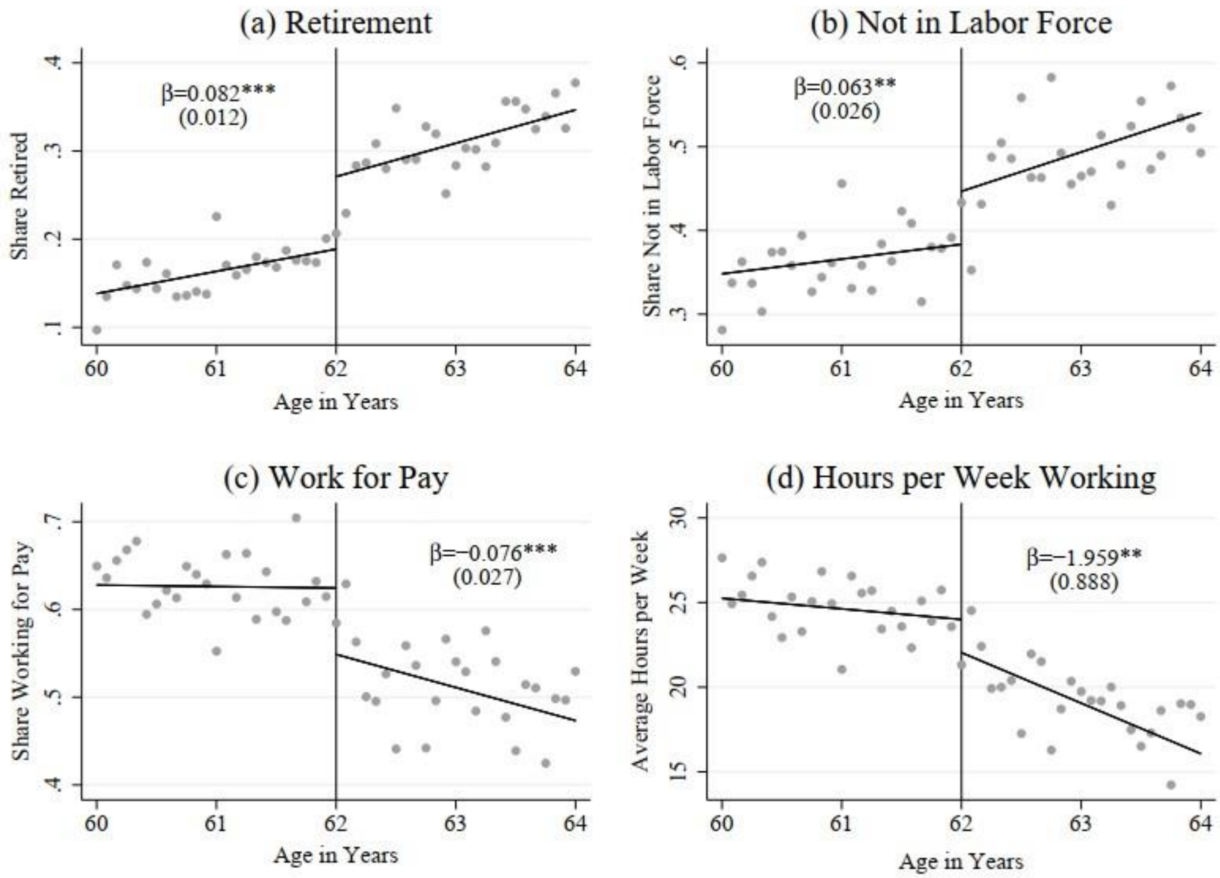


Figure 2: People Join Social Security Discontinuously at 62



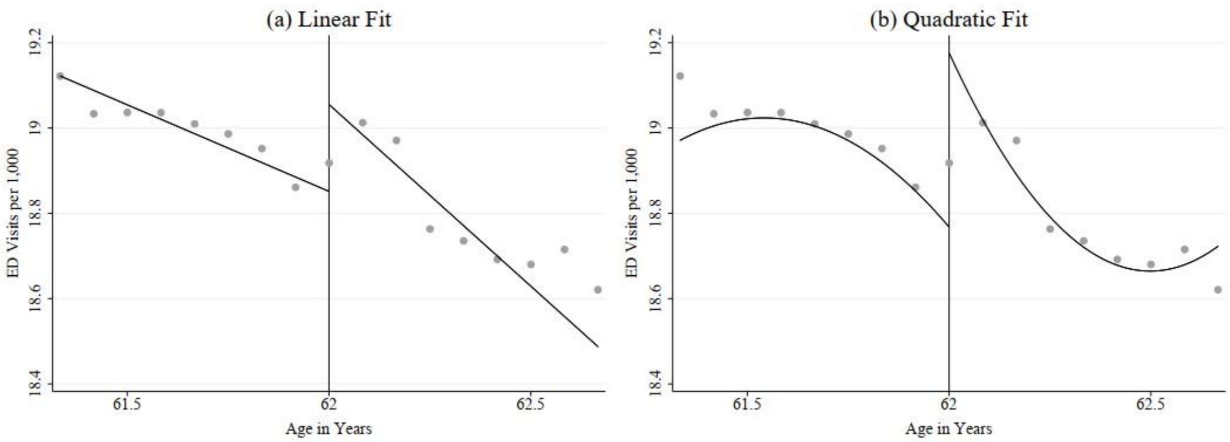
Note: This figure plots the share of the U.S. population currently receiving Social Security Retirement, Disability, or Survivor Insurance by age. Linear fit generated by estimating equation (2) with a 24-month bandwidth and triangular kernel. Data are from the Health and Retirement Study waves 2006-2018. Each age bin corresponds to the calendar month that people turned each age (e.g., share receiving income from Social Security in the calendar month people turn 62). Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01.

Figure 3: People Leave Work Discontinuously at 62



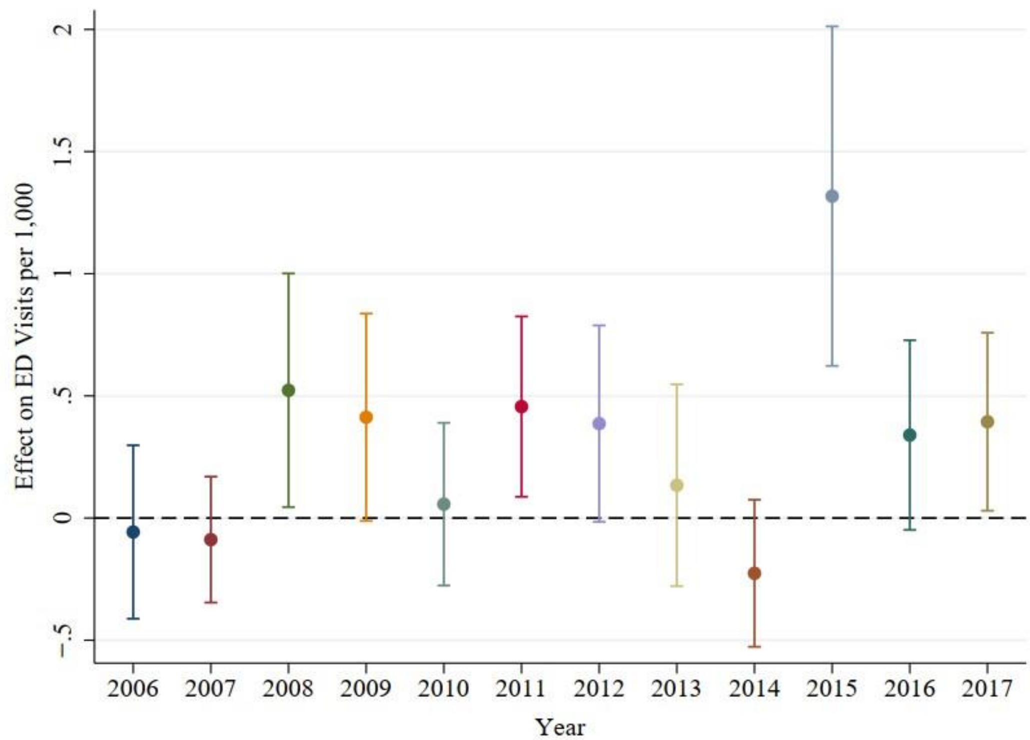
Note: This figure plots the share retired, not in the labor force, and working for pay, by age. It also plots the average number of hours people report working by age. Linear fits are generated by estimating equation (2) with 24-month bandwidths using triangular kernels. Data are from the Health and Retirement Study waves 2006-2018. Each age bin corresponds to the calendar month that people turned each age (e.g., share retired in the calendar month people turn 62). Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01.

Figure 4: Aggregate ED Visits Increase Visibly at 62



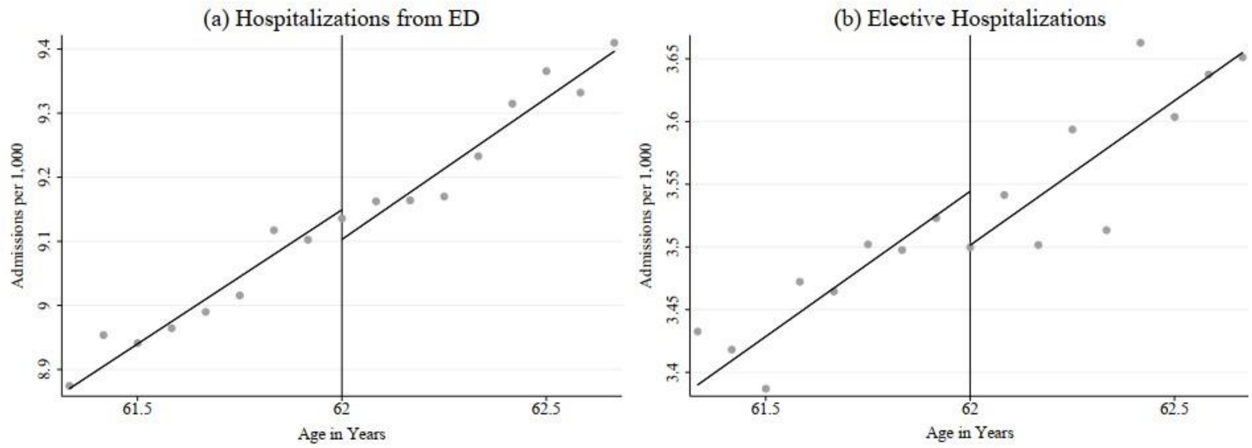
Note: This figure plots ED visits per 1,000 population by age. Data are from HCUP NY SEDD and CA OSHPD ED visit data, years 2006-2017. Each age bin corresponds to the calendar month that people turned each age (e.g., ED visits in the calendar month people turn 62). Polynomial fits generated from estimating equation (1) with 8-month bandwidths and triangular kernels.

Figure 5: The Effect on ED Visits Remains Stable Over the Time Period



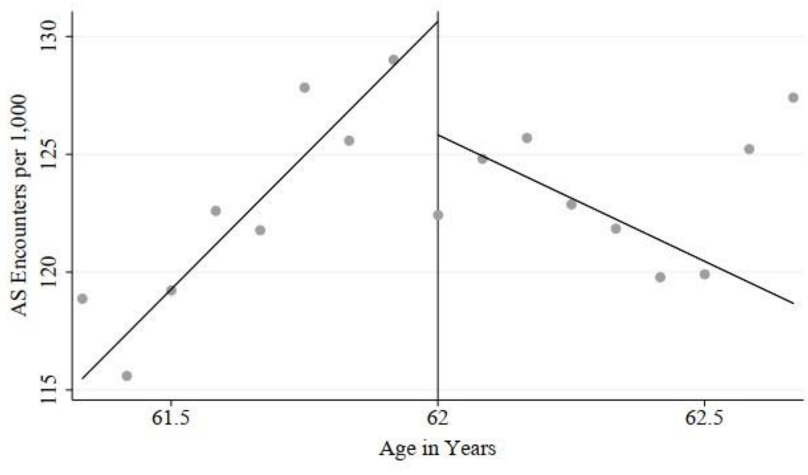
Note: This figure plots the RD effect on ED visits per 1,000 population by year. Data are from HCUP NY SEDD and CA OSHPD ED visit data, years 2006-2017. Each estimate is made by estimating equation (1) using a linear fit, triangular kernel, and CCT optimal bandwidth. 95% confidence intervals generated using robust standard errors.

Figure 6: Aggregate Inpatient Hospitalizations Do Not Change at 62



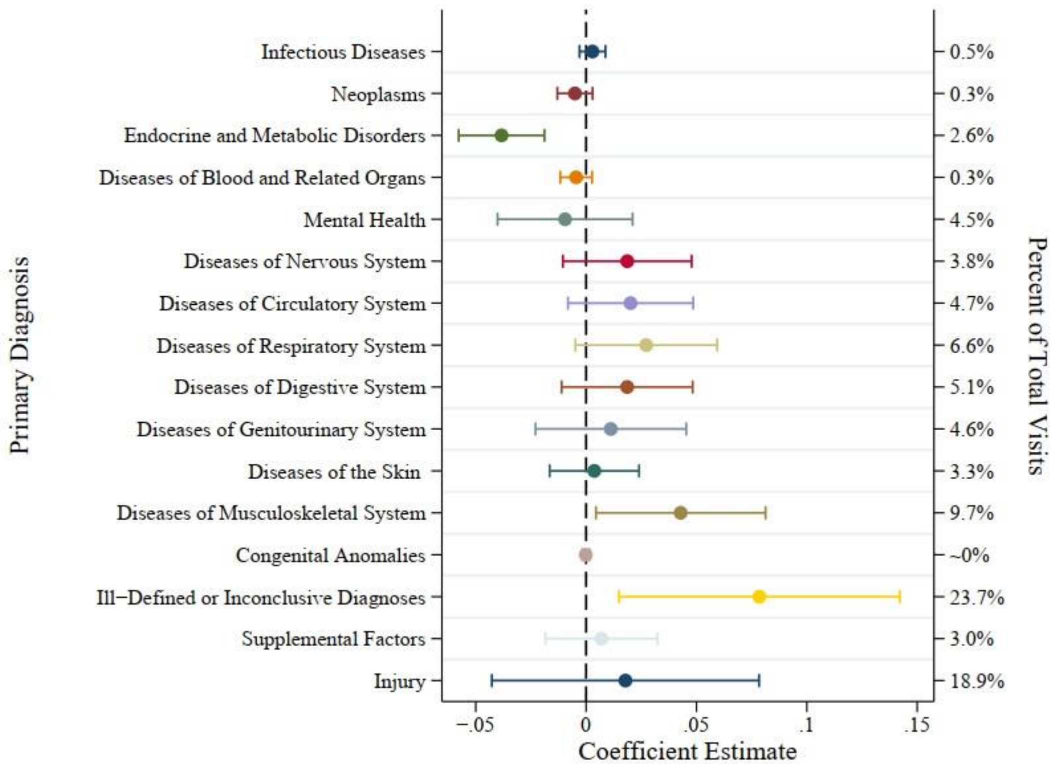
Note: This figure plots inpatient admissions per 1,000 population by age. Data are from HCUP NY SID and CA OSHPD PDD data, years 2006-2017. Each age bin corresponds to the calendar month that people turned each age (e.g., admissions in the calendar month people turn 62). Linear fit generated from estimating equation (1) with linear fit, triangular kernel, and 8-month bandwidth.

Figure 7: Ambulatory Surgery Encounters in CA Do Not Change at 62



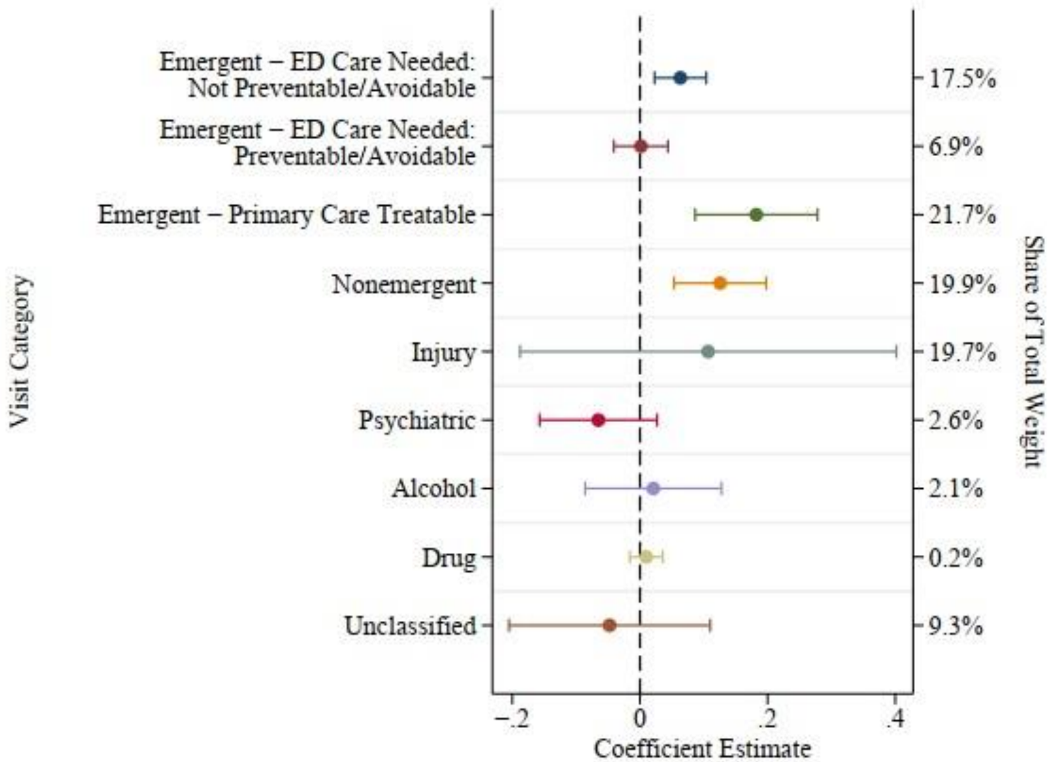
Note: This figure plots ambulatory surgery ancounters per 1,000 population by age. Data are from CA OSHPD AS data, years 2006-2017. Each age bin corresponds to the calendar month that people turned each age (e.g., admissions in the calendar month people turn 62). Linear fit generated from estimating equation (1) with linear fit, triangular kernel, and 8-month bandwidth.

Figure 8: Effect on ED Visits Not Driven by Any Particular Diagnoses



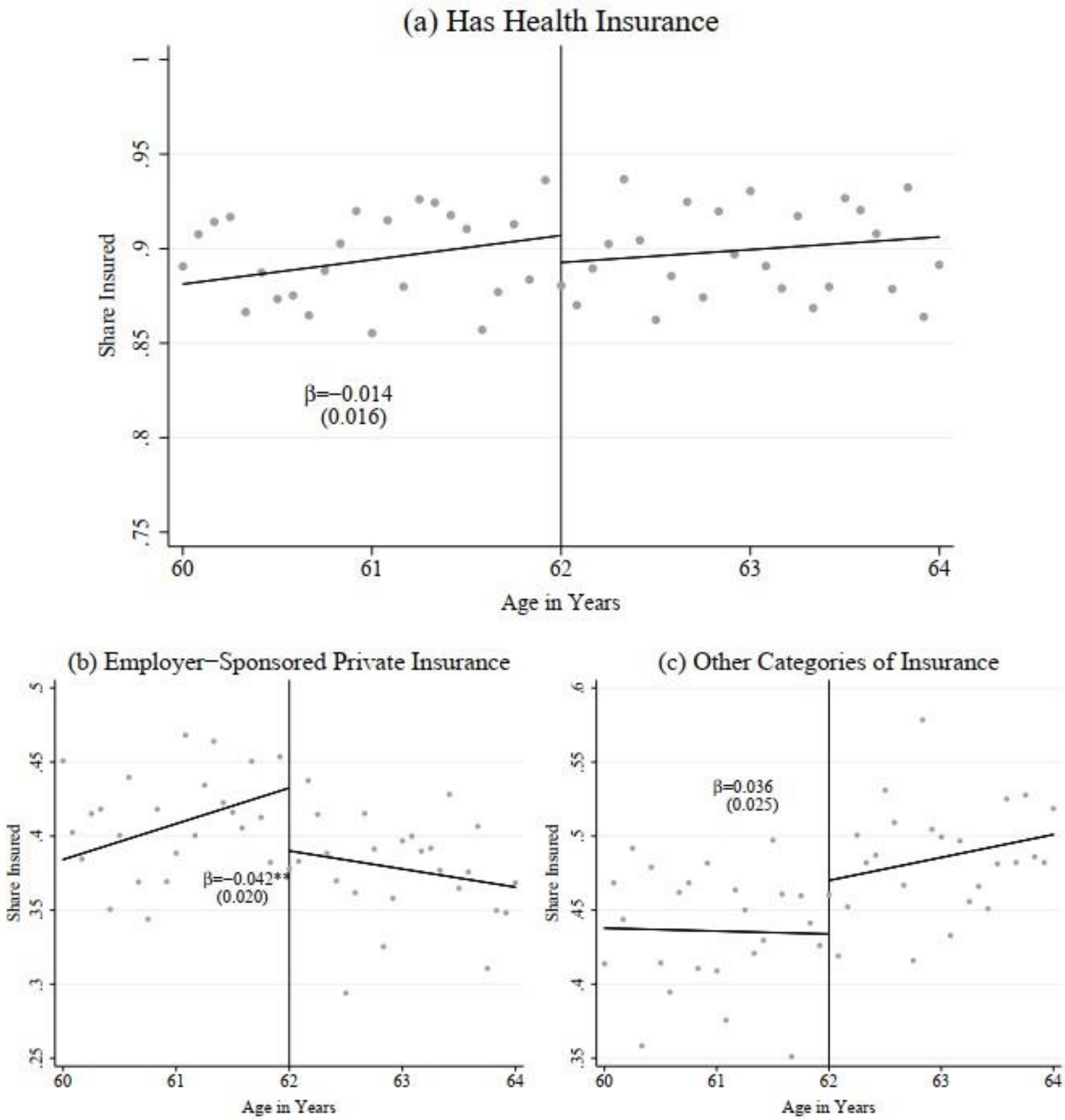
Note: This figure plots the RD effect on ED visits per 1,000 population by primary diagnosis. Data are from HCUP NY SEDD and CA OSHPD ED visit data, years 2006-2017. Each estimate is made by estimating equation (1) using a linear fit, triangular kernel, and CCT optimal bandwidth. Percent of total visits is calculated based on the month before people turn 62. 95% confidence intervals generated using robust standard errors.

Figure 9: Effect on ED Visits Drive by Emergent and Nonemergent Conditions



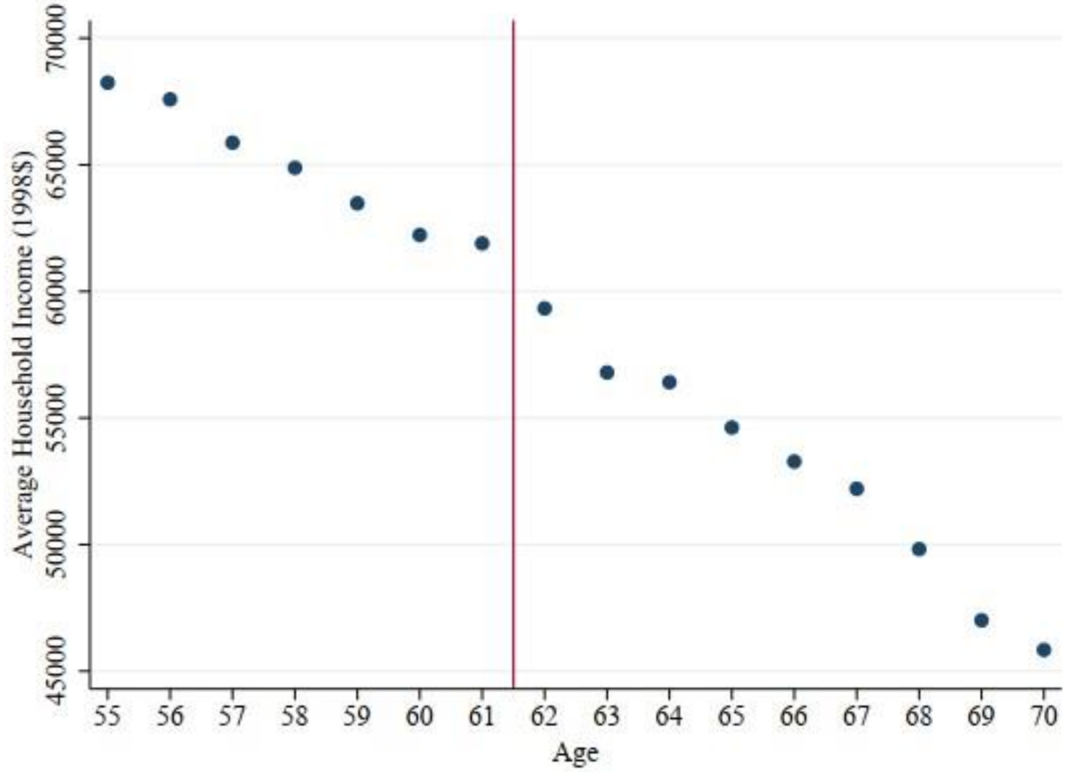
Note: This figure plots the RD effect on ED visits per 1,000 population according to NYU algorithm classification. The outcome for each category of visit is the sum of the probability weight across all visits in each age-state-year bin. Data are from HCUP NY SEDD and CA OSHPD ED visit data, years 2006-2017. Each estimate is made by estimating equation (1) using a linear fit, triangular kernel, and CCT optimal bandwidth. Percent of total visits is calculated based on the month before people turn 62. 95% confidence intervals generated using robust standard errors.

Figure 10: Employer-Sponsored Insurance Decreases at 62, but Overall Rates Stay the Same



Note: Panel (a) plots the share of share of people with at least one health insurance plan by age. Panel (b) plots the share of people with an employer-sponsored private insurance plan. Panel (c) plots the share of people with at least one form of another type of insurance. Data are from the Health and Retirement Study waves 2006-2018. Linear fits generated by estimating equation (2) with 24-month bandwidth and triangular kernel. Each age bin corresponds to the calendar month that people turned each age (e.g., fraction insured in the calendar month people turn 62). Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01.

Figure 11: Average Household Income Does Not Change Meaningfully at 62



Note: This figure plots average household income by year of age from the CPS-ASEC years 2006-2017.

Table 1: How the Effects on Social Security and Labor Market Outcomes Vary by Sex and Race

Outcome:	(1) **Social Security**	(2) **Retired**	(3) **Not in Labor Force**	(4) **Work for Pay**
Sex:				
Male	0.159***	0.082***	0.067**	-0.085***
	(0.050)	(0.022)	(0.033)	(0.032)
Female	0.173***	0.078***	0.061	-0.071*
	(0.041)	(0.028)	(0.038)	(0.041)
Race:				
White	0.176***	0.099***	0.093***	-0.099***
	(0.041)	(0.015)	(0.031)	(0.032)
Hispanic	0.145***	0.111***	-0.017	-0.030
	(0.058)	(0.039)	(0.072)	(0.075)
Black	0.097*	-0.021	-0.057	0.027
	(0.052)	(0.043)	(0.045)	(0.042)
Other Races/Ethnicities	0.350***	-0.040	0.068	-0.070
	(0.095)	(0.063)	(0.089)	(0.086)

Note: Regressions are estimated using equation (2) with the Health and Retirement Study waves 2006-2018. All regressions use triangular kernel, linear polynomial, and 24-month bandwidths. Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01.

Table 2: Regression Estimates of the Effect on ED Visits per 1,000 Population

	(1)	(2)	(3)	(4)	(5)	(6)
Sample: Both States						
Aged 62+	0.204***	0.409***	0.292***	0.397***	0.357***	0.315***
	(0.064)	(0.108)	(0.074)	(0.105)	(0.096)	(0.109)
			[0.001]	[0.001]	[0.000]	[0.005]
BW	8	8	5.5	8.1	3.9	8.3
Sample: New York Only						
Aged 62+	0.234**	0.550***	0.346***	0.548***	0.400***	0.569***
	(0.096)	(0.162)	(0.107)	(0.162)	(0.129)	(0.176)
			[0.001]	[0.002]	[0.002]	[0.001]
BW	8	8	6.1	7.7	4.5	7.8
Sample: California Only						
Aged 62+	0.174**	0.267*	0.213**	0.232*	0.208**	0.216
	(0.086)	(0.147)	(0.102)	(0.130)	(0.101)	(0.135)
			[0.121]	[0.157]	[0.082]	[0.187]
BW	8	8	5.6	9.8	5.6	9.9
Poly. Deg.	1	2	1	2	1	2
CCT?			X	X	X	X
Kernel	Triangular	Triangular	Triangular	Triangular	Uniform	Uniform

Note: Regressions are estimated using equation (1) using NY HCUP SEDD and CA OSHPD ED data for years 2006-2017 with ED visits per 1,000 population as the outcome. The odd-numbered columns estimate the model using a linear fit and the even-numbered columns estimate the model with a quadratic fit. Columns (1)-(2) use a set bandwidth of 8 months while columns (3)-(6) use CCT optimal bandwidths. Columns (1)-(4) use triangular kernels and columns (5)-(6) use uniform kernels. The sample range for CCT optimal bandwidth calculations is 24 months on either side of the cutoff. Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01. Brackets contain p-values from bias-corrected confidence intervals.

Table 3: How RD Estimates for Healthcare Encounters Vary by Sex and Race

Outcome:	(1) **ED Visits**	(2) **Inpatient Admissions - ED**	(3) **Inpatient Admissions - Elective**	(4) **AS Encounters**
Sex:				
Male	0.011**	-0.001	-0.012	0.015
	(0.005)	(0.009)	(0.010)	(0.013)
	[0.025]	[0.926]	[0.207]	[0.141]
Female	0.015**	-0.004	-0.016	-0.005
	(0.006)	(0.008)	(0.011)	(0.008)
	[0.033]	[0.491]	[0.151]	[0.726]
Race:				
White	0.024***	0.012	-0.010	0.010
	(0.006)	(0.010)	(0.010)	(0.010)
	[0.001]	[0.195]	[0.334]	[0.246]
Hispanic	-0.004	0.005	-0.006	-0.009
	(0.009)	(0.011)	(0.031)	(0.013)
	[0.753]	[0.658]	[0.930]	[0.407]
Black	0.004	-0.037***	0.002	0.000
	(0.010)	(0.013)	(0.032)	(0.022)
	[0.574]	[0.006]	[0.817]	[0.824]
Other Races/Ethnicities	0.009	-0.026*	-0.031	-0.012
	(0.008)	(0.015)	(0.029)	(0.014)
	[0.335]	[0.111]	[0.432]	[0.628]

Note: Regressions are estimated using equation (1) with triangular kernel, linear polynomial, and CCT optimal bandwidths. The sample range for CCT optimal bandwidth calculations is 24 months on either side of the cutoff. All outcomes are specified as the natural logs of various types of healthcare encounters. ED visit data are taken from NY HCUP SEDD and CA OSHPD ED data. Inpatient hospitalization data are taken from NY HCUP SID and and CA OSHPD PDD. Ambulatory surgery data are taken from CA OSHPD AS data. All data from years 2006-2017. Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01. Brackets contain p-values from bias-corrected confidence intervals.

Table 4: Regression Estimates of the Effect on Inpatient Hospitalizations per 1,000 Population

	(1)	(2)	(3)	(4)
Sample: Hospitalizations from the ED				
Aged 62+	-0.047	-0.022	-0.034	-0.027
	(0.047)	(0.081)	(0.057)	(0.079)
			[0.589]	[0.798]
BW	8	8	5.5	8.2
Sample: Elective Hospitalizations				
Aged 62+	-0.040	0.002	-0.038	0.010
	(0.026)	(0.004)	(0.025)	(0.042)
			[0.160]	[0.705]
BW	8	8	8.7	7.4
Poly. Deg.	1	2	1	2
CCT?			X	X

Note: Regressions come from estimating equation (1) using using NY HCUP SID and CA OSHPD PDD data for years 2006-2017. Outcomes are inpatient admissions per 1,000 population. All regressions use triangular kernels. The sample range for optimal bandwidth calculations is 24 months on either side of the cutoff. Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01. Brackets contain p-values from bias-corrected confidence intervals.

Table 5: Regression Estimates of the Effect on Ambulatory Surgery per 1,000 Population

	(1)	(2)	(3)	(4)
Sample: Aggregate				
Aged 62+	-0.029	0.014	0.035	0.137
	(0.060)	(0.104)	(0.074)	(0.105)
			[0.444]	[0.172]
BW	8	8	5.4	7.9
Poly. Deg.	1	2	1	2
CCT?			X	X

Note: Regressions come from estimating equation (1) using CA OSHPD AS data for years 2006-2017. Outcomes are ambulatory surgery encounters per 1,000 population. All regressions use triangular kernels. The sample range for optimal bandwidth calculations is 24 months on either side of the cutoff. Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01. Brackets contain p-values from bias-corrected confidence intervals.

1.9. APPENDIX A: VARIABLE CONSTRUCTION

Population Denominator

I model my construction of the population denominator on the method developed in Arenberg et al. (2020). In this method, I approximate the population for each state-by-year-by-age-by-calendar month cell by combining historical vital statistics data on births by month between 1941-1958 with adjustments from state population estimates from the 10% 2010 Decennial U.S. Census (Ruggles et al., 2020). Specifically, the approximation for each cell's population value is written in equation (A.1):

$$\widehat{pop}_{asym} = births_{asym} * \frac{pop2010_{asyq}}{births_{asyq}} \tag{3}$$

$\widehat{pop}_{asym}$ is an estimate of the cohort size of people aged a months old in state s in year y and were born on month m. $births_{asym}$ is the number of people of age (in months) a in state s on year y that were born in calendar month m. Births alone is an insufficient measure of population size between 2006-2017 since people may have either died or moved between when they were born and this time period. Therefore, I adjust monthly birth counts each quarter by $\frac{pop2010_{asyq}}{births_{asyq}}$, the ratio of the 2010 population of each age cohort by state and quarter-of-birth[18] to $births_{asyq}$, which is the number of births aggregated to quarter-of-birth instead of month-of-birth. Since my outcomes are aggregated by state, age, and year, I must take the sum of $\widehat{pop}_{asym}$ across calendar months of birth to calculate the denominator for regressions. Thus, my final denominator is displayed in equation (A.2), where M is a set containing all months of birth for

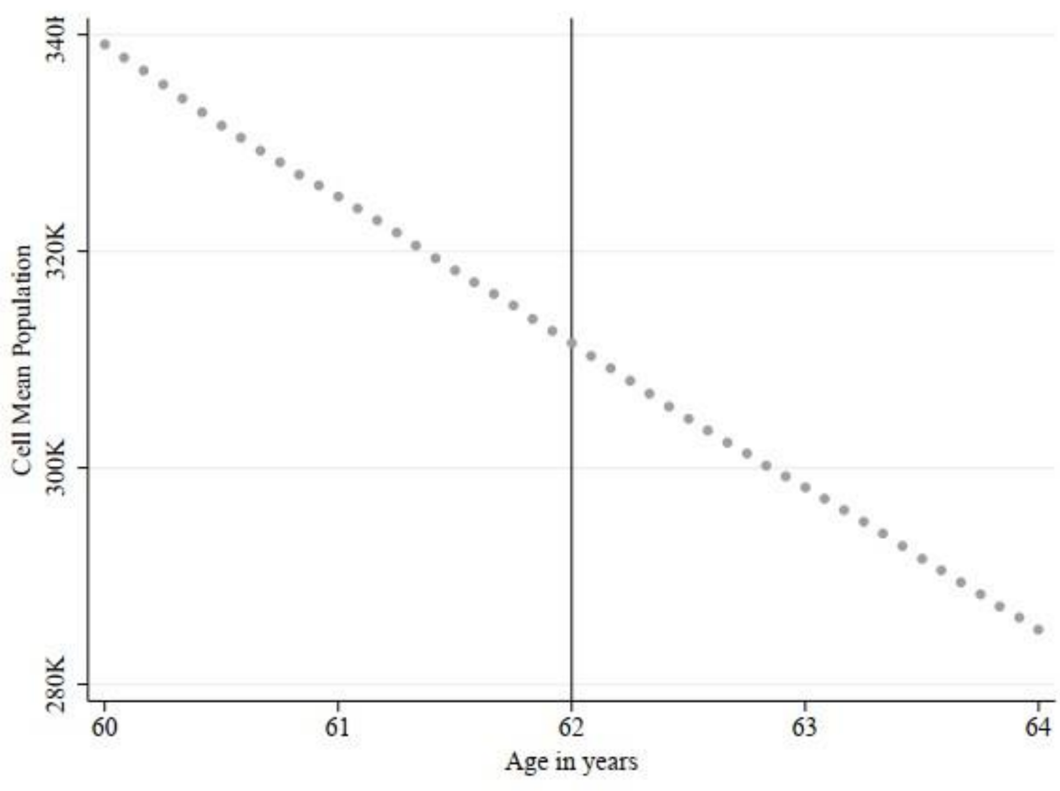
[18] Estimated by multiplying the 10% Census sample's population count by 10.

people of age a in state s and year y. The population estimate for each age, state, and year cell can be interpreted as the total number of people in a given state and year that are ever a given age (e.g., the total number of people who are ever 62 and 1 month old in NY in 2006).

$$\overline{pop}_{asy} = \sum_{m \in M} \widehat{pop}_{asym} \tag{4}$$

Appendix Figure A.1 shows the smoothness of the population estimates within a two-year radius of the age 62 cutoff. Each dot represents the mean population for each month of age across years and states. This figure indicates that my method does not produce “jumpy” population estimates that would inappropriately affect RD estimates.

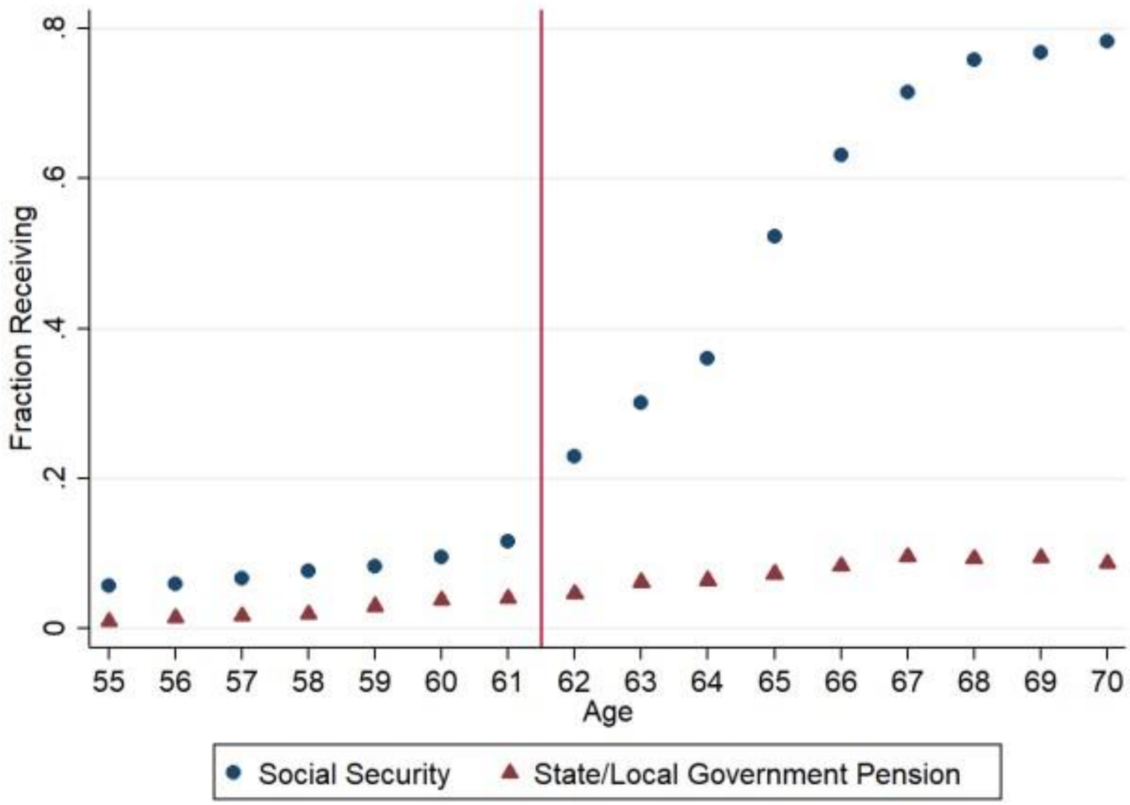
Figure A.1: The Population Denominator is Smooth Through the Cutoff



Note: This figure plots the estimated population counts by month of age. Estimates are derived from from 1941-1958 U.S. vital statistics data combined with the publicly available 10% extract of the 2010 U.S. Census.

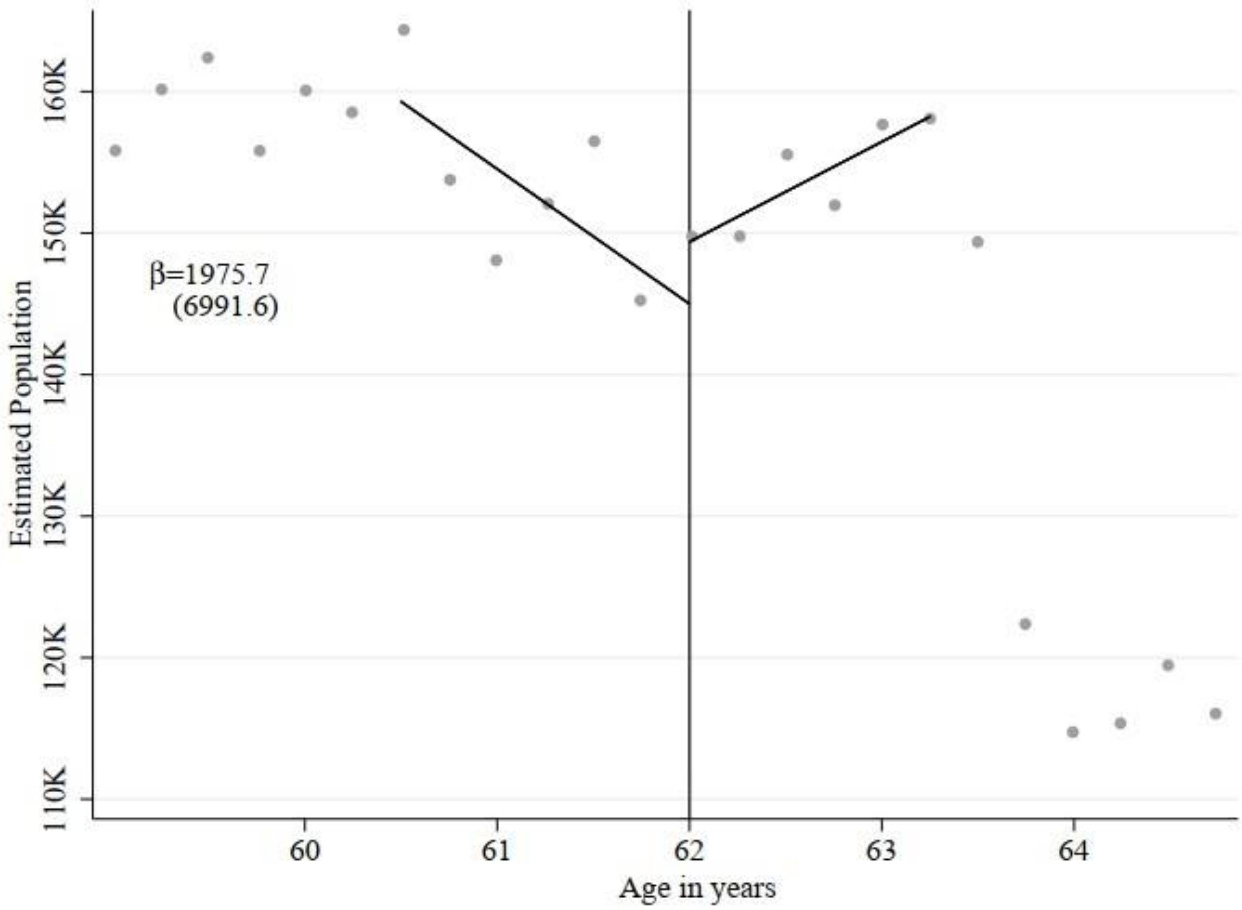
1.10. APPENDIX B: SUPPLEMENTAL RESULTS

Figure B.1: Social Security Claiming Increases Meaningfully in NY and CA at 62, State and Local Pension Claiming Do Not



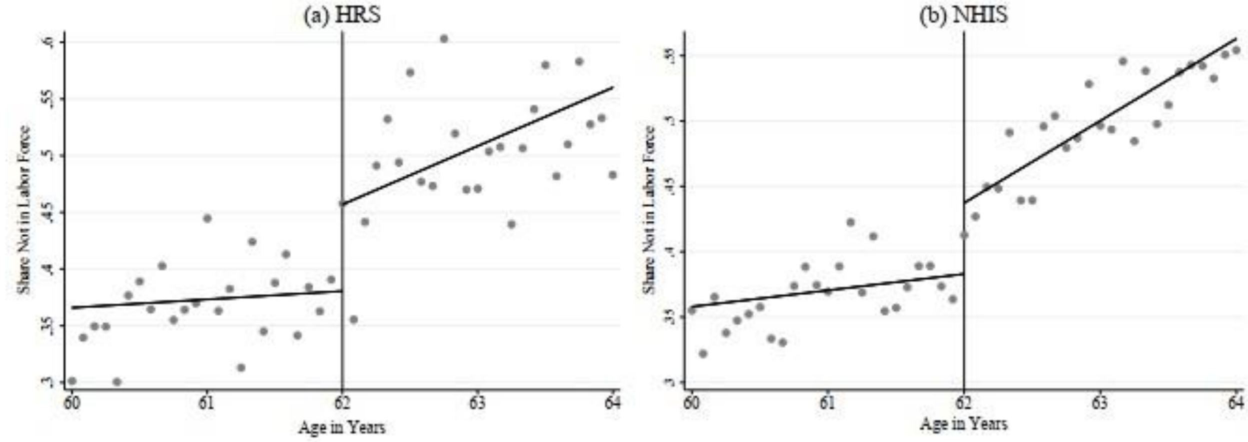
Note: This figure plots rates of Social Security and State/Local Government Pension receipt in NY and CA by age in 2006-2017. Data are taken from the CPS-ASEC 2007-2018 waves, which asks about people's sources of retirement income from the previous year. Since the CPS-ASEC is conducted every March and asks about the previous year, ages are calculated by subtracting one year from the respondent's current age in years.

Figure B.2: The Aggregate Population in NY and CA Does Not Change Discontinuously at 62



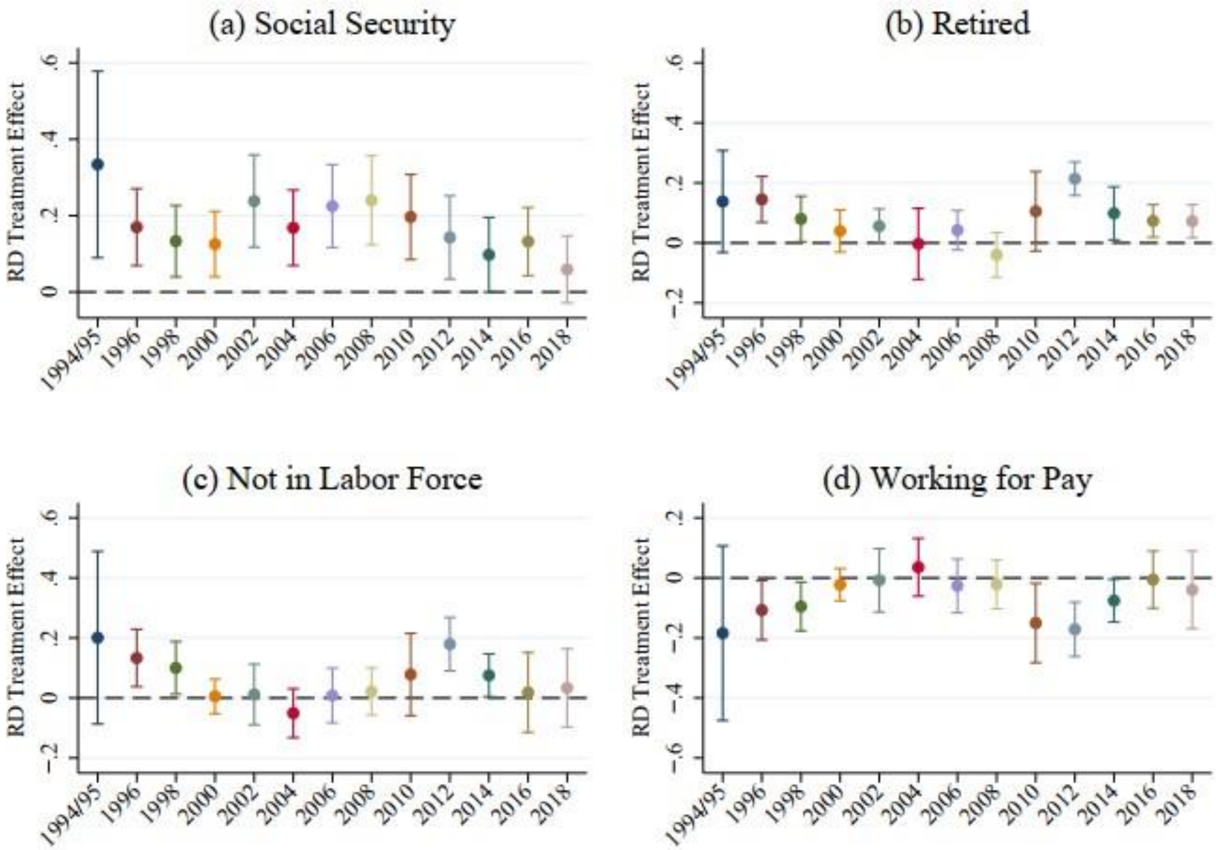
Note: This figure plots the estimated population size in NY and CA by quarter of age using the 10% sample of the 2010 Decennial Census. Population counts for each age bin are multiplied by 10 to account for the 10% random sample. Linear fits are estimated using triangular kernels and 6-quarter (18-month) bandwidths.

Figure B.3: The Discontinuity in Labor Force Participation Occurs Immediately at 62



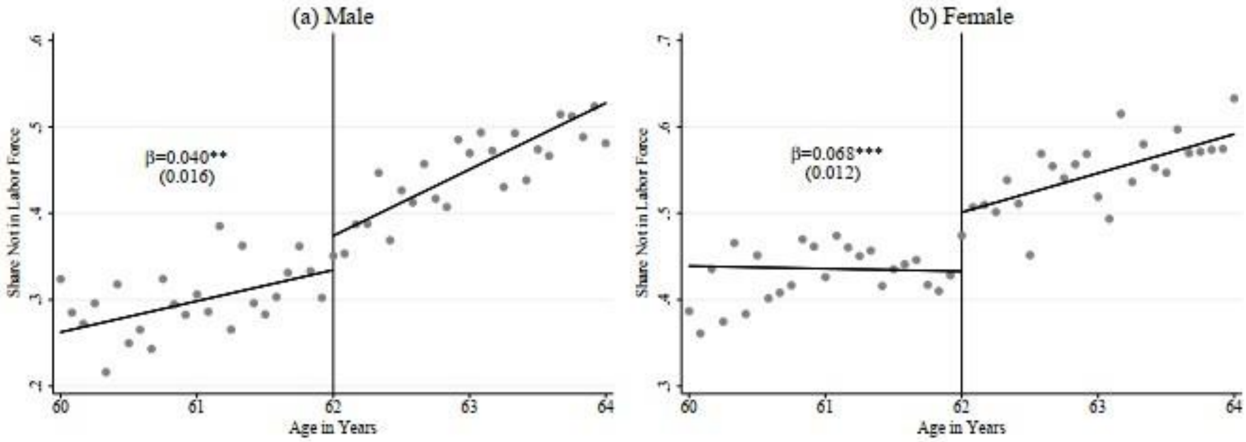
Note: This figure plots the share of the population currently not in the labor force. Linear fit obtained form estimating equation (2) with a 24-month bandwidth and triangular kernel. Data are from the Health and Retirement Study waves 2006-2014 and National Health Interview Survey waves 2006-2014. Each age bin corresponds to the calendar month that people turned each age (e.g., share working in the calendar month people turn 62).

Figure B.4: Effect on Social Security and Labor Market Outcomes by Year



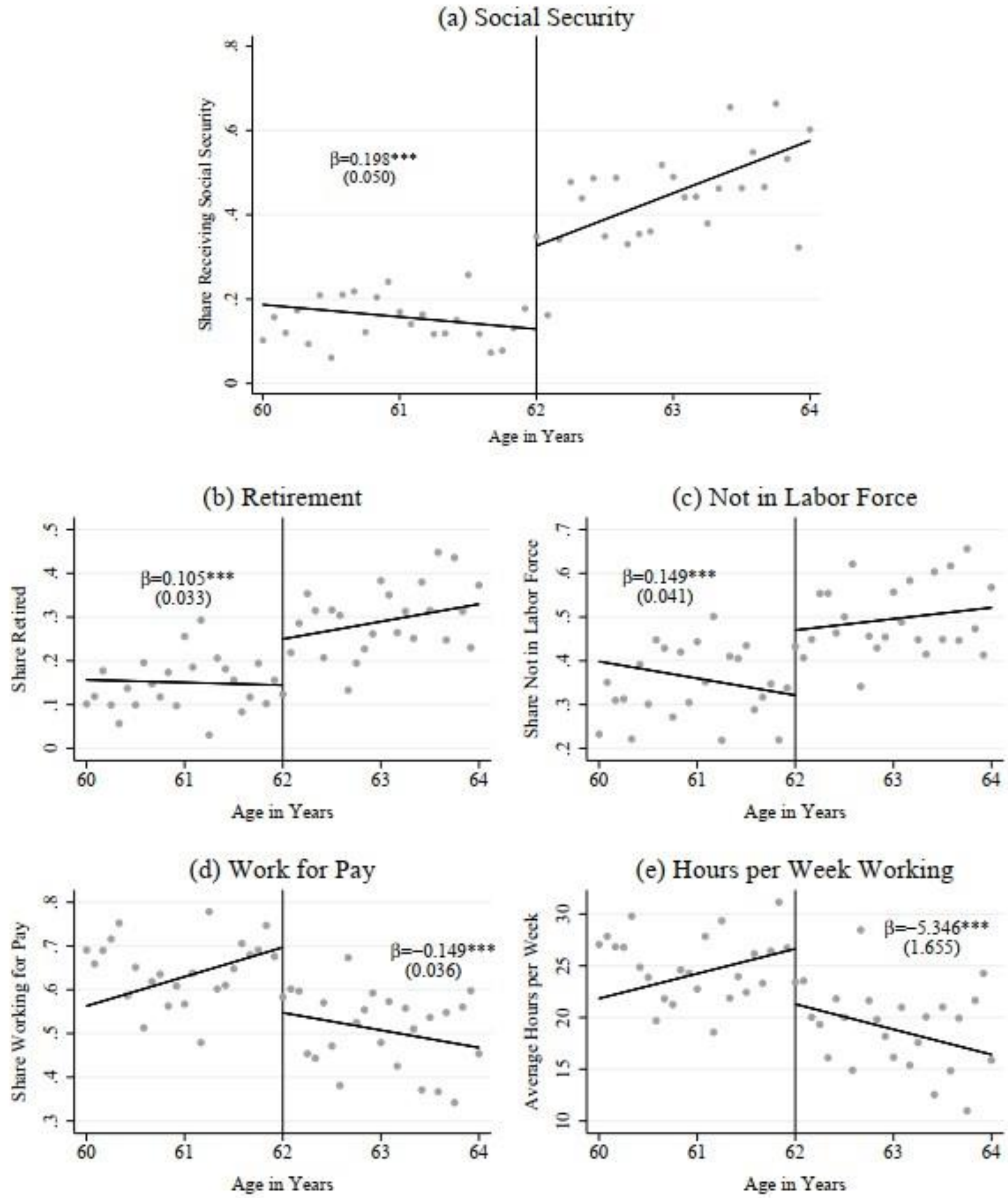
Note: This figure plots the effect of turning 62 on Social Security receipt and labor market outcomes by year by estimating β using equation (2). Data are from the Health and Retirement Study waves 1994/95-2018. Each specification controls for a dummy for the month people turn 62. 95% confidence intervals are derived from robust standard errors.

Figure B.5: The Discontinuity in Female Labor Force Participation is Also Present in the NHIS



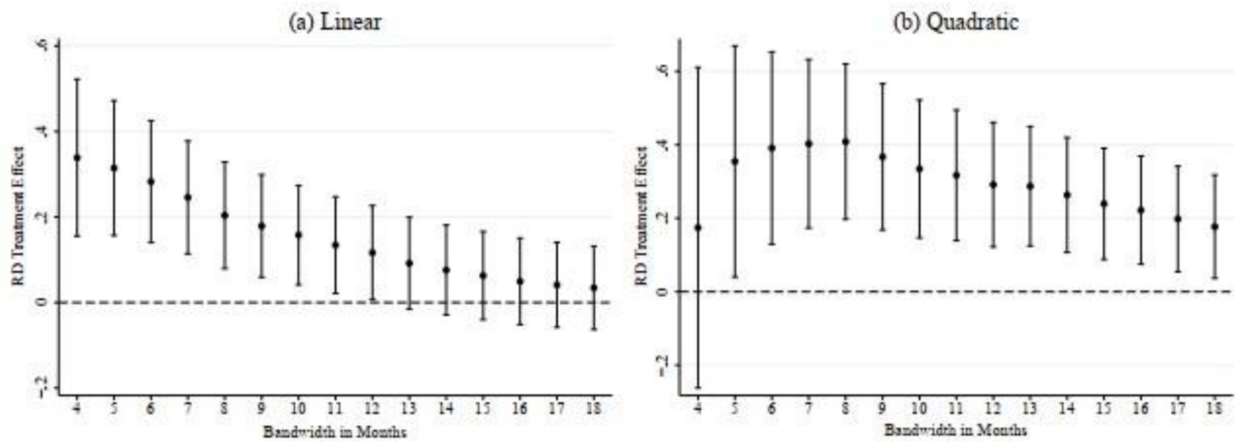
Note: This figure plots the share of the population currently not in the labor force. Linear fit obtained from estimating equation (2) with a 24-month bandwidth and triangular kernel. Data are from the National Health Interview Survey waves 2006-2014. Each age bin corresponds to the calendar month that people turned each age (e.g., share working in the calendar month people turn 62). Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01.

Figure B.6: The Effects on Social Security Uptake and Labor Market Outcomes: Pacific and Middle Atlantic Census Divisions



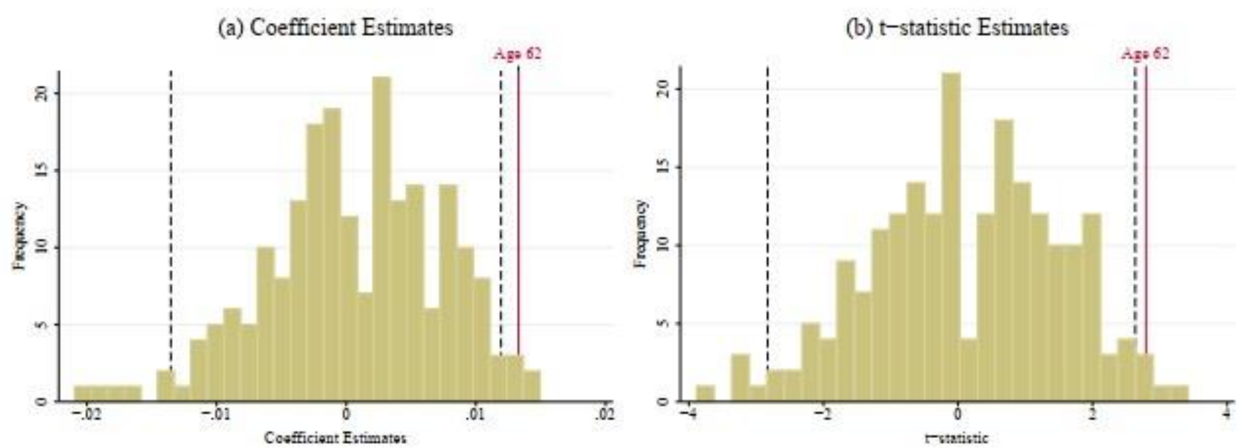
Note: These figures plot the discontinuity in Social Security receipt and labor outcomes for people in the Pacific and Middle Atlantic Census Divisions. Linear fit obtained from estimating equation (2) with a 24-month bandwidth and triangular kernel. Data are from the Health and Retirements Study waves 2006-2018. Each age bin corresponds to the calendar month that people turned each age (e.g., share working in the calendar month people turn 62). Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01.

Figure B.7: Estimated Effect on ED Visits is Robust to Many Bandwidths, Particularly Narrow Ones



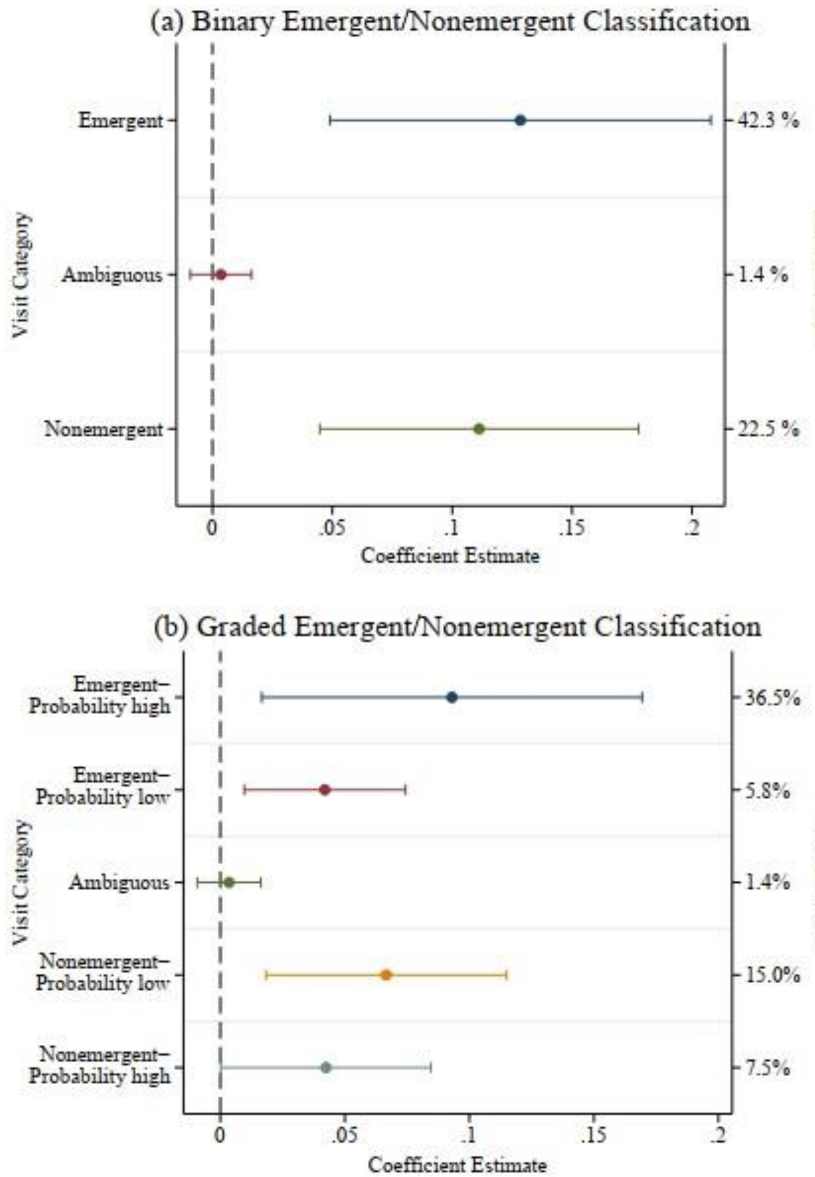
Note: This figure plots RD estimates using equation (1) by bandwidth in months and polynomial fit. Data come from NY HCUP SEDD and CA OSHPD ED for years 2006-2017. Outcomes are ED visits per 1,000 population. All regressions estimated with triangular kernels. 95% confidence intervals are calculated using robust standard errors.

Figure B.8: The Effect on ED Visits and Robust t-Statistic are Larger in Magnitude than 95% of Placebo Estimates



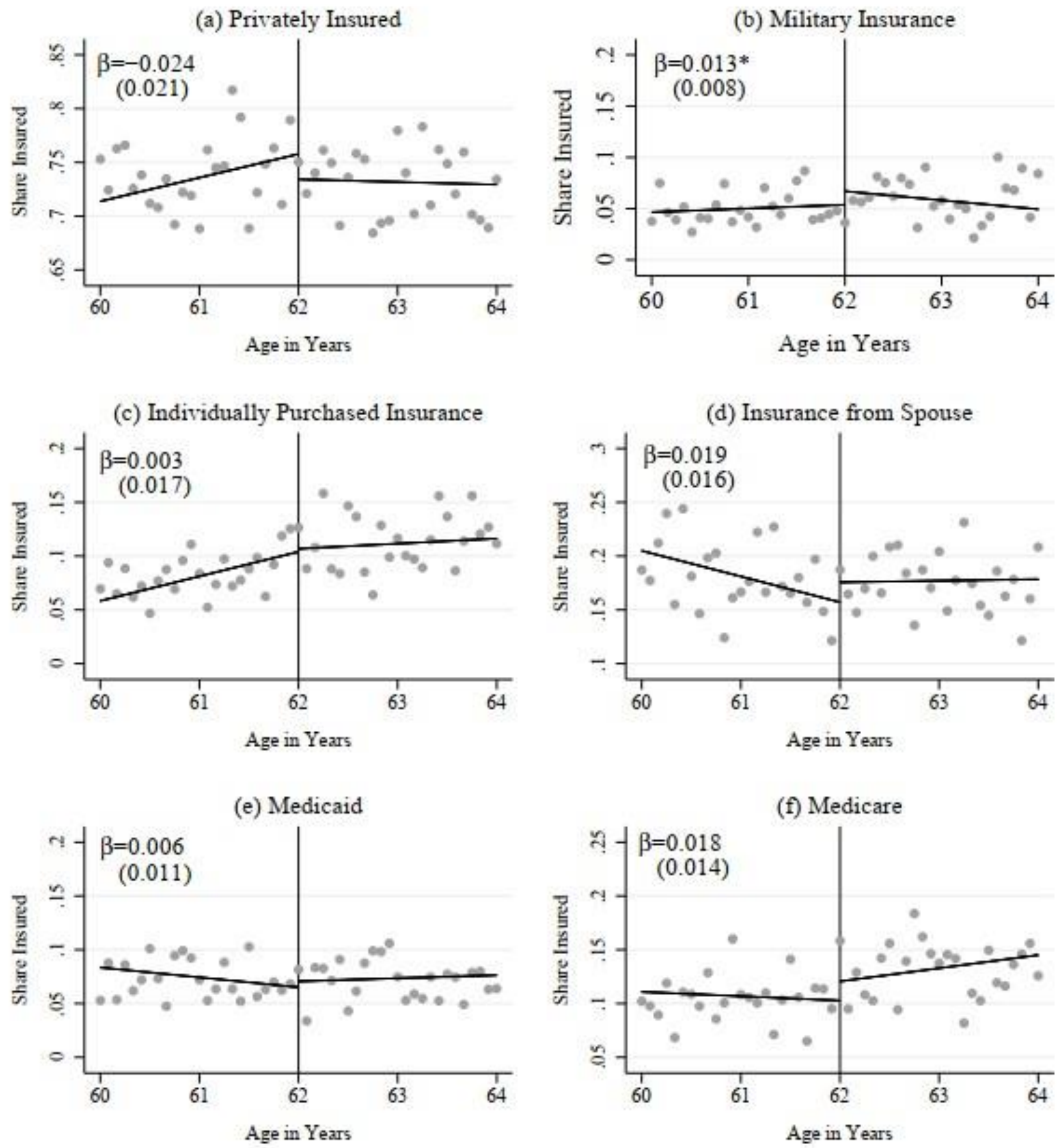
Note: This figure plots the distribution of RD coefficients and robust t-statistics, estimated by equation (1), at the age 62 cutoff and a set of placebo age cutoffs within a +/- 10-year radius. Data come from NY HCUP SEDD and CA OSHPD ED data for years 2006-2017. The outcomes are the natural log of aggregate ED visits. All regressions estimated with linear fits, CCT optimal bandwidths, and triangular kernels. The sample range for CCT optimal bandwidth calculations is 24 months on either side of the cutoff. I do not include coefficients using placebo cutoffs within 8-month radiuses of age 62 or age 65, or a cutoff at 65 itself, in order to avoid treatment effect contamination (Calonico and Titiunik, 2021). 95% of the estimated coefficients and t-statistics fall within the dotted black lines, and the solid red line indicates the estimate at age 62.

Figure B.9: Effect on Emergent and Nonemergent Visits Is Robust To Alternative Classification Methods



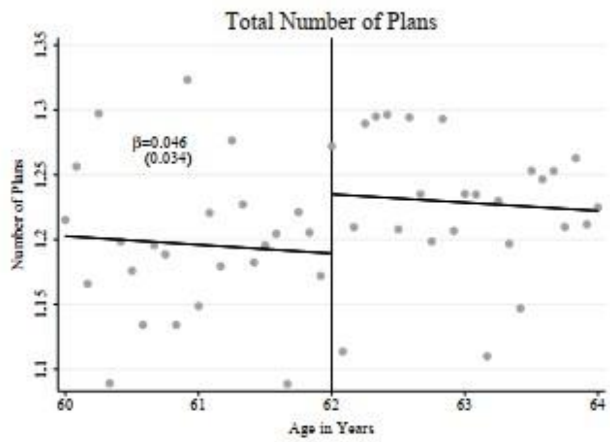
Note: This figure plots the RD effect on ED visits per 1,000 population as classified by the NYU algorithm via Johnston et al. (2017). Panel (a) classifies visits as emergent if the probability weights in each emergent category sum to greater than .5 and classifies visits as nonemergent if the same holds true for the nonemergent probability weight. Panel (b) does the same, but adds gradation of low and high probabilities for each category depending on whether the probabilitity weights sum between .5 to .75 or .75 to 1. Cases in which the probability weights for emergent categories and nonemergent categories are both .5 are classified as ambiguous. Data are from HCUP NY SEDD and CA OSHPD ED visit data, years 2006-2017. Each estimate is made by estimating equation (1) using a linear fit, triangular kernel, and CCT optimal bandwidth. Percent of total visits is calculated based on the month before people turn 62. 95% confidence intervals generated using robust standard errors.

Figure B.10: Effect on Insurance Coverage by Type of Payer



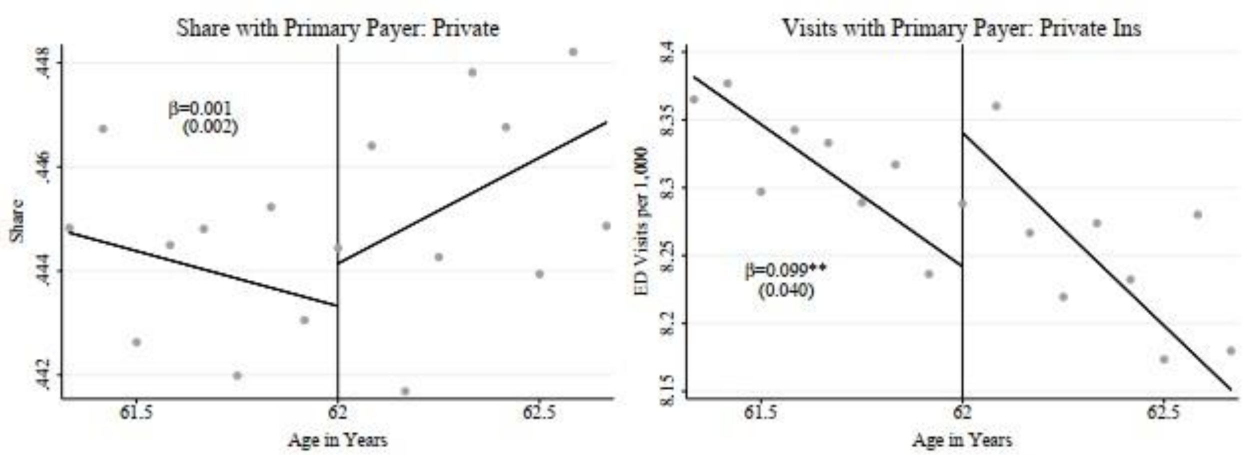
Note: This figure plots the rates of of insurance coverage by age and type of payer. Data are from the Health and Retirement Study waves 2006-2018. Linear fits are generated by estimating equation (2) with 24-month bandwidths and triangular kernel. Each age bin corresponds to the calendar month that people turned each age (e.g., share retired in the calendar month people turn 62). Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01.

Figure B.11: Effect on Number of Insurance Policies



Note: This figure plots the rates of of insurance coverage by age. Data are from the Health and Retirement Study waves 2006-2018. Linear fits are generated by estimating equation (2) with 24-month bandwidths and triangular kernel. Each age bin corresponds to the calendar month that people turned each age (e.g., share retired in the calendar month people turn 62). Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01.

Figure B.12: Effect on Share and Rate of ED Visits Among Privately Insured



Note: This figure plots the estimated discontinuities in the share and rate per 1,000 population of privately insured patients. Data come from NY HCUP SEDD and CA OSHPD ED data for years 2006-2017. Linear fits are generated by estimating equation (1) with 8-month bandwidths and triangular kernel. Each age bin corresponds to the calendar month that people turned each age (e.g., share retired in the calendar month people turn 62). Parentheses contain robust standard errors where * p <.1, ** p<.05, *** p<.01.

CHAPTER 2

HOW ACCESS TO ADDICTIVE DRUGS AFFECTS THE SUPPLY OF SUBSTANCE ABUSE TREATMENT: EVIDENCE FROM MEDICARE PART D

2.1. INTRODUCTION

In 2005, the Congressional Budget Office noted that by improving access to prescription drugs, Medicare Part D had the potential to increase the incidence of adverse drug events (Zhang et al., 2009). These concerns appear to have been well-founded. Recent analysis has determined that Medicare Part D exacerbated the opioid epidemic by increasing access to prescription opioids, thereby driving people into treatment for opioid-use disorder (OUD) (Powell et al., 2020). If this increase in addiction led to greater aggregate demand for substance abuse treatment (SAT), it is possible that Medicare Part D could have also induced the entry of new SAT providers to treat these individuals[1]. This is particularly important because SAT is perhaps the most effective available method of treating OUD, especially when conducted using medication-assisted treatment (MAT) (National Academies of Sciences, Engineering, and Medicine, 2019). Given the high costs to society imposed by substance-use disorders (Caulkins et al., 2014), and the current shortage of MAT availability (Jones et al., 2015; Dick et al., 2015), it is policy-relevant to understand whether the supply of SAT providers, both public and private, has been responsive to changes in population-level opioid addiction rates. If not, then it is possible that further policy intervention

[1] Gowrisankaran and Town (1997) model hospital profits to scale positively with the number of "sick" patients in a community. Their model predicts that an increase in these patients can induce hospital entry. I use this framework as motivation for my analysis of opioid addiction, acknowledging that differences exist between specialty SAT facilities and hospitals.

may be necessary to bolster the supply of treatment providers and MAT in addition to previous and ongoing efforts.

Although the question of whether increases in opioid addiction cause increases in the supply of SAT is fundamental for understanding the role of policymakers in addressing the opioid crisis, there exists little to no research on this topic. Furthermore, assessing this relationship empirically is not straightforward since simple correlations may produce misleading conclusions. For example, high addiction rates may be the product of local economic downturns, which could independently cause closures or prevent openings of SAT clinics. It is also possible that people are most likely to seek treatment when their opioid supply is cut off, which could attenuate the correlation between increases in opioid access and demand for SAT. To address these challenges, I use a quasi-experimental approach to circumvent these methodological issues and identify the causal impact of population-wide addiction on treatment capacity.

Estimating the effects of Medicare Part D on the supply of SAT is an empirical challenge since the program was implemented simultaneously for all Medicare beneficiaries on Jan 1st, 2006. In order to circumvent this difficulty, I follow the approach taken by Alpert et al. (2015) and Powell et al. (2020) by comparing states with higher shares of the population aged 65+ to states with lower shares, before and after the program's introduction in 2006. The premise of this method is that states with more Medicare-eligible individuals should have greater Part D enrollment (and, therefore, opioid use), on average, than states with fewer eligible people. Indeed, empirical tests confirm that high-share states had significantly greater Part D enrollment counts than low-share states. I exploit data on the near-universe of specialty licensed SAT facilities in the United States taken from the National Survey of Substance Abuse Treatment Services (N-SSATS) to estimate how Medicare D has affected the supply of SAT providers and services. Estimates suggest that a

10% increase in the supply of prescription opioids due to Part D increased residential and hospital inpatient SAT facilties by approximately 2.5% and beds by approximately 2.3%, compared to averages before Part D began in 2006. I also find that residential/hospital inpatient client counts increased by a similar magnitude, which implies that little to no additional within-facility crowding has resulted from this increase in addiction. Additionally, Part D has increased the number of SAT facilities offering medication-assisted treatment for OUD by 10.4%, an effect which was nearly entirely driven by the adoption of naltrexone.

This paper contributes to several strands of literature. Primarily, it contributes to our understanding of the determinants of healthcare access. The majority of the work on this topic has focused on the geographic distribution of physicians and the influence of public incentives on said distribution (Bärnighausen and Bloom, 2009). On the other hand, very little research has focused on whether treatment availability responds to demand, particularly with respect to SAT. The evidence that does exist in this space centers around insurance expansions rather than changes in opioid availability. For example, Maclean et al. (2018) find that state-mandated private insurance coverage of SAT services reduces provider participation in certain (public) insurance markets. On the other hand, Hamersma and Maclean (2021) and Meinhofer and Witman (2018) find that public insurance expansions change the types of services available from SAT providers. However, the effects of changes in opioid availability are likely to differ from those of specific insurance expansions if the newly addicted people are covered by a diversity of insurers or are uninsured. Furthermore, changes in opioid access could also affect other outcomes, such as employment rates, which may have independent impacts on health insurance status and, therefore, demand for treatment. Thus, it is difficult to project what the effect of increased addiction would be on the supply of SAT providers based solely on studies of insurance expansions. This paper also contributes to the literature on Medicare Part D's implications for supply-side actors (Dranove et al., 2020; Dranove et al., 2014; Blume-Kohout and Sood, 2013; Hu et al., 2017)[2]. Furthermore, it is the first paper to analyze Part D's effects on provider behavior that is unrelated to prescribing tendencies.

2.2. BACKGROUND

2.2.1. Medicare Part D and Opioid Diversion

Medicare Part D is an opt-in prescription drug insurance program for Medicare enrollees. Passed as part of the Medicare Modernization Act on December 8th, 2003, and implemented on January 1st, 2006, Part D was the largest expansion to the Medicare program since its inception in 1966. Before that, only certain non-prescription drugs were covered by Medicare Parts A and B, and enrollees had to have either their own supplemental drug coverage or pay out-of-pocket for prescription medication. This left approximately 25% of enrollees aged 65+ without prescription drug coverage as of 2003 (Safran et al., 2005). Take-up of Part D was rapid (Cubanski et al., 2019). Approximately 22 million people (51% of Medicare enrollees) opted-in after just the first year, which grew to about 43.5 million people (72% of Medicare enrollees) by 2018. Half-way through the first year of implementation, the fraction of seniors without prescription drug coverage had already dropped to around 10% (HHS, 2006). A wide body of research has examined how Part D decreased out-of-pockets costs and increased drug utilization among seniors (Duggan and Morton, 2010; Ketcham and Simon, 2008; Lichtenberg and Sun, 2007; Yin et al. 2008). While these studies

[2] Dranove et al. (2020), Dranove et al. (2014), and Blume-Kohout and Sood (2013) are primarily concerned with the policy's effects on prescription drug R&D. Hu et al. (2017) examines Part D's effects on the number of prescriptions made per physician visit.

typically assume that Medicare Part D's effects began in 2006, Alpert (2016) demonstrates that the program's announcement may have influenced utilization among the elderly in the years leading up to implementation (although this does not appear to have been an issue in the outcomes I study).

Medicare Part D has covered prescription opioids since its implementation in 2006 and appears to have increased uptake of the drug among enrollees. Powell et al. (2020) found that after Part D went into effect, people aged 66-71 experienced a 28% increase in opioid prescriptions relative to people aged 59-64. Furthermore, there is evidence that Medicare Part D also increased the prescription drug consumption of those below age 65. Alpert et al. (2015) find that non-elderly people in counties with higher shares of the population eligible for Medicare experienced larger increases in prescription drug consumption after 2006 than non-elderly people in counties with lower shares of Medicare-eligible people. Powell et al. (2020) translate this design to that state level and find that states with greater shares of the population aged 65+ experienced larger increases in opioid distributions per capita after Part D was implemented. Lastly, and importantly, they find that a 10% increase in opioid disbursement caused by Part D increased opioid mortality by 7.1% and opioid SAT admissions by 9.6%. Counterintuitively, these effects were entirely driven by people aged < 65. When taken in conjunction with the effects on opioid prescriptions for people aged 66+, Powell et al. (2020) conclude that a substantial portion of the opioids distributed through Part D were diverted away from their intended recipients and towards abuse. Given Part D's impacts on prescription opioid treatment admissions, it stands to reason that the policy may have increased aggregate demand for SAT, thereby encouraging entry among SAT providers.

2.2.2. Substance Abuse Treatment Capacity in the United States

The main platforms for delivering substance abuse treatment in the United States are community-based specialized SAT facilities (NIDA, 2018). These facilities offer several types of programming, which can be separated into two major treatment settings: outpatient and inpatient. Outpatient treatment ranges from simple drug education and counseling once per week, to intensive programs that meet every day (SAMHSA, 1997). Inpatient treatment can be separated into two further categories: residential and hospital. Residential inpatients stay in facilities under 24-hour supervision, the length of stay depending greatly on the individual. Hospital inpatients, on the other hand, usually require care for acute issues including severe overdoses, withdrawal, or complications from comorbidities. Stays in hospitals are typically shorter than stays in residential facilities. The majority of SAT clients are in outpatient treatment, representing about 90% of clients in treatment on March 31st, 2017 according to the N-SSATS. However, according to the Treatment Episode Data Set, outpatients only make up about 60% of new SAT admissions, implying that outpatients have longer stints of treatment than residential and hospital inpatient clients. Finally, one of the most important differences for this study between residential/hospital inpatient and outpatient care is how patient capacity is measured. For residential/hospital inpatient treatment, capacity is mostly determined by the number of beds per facility dedicated to SAT patients and, to a lesser degree, staffing levels. For outpatient treatment, determining capacity is less clear since the maximum number of clients does not depend on beds. Unlike beds, staffing data are not regularly collected by the N-SSATS[3].

An important component of modern SAT is the use of specialized medications to help people overcome addiction. Called medication-assisted treatment (MAT), three varieties have been approved by the FDA to combat opioid-use disorders (SAMHSA, 2020a). Methadone, first

[3] The N-SSATS only collected detailed staffing data in a special supplement in 2016.

approved by the FDA in 1947 and used for the treatment of addictions since the 1970s, is an opioid that is used to wean people off of stronger narcotics like heroin and prescription pain relievers (Retting and Yarmolinsky, 1995). Buprenorphine is similar to methadone in that it is an opioid that produces weaker euphoric effects than traditional narcotics and in that it helps users move away from more dangerous substances. It was approved by the FDA for the treatment of opioid-use disorder in late 2002 (SAMHSA, 2020a). Lastly, naltrexone was approved by the FDA for treatment of opioid dependence in 1984 (Krupitsky et al., 2010). Unlike the other two MATs, naltrexone is not an opioid substitute. Instead, it binds to opioid receptors and blocks the euphoric effects of the drugs. The effectiveness of these medications at combatting opioid addiction is empirically well-established (National Academies of Sciences, Engineering, and Medicine, 2019). However, despite these clinical benefits, MATs have not yet been fully embraced by SAT providers. As of 2018, only 44% of providers offered at least one of them as part of their service according to the N-SSATS. However, this represents a 123% increase over the offer rate in 2002.

2.3. DATA AND EMPIRICAL STRATEGY

2.3.1. Specialty SAT Facility Data

The N-SSATS is an annual survey of specialty SAT facilities that collects data on many of this study's outcomes. The sampling frame for the N-SSATS is called the Inventory of Behavioral Health Services (I-BHS), which includes a registry of the universe of SAT facilities known to the Substance Abuse and Mental Health Services Administration (SAMHSA)[4] . Data collection for the N-SSATS occurs between late March and early December each year, during which time

[4] SAMHSA defines a SAT facility as an entity, public or private, that provides substance abuse treatment.

surveys are sent to all facilities on the I-BHS. New facilities that are discovered by SAMHSA during this period (or reported by state substance abuse agencies) are also included in the N-SSATS and subsequently placed on the I-BHS for the following year. Facilities fill out a questionnaire of services they provide, counts of clients currently receiving treatment, and counts of beds dedicated to SAT clients (if applicable)[5]. Since the N-SSATS is voluntary for facilities, there is a certain amount of non-response each year. These response rates are recorded by state and year, and are nationally around 90% or above in each year during my sample period.

This study uses the N-SSATS for three categories of outcomes: facility counts, residential and hospital inpatient bed counts, and aggregate client counts. The N-SSATS collects data at the facility level on the types of treatment setting (inpatient, outpatient, both) and varieties of MAT for OUD they offer (methadone, buprenorphine, and/or naltrexone), as well as whether the facility is an officially licensed opioid treatment program (OTP) by SAMHSA. The N-SSATS also collects data on the number of beds designated for SAT patients in both residential and hospital inpatient facilities. These data were made publicly available as exact counts at the facility level until 2007, but in subsequent survey years the counts were censored into bins to protect facility privacy. However, by request to SAMHSA, I have obtained access to aggregate data containing exact counts of residential and hospital inpatient beds at the state-year level through 2017[6] . Lastly, I use the N-SSATS client counts by treatment setting (outpatient or residential/hospital inpatient), which I obtained in state-year aggregates from the N-SSATS Annual Reports (SAMHSA, 2020b)[7] . It is

---

[5] The reference date for residential and hospital inpatient beds and client counts was October 1 until the 2002 survey, when it changed to March 30/31. For outpatient clients, the reference period is/was the month before those dates.

[6] N-SSATS stopped collecting bed counts in even years after 2013, meaning that counts for 2014, 2016, and 2018 are unavailable for analysis. Additionally, some observations show implausibly large counts of beds for individual state-years. I have confirmed with SAMHSA that at least some of these temporary spikes can be attributed to individual facilities in these states, which suggests that the data were reported erroneously. I discuss my method for dealing with such outliers in Appendix A.

[7] As with the counts of inpatient beds, client counts ceased being collected during even years after 2013. The one exception to this is in 2016, when client counts were collected for that year alone. Inpatient client counts were also

important to note that I cannot disaggregate these client data by age or by substances used at admission (opioids, alcohol, etc.). I use these counts of facilities, beds, and clients to construct rates per 100,000 people. I also divide these rates by state-year facility response rates (described above) in order to better-account for non-response[8] . The time period I use for most of my analysis is 2000-2018, starting when the N-SSATS eligibility criteria were finalized[9].

Additionally, since it is possible that Medicare Part D increased the aggregate number of SAT clients, but not facilities or beds, I construct "utilization rates" of clients-per-facility and clients-per-bed to examine the extent to which this occurred. Higher utilization rates indicate that facilities became more crowded, on average[10]. I also construct a ratio of inpatient beds-per-facility to examine the degree to which increases in beds are being driven by new facility construction. For utilization rates of residential/hospital inpatient facilities and beds, I consider only facilities that report both client and bed counts in a given year, which I draw from SAMHSA's N-SSATS Annual Reports (SAMHSA, 2020b)[11]. However, since no such counts are available for outpatient facilities in the Annual Reports or using the publicly-available N-SSATS data, I divide the total reported outpatient clients by the number of outpatient facilities to create outpatient utilization rates.

---

subject to apparent misreporting in certain state-years, similar to the misreporting for bed counts (as discussed in footnote 6). I discuss my method for dealing with such outliers in Appendix A.

[8] Separate response rates are not available by type of facility. However, this will not introduce bias as long as the difference between the pooled facility response rate and the response rates by facility type are uncorrelated with the treatment variable. Additionally, I show that there is no significant correlation between Medicare Part D and the pooled response rate in Appendix Table A.2.

[9] I explain the evolution of the N-SSATS eligibility criteria in further detail in Appendix B.

[10] Note that it is common for clients-per-bed to be greater than 100% because clients can occupy beds that are not designated for SAT. For example, in 2017 about 10% of residential and 20% of hospital inpatient facilities had utilization rates greater than 100% (SAMHSA, 2020b).

[11] After 2012, these counts are only reported every odd year. The 2013 data are excluded because the report censors several state cells for hospital inpatients due to small sample sizes.

2.3.2. Detailed SAT Admission, Medicare Part D Enrollment, and Prescription Drug Distribution Data

The Treatment Episode Data Set (TEDS), also conducted by SAMHSA, provides detailed information on individual admissions to all specialized SAT facilities that receive public funding (SAMHSA, 2020c). While this only represents a fraction of the facilities reporting to the N-SSATS, the TEDS provides more detailed information about individual admissions such the individual's age and substances used. I categorize a patient as being admitted for prescription opioid abuse if they list any of the substances used as “non-prescription methadone” or “other opiates and synthetics,” the latter of which includes prescription opioids. Additionally, I focus on admissions of patients aged 12-54, which is the population that was induced into SAT by Medicare Part D (Powell et al., 2020). I obtain data on Medicare Part D enrollment from the Center for Medicare & Medicaid Services (CMS) Statistical Supplements (years 2006-12) and CMS Program Statistics (2013-18). Enrollment counts are aggregated to state-year levels.

Lastly, I obtain data on prescription opioid distributions from US Drug Enforcement Administration (DEA) Automated Reports and Consolidated Ordering System (ARCOS), years 2000-2017. As mandated by the Controlled Substance Act of 1970, these data track quantities of Schedule II, and select Schedule III and IV, substances as they are supplied from manufacturers to retail distributors. I follow Powell et al. (2020) in defining a measure of prescription opioid supply distributed to a state in a given year that consists of the total morphine equivalent doses (MEDs) of the following medications: fentanyl, hydrocodone, hydromorphone, meperidine, methadone, morphine, oxycodone, codeine, dihydrocodeine, levorphanol, oxymorphone, and tapentadol[12].

[12] MED conversion executed according to the conversion factors published by CMS here: https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Opioid-Morphine-EQ-Conversion-Factors-March-2015.pdf

2.3.3. Data on Buprenorphine Licensing

Although the N-SSATS documents whether or not specialty SAT facilities administer MAT, it does not account for the many providers who operate out of non-specialty settings. In order to determine if Medicare Part D had an effect on the provision of buprenorphine across all treatment settings, I use data on the granting of Drug Addiction Treatment Act of 2000 (DATA 2000) waivers by SAMHSA, which first began in 2002. These waivers are given to doctors, nurse practitioners, or physician's assistants who meet certain qualifications to enable them to treat opioid patients with buprenorphine. Each waiver specifies the number of patients (30, 100, or 275) that the practitioner is allowed to treat at once. Appendix Figure A.2 displays the cumulative number of waivers and DATA 2000 patient capacity granted by year since 2002. Since SAMHSA does not record the exit of providers who had formerly been granted DATA 2000 waivers, I measure the stock of DATA 2000 waivers capacity in two ways: based on the number of waivers ever granted and based on the number of waivers granted to providers still practicing in 2020. The former of these measurements is somewhat of an overestimate of capacity (especially later in the sample period) and the latter is an underestimate (especially earlier in the sample period).

2.3.4. Population and Control Variables

I use state-by-year population data from the Surveillance, Epidemiology, and End Results Program (SEER) for weighting regressions and two control variables (natural log of the population and fraction of the population “white”) (SEER, 2019). Additionally, I control for a number of time-varying state-level policies which have been shown to have affected access to SAT and/or provider behavior (MacLean et al., 2018; Maclean and Saloner, 2019; Meinhofer and Witman, 2018; Wen et al., 2017). Specifically, these are the Affordable Care Act (ACA) Medicaid expansions, Health Insurance Flexibility and Accountability (HIFA) waivers, and State SAT Parity Laws[13]. I also control for the implementation of Pain Clinic Laws, Must-Access Prescription Drug Monitoring Programs (PDMPs), and Medical Marijuana Laws, all of which are associated with reductions in opioid prescribing (Rutkow et al., 2015; Powell et al., 2018, Sacks et al., 2021). Lastly, I control for state unemployment rates over time using data from the Bureau of Labor Statistics.

2.3.5. Empirical Strategy

The main challenge in estimating the impact of a national policy like Medicare Part D is assigning treatment and control groups. I follow the approach utilized in Powell et al. (2020), which exploits the fact that after the program was implemented, states with higher percentages of the population eligible for Medicare also had higher percentages of people eligible for Medicare Part D. This research design amounts to a differences-in-differences approach, where the first "difference" compares states before/after 2006, and the second "difference" compares states with varying fractions of the population aged 65+[14]. I fix the cross-sectional variation in the fraction of population aged the 65+ to its 2003 values, the year Medicare Part D was passed, in order to avoid incorporating systematic migration resulting from the policy (Figure 1 summarizes this variation across states). Thus, the treatment effect is identified by comparing outcomes of interest across states by the share of their 2003 population aged 65+ (*elderly share)*, before and after 2006.

Below is the specification for the main differences-in-differences design:

$$y_{st} = \alpha + \gamma\left[\%aged \geq 65_{s,2003} \times Post_t\right] + X'_{st}\beta + \delta_s + \eta_t + \varepsilon_{st} \tag{1}$$

[13] I use the same implementation dates as are listed in the above studies.

[14] This approach does not take into account people qualified for Medicare via SSDI who would have gained access to Medicare Part D. However, many who had received SSDI for 24 consecutive months were also eligible for Medicaid, which has its own prescription drug benefit.

The variable y_{st} stands in for the outcomes described above, all of which vary by state (s) and year (t). Regressions using admission, client, facility, or bed counts per 100,000 population are weighted using state population. Regressions using outcomes such as clients-per-bed, clients-per-facility, or beds-per-facility, are weighted by the outcome's denominator (e.g., regressions using beds-per-facility are weighted by state facility count). The treatment variable is as described in the previous paragraph, where $Post_t$ equals 1 in the years 2006+ and 0 otherwise. X'_{st} is a vector of control variables, added in alternate specifications. This may include a variety of policy indicators (ACA Medicaid Expansion, HIFA Waivers, State SAT Parity Laws, Strong PDMPs, Pain Clinic Laws, and Medical Marijuana Laws) and other time-varying controls (state unemployment rate, the natural log of the population, and fraction of population "white")[15]. I also include state and year fixed effects which control for the independent effects of the fixed share of population aged 65+ and national effect of Medicare Part D, respectively. Standard errors are clustered at the state level. Given the regional differences in elderly share apparent in Figure 1 (e.g., older populations in the east and younger populations in the west), one may be concerned that the variable merely proxies for regional differences in outcomes trends. In order to address these concerns, I estimate alternate specifications in which I include region-by-year fixed effects in the regression. In these specifications, γ is identified only from variation driven by intra-regional differences in elderly share across states and the pre/post 2006 temporal variation. Since several of the outcomes I study are count variables (e.g., clients, facilities, and beds), I also estimate Poisson models in addition to the linear models[16].

[15] Since these controls vary by time, I want to confirm that none of them are related to the treatment variable. Additionally, since I divide all of my N-SSATS outcomes by the survey response rate, I also need to verify that the response rate is not related to treatment either. I show the results of estimating equation (1) (excluding $X'_{st}\beta$ from all specifications) using the control variables and response rate as the outcomes in Appendix Table A.2.

[16] Poisson models assume effects that are in proportion to the outcome means, which is a plausible alternative to the linearity assumption in linear regression.

Additionally, I use an event study version of equation (1) in which the effect of $\%aged \geq 65_{s,2003}$ is interacted with year fixed effects. This model is specified in equation (2), where $T = \{t_0, \ldots, 2018\} \backslash 2005$ and t_0 is either 2000 and 2002, depending on the outcome. I omit the year before Part D is implemented, 2005, as the baseline period.

$$y_{st} = \alpha + \sum_{t \in T} \gamma_t \left[\%aged \geq 65_{s,2003} \times \eta_t\right] + X'_{st}\beta + \delta_s + \eta_t + \varepsilon_{st} \qquad (2)$$

2.4. RESULTS

First, I provide evidence that Medicare Part D constituted an exogenous shock to aggregate demand for SAT through its effect on prescription opioid abuse. Then, I demonstrate the subsequent effects of the policy on the availability of SAT. Last, I provide instrumental variable estimates that contextualize the main results on the supply of SAT in terms of total prescription opioids distributed.

2.4.1. Demand for Substance Abuse Treatment

I replicate three key findings from Powell et al. (2020) and display the results in Figure 2. Panel A shows estimates from yearly cross-sectional regressions of elderly share of the percent of state population enrolled in Part D on elderly share. These coefficients show that, after Part D became effective in 2006, an additional percentage point of elderly share is associated with an approximately 0.4 percentage point increase in Part D enrollment. This effect grew over time through 2018. Panel B shows that, beginning in 2006, increases in elderly share were associated with an increase in prescription opioid distributions to states. This effect grew through 2010 and then began to decline, a shift which coincides temporally with the decrease in prescription opioid abuse observed after the Oxycontin reformulation (Alpert et al., 2018). Panel C displays the effect of Part D on prescription opioid SAT admissions per 100,000 population. Similar to the trends in opioid distribution, after 2006 a 1% increase in elderly share caused an increase in admissions by about 20 per 100,000 population at its peak in 2011. After 2011, this differential begins to fall back to 2005 levels, which coincides with the reformulation as well. The parallel pre-treatment trends in opioid admissions through 2005 indicate that states with varying levels of elderly shares were experiencing similar trajectories of opioid abuse before Part D began.

Next, I demonstrate that Medicare Part D's effect on prescription opioid admissions increased SAT clients in aggregate rather than just crowding out other types of admissions. Figure 3 displays event study coefficients from estimating equation (2) using outpatient and residential/hospital inpatient aggregate client counts per 100,000 population as the outcomes. Panels A and B demonstrated that, starting in 2006, a percentage point increase in elderly share is associated with increases in both categories of SAT clients after several years of parallel pre-treatment trends. Summing the treatment effects across event studies indicates increases in aggregate clients similar to the effects on prescription opioid admissions from Figure 2, Panel C, through 2011. However, unlike the estimates for only prescription opioid admissions, the relative increases in aggregate clients do not decline after 2011. These effects are summarized via differences-in-differences models by estimating equation (1) with client counts as the outcomes with results displayed in Table 1. The preferred specification in column (3) indicates that a percentage point increase in elderly share is associated with a 2.5% increase in residential and hospital inpatient clients and a 3.6% increase in outpatient clients relative to the 2000-2005 pre-treatment means. These estimates are robust to excluding non-policy controls in column (2) and the exclusion of all controls in column (1). Additionally, the results are robust to inclusion of region-by-year fixed effects and use of Poisson model in columns (4) and (5), respectively.

2.4.2. Residential/Hospital Inpatient and Outpatient SAT Availability

In this section, I show how Medicare Part D affected the availability of SAT facilities and beds. Since client capacity at residential and hospital inpatient facilities is constrained by the number of open beds, as unlike outpatient facilities, I choose to split my analysis along these lines to allow for differing effects of Part D. I estimate event studies for residential and hospital inpatient facilities and beds per 100,000 population and display the results in Figure 4, Panels A and B. Both estimates show insignificant pre-period coefficients before Part D is introduced, after which states with higher shares of the population aged 65+ experience increases in facilities and beds relative to states with lower shares. Similar to the results for client counts, and contrasting with the results for opioid admissions, these effects on treatment availability do not contract again after 2011 and actually continue to increase slightly through 2018. I summarize these event study results by estimating equation (1) with facility and beds counts per 100,000 population as the outcomes. The results are displayed in Table 2, Panels A and B. The preferred specification (column (3)) indicates that an additional percentage point of elderly share is associated with a 2.3% increase in residential and hospital inpatient facilities per 100,000 population and a 2.0% increase in beds per 100,000 population, relative to the 2000-2005 pre-treatment means. The results are also robust to exclusion of control variables, inclusion of region-by-year fixed effects, and use of Poisson models.

I also test whether, on average, the effects on bed capacity were driven by facilities altering their bed stocks or if they were entirely driven by changes in the number of facilities. I do this by estimating equation (1) with the outcome being the state-level ratio of beds over residential and hospital inpatient facilities. Since a certain number of residential and hospital inpatient facilities fail to report bed and/or client data, these models are estimated on a sample of only facilities that report both pieces of information in a year. These estimates are displayed in Table 2, Panel C. Across all specifications, there are no statistically significant effects of elderly share on beds per facility. Looking at the preferred specification in column (3), the reported confidence interval allows me to rule out effects as small as an increase of 0.23 beds per facility at the 95% level. Exclusion of controls or inclusion of region-by-year fixed effects alter this estimate very little. Since the point estimate in Panel B, column (3) is 0.877, this suggests that the majority of Part D's effect on the number of residential and hospital inpatient beds is due to an increase in the number facilities, as opposed to changing the number of beds per facility. I examine whether this increase in the availability of residential and hospital inpatient SAT services was commensurate with the increase in clients documented above. In order to determine whether facilities became more crowded with residential and hospital inpatient clients due to Part D, I estimate equation (1) with utilization rates for facilities and beds as the outcomes and display the results in Appendix Table A.3, Panel A. These estimates suggest any increase in utilization rates that may have occurred were small and the confidence interval in column (5) rules out effects on clients-per-bed greater than 1.7% of the pre-treatment mean at the 95% level.

Next, I examine the impact of Medicare Part D on outpatient treatment facilities. Since outpatient clients do not take up beds, the only measure of outpatient treatment capacity available is the number of outpatient facilities per 100,000 population. I estimate equation (2) using this outcome and display the lead and lag coefficients in Figure 5. None of the pre-period coefficients are significantly different from zero, though there is a slight downward trend leading up to implementation 2006. After Part D began, states with higher elderly shares experienced growth in the number of outpatient facilities over time. However, many of these coefficients are imprecisely estimated with large confidence intervals. I estimate equation (1) for outpatient facilities and display the difference-in-difference estimates in Table 3. The treatment effect appears to be statistically insignificant using the linear models shown in columns (1)-(4) and only significant at the 10% level when using a Poisson model in column (5). However, although the point estimates for outpatient facilities are all statistically insignificant at the 95% level, the Poisson estimates are not significantly different than the Poisson estimates for residential and hospital inpatient facilities at even the 90% level[17]. Therefore, I cannot formally rule out the null hypothesis that the effects of Part D on residential/hospital inpatient and outpatient facilities are the same. Appendix Table A.3, Panel B shows the results from estimating equation (1) using outpatient utilization rates as the outcomes. These results suggest that Part D may have increased outpatient clients per facility, although estimates are sensitive to the specification.

These effects on the supply of residential/hospital inpatient and outpatient SAT may explain some of the persistence in the effect of Part D on aggregate client counts as displayed in Figure 3. Previous evidence indicates that the opening of new SAT facilities in a given area results in a persistent increase in local treatment admissions (Swenson, 2015). Even if the effect on prescription opioid admissions began to decrease after 2011, the newly available treatment facilities resulting from Part D could have still been used by people with other conditions.

2.4.3. Medication-Assisted Treatment Availability

[17] I compare effect sizes using Poisson models since the estimates for γ can be interpreted as percent increases relative to their respective means.

Section 2.4.1 of this paper shows that Medicare Part D increased demand for SAT for OUD. Since the results in the previous section suggest that this increase in demand caused an expansion in residential and hospital inpatient SAT facilities and beds, it stands to reason that it may have also affected the number of providers offering MAT. To test this hypothesis, I analyze whether the introduction of Medicare Part D increased the number of specialty facilities offering at least one form of MAT for OUD. I begin the analysis period in 2002 since the N-SSATS did not track naltrexone offerings until that date. Additionally, bear in mind that since buprenorphine was not approved by the FDA to treat opioid-use disorder until late 2002, its offering was not tracked by the N-SSATS until 2003. Figure 6 contains the event study results from estimating equation (2) with the outcome being the number of facilities per 100,000 population that offer at least one form of MAT. Between 2002 and 2005, outcomes trended similarly across states with varying elderly shares. Then, starting immediately in 2006, states with higher elderly shares experienced increases in facilities with MAT compared to states with lower elderly shares. This difference increased gradually throughout the sample period to its peak in 2018.

Next, I estimate differences-in-differences models using equation (1) to summarize the relationship between elderly share and facilities with MAT for OUD, and display the results in Table 4, Panel A. I also estimate additional models examining whether Part D affected the number of facilities offering multiple forms of MAT, zero forms of MAT, or an OTP. Panel A, columns (1)-(3) indicate that elderly share is significantly related to the number facilities that provide MAT for OUD after 2006. These results are robust to the addition of controls and region-by-year fixed effects. According to the preferred specification in column (2), a percentage point increase in elderly share is associated with a 6.9% increase in facilities that offer MAT relative to the pre-treatment mean. Additionally, there are no statistically significant effects on facilities that offer no forms of MAT, or on facilities that offer at least two forms. There is also no significant impact on the number of facilities with accredited opioid treatment programs (OTPs). These results suggest that the introduction of Medicare Part D, and the resulting shock to demand for opioid addiction treatment, played a role in moving the marginal facility to adopt MAT (either within existing facilities or within newly opening ones). Table 5 indicates the specific types of MATs that are driving this result by estimating equation (1) with facilities offering naltrexone, buprenorphine, and/or methadone per 100,000 population as the outcomes. Panel B, Columns (1)-(3) indicate that the effects on facilities offering MAT are nearly entirely driven by facilities newly offering naltrexone. These estimated treatment effects are very similar to those for facilities with MAT in Panel A. On the other hand, it appears that Part D did not have a significant effect on facilities offering buprenorphine or facilities offering methadone. This result is of note because naltrexone has been shown to be highly effective at improving treatment outcomes for opioid users[18].

As an alternative measure of buprenorphine availability, I also estimate the effects of Part D on state-level DATA 2000 buprenorphine patient capacity. Appendix Figure A.3 shows the event study results from estimating equation (2) using as the outcome the total number of buprenorphine clients per 100,000 population ever granted under DATA 2000 by state. The sample period begins in 2002, which is the first year that SAMHSA began granting DATA 2000 waivers. Through 2005, states with high and low elderly shares trended closely. Starting in 2006, states with higher elderly shares began to see increases in capacity relative states with lower elderly shares, though these differences remained statistically insignificant throughout much of the post-treatment period. Starting in 2016, when SAMHSA began granting waivers that permitted a patient

[18] Evidence from clinical trials of extended release Naltrxone have been shown that application during treatment of OUD can increase treatment retention by 75%-78% and reduce relapse rates 94% (Comer et al., 2006; Krupitsky et al., 2011).

capacity of 250, the effect of having a higher elderly share become much more pronounced. However, the results from the differences-in-differences models displayed in Appendix Table A.4, column (4) indicate that these results are not robust to controlling for region-by-year trends in DATA 2000 client waivers. Estimating equation (1) using clients allowable under DATA 2000 as the outcome, but restricted to professionals who are still practicing in 2020, returns smaller coefficients that are less statistically significant. In sum, there is some evidence that Medicare Part D may have increased the number of buprenorphine clients allowable under DATA 2000, but results are sensitive to the specification.

2.4.4. Effects on Facilities by Ownership Status

Previous work has documented that the effects of health policies on SAT providers can vary by ownership status (i.e., public, non-profit, and for-profit) (Hamersma and Maclean, 2021; Maclean et al., 2018). Therefore, there is also reason to believe that Medicare Part D and its subsequent effect on opioid addiction could have had varying effects on the supply of SAT according to the ownership statuses of facilities. This analysis will shed light on the degree to which the public versus the private sector is driving the observed effects on providers. Table 6 displays the results of estimating equation (1) using facilities per 100,000 population by treatment environment and ownership status. Panel A shows that only publicly-owned residential and hospital inpatient facilities were significantly increased by Part D, although the coefficients for each type of privately-owned facility are positive and relatively large. Panel B shows a similar pattern for outpatient facilities. Panel C examines Part D's effects on facilities offering MAT by ownership type and shows that the vast majority of the effect is concentrated among non-profit facilities. Estimates suggest that a percentage point increase in elderly share after 2006 is associated with 0.051 more non-profit private facilities offering MAT per 100,000 population. In sum, these results suggest that both private and public facilities increase treatment capacity in response to demand. However, it is difficult to specify precisely the mechanism by which these increases work. For example, the significant increase in private non-profit facilities offering MAT could be due to market-based forces or due to mediating government policies which have encouraged MAT adoption.

2.4.5. Scaling the Effects of Part D by its Impact on Prescription Opioid Distribution

I argue that Part D increased the supply of SAT by expanding the availability of prescription opioids. This section uses a two-stage least squares (2SLS) approach to rescale the effect of Part D on SAT provides to be in terms of quantity of medication distributed. Table 7, columns (1) and (3) show the results of "naive" regressions of the outcomes of interest on MEDs of prescription opioids distributed per capita. Columns (2) and (4) show the 2SLS results from instrumenting for prescription opioid distribution with Medicare Part D. Translating the 2SLS results into percentage terms: a 10% increase in MEDs per capita is associated with increases in residential/inpatient facilities per 100,000 population by 2.6%, in beds per 100,000 by 2.3%, in outpatient facilities per 100,000 by 1.6% (at 10% significance), and in facilities with MAT per 100,000 by 8.6%, all relative to the pre-treatment means[19]. All of these effects are larger than the "naive" OLS estimates in columns (1) and (3). This disparity either suggests that there are downward biases inherent to the OLS models or that there are alternative channels through which Part D positively affects the availability of SAT.

---

[19] Table 7 also shows the F-statistics on elderly share for the corresponding first stage regressions for each 2SLS estimate.

2.4.6. Robustness Checks

2.4.6.1. Leave-One-Out Tests

The research design used in this paper weighs a state's contribution to the treatment effect as being proportionate to its elderly share. Therefore, states with particularly large or small values of elderly share have the potential to drive a substantial portion of the treatment effects. I test directly for this possibility by re-estimating equation (1) with the main outcomes of interest while systematically removing one state at a time. I display point estimates and 95% confidence intervals from these regressions in Appendix Figure A.4 for the following outcomes: residential and hospital inpatient facilities, beds, outpatient facilities, and facilities with at least one form of MAT for opioid-use disorder, per 100,000 population. The only states whose exclusions meaningfully impact the point estimates are Florida and Texas[20]. Furthermore, in only one instance among the former three outcomes (residential and hospital inpatient beds) is there a case in which the exclusion of a state (Texas) from the sample renders the treatment effect statistically insignificant at the 5% level. In Panels A and D, on the other hand, estimates are robust to the removal of any one state. Estimates in Panel C are statistically insignificant at the 5% level in almost all cases.

2.4.6.2. Are These Effects Driven by Other Medicare Policies?

Medicare Part D reduces the cost of obtaining prescription opioids. However, Medicare Part D can also pay for certain prescription MATs (HHS, 2016)[21]. Therefore, one concern may be that the

[20] Possible reasons for this include the relatively extreme levels of elderly share in Florida (high) and Texas (high) and their large populations. Indeed, re-estimating the leave-one-out tests without population weights mutes the impact of excluding either of these states. Furthermore, this phenomenon is also present in the mortality estimates in Powell et al. (2020) to a milder degree.

[21] These include some forms of buprenorphine and naltrexone, though not methadone when prescribed as a MAT (Congressional Research Service, 2020).

effects of Part D on the availability of SAT are not driven by the policy's effect on prescription opioids, but instead by the effects of other Part D provisions. I show that this mechanism is unlikely. Appendix Table A.5 contains the results of estimating the effect of Part D on the number of Medicare admissions and facilities that accept Medicare per 100,000 population. Column (1), which uses as the outcome the number of SAT admissions per 100,000 that report having non-Medicaid public insurance (i.e., Medicare, CHAMPUS, etc.), shows an insignificant effect of elderly share[22]. Columns (2)-(4) show the effects of Part D on the number of various types of facilities that accept Medicare. None of these effects are statistically significant, and the point estimates are all smaller than the effects on each type of facility as estimated in sections 2.4.2 and 2.4.3. Furthermore, since Medicare acceptance rates are substantial and remain relatively unchanged across the sample period (about 37% of all facilities in both 2000 and 2018), low participation rates are not inducing power issues in these analyses. These results indicate that other Medicare Part D provisions are not likely to be driving the main results.

2.4.6.3. Alternative Data: County Business Patterns

An alternative source of data on SAT facilities is the County Business Patterns (CBP). These data, collected by the United States Census Bureau, contain yearly information on establishments extracted from their Business Register. The Business Register, in turn, is a database of all single and multi-establishment employer companies that are known to the Census Bureau. The strength of the CBP for this project's purposes, when compared to the N-SSATS, is that they should contain a more complete count of private substance abuse treatment establishments since they are not similarly subject to non-response from participants. However, the CBP also comes with at least

[22] The reduced number of observations is due to certain state-years not reporting health insurance status of clients.

two drawbacks. The main drawback is that they exclude most establishments reporting government employees. This is particularly challenging for my analysis, since my estimates by ownership status indicate that a large share of the effects on residential/hospital inpatient and outpatient facilities are driven by publicly-owned facilities. The second drawback is that the North American Industry Classification System (NAICS) codes used to identify both outpatient SAT facilities (621420) residential SAT facilities (623220) group together substance abuse and mental health facilities. Although many specialty mental health facilities also run programs for substance-use disorder, as of 2017 about 44% of mental health facilities did not offer substance abuse treatment (SAMHSA, 2018a). Furthermore, SAMHSA estimates the existence of approximately 15,000 SAT facilities and 13,000 mental health facilities that same year. This implies that the share of mental health facilities without SAT within each NAICS code is non-trivial (SAMHSA, 2018b). Both of these drawbacks suggest that estimates using the CBP may differ meaningfully from estimates using the N-SSATS.

Appendix Figure A.5, Panels A and B display the event study coefficients from estimating equation (2) with outpatient and residential/hospital inpatient facilities as the outcomes, respectively. These graphs display patterns that are qualitatively similar to the estimates using the N-SSATS but the confidence intervals tend to be wider. Furthermore, each graph displays the treatment effects estimated using equation (1) along with the CBP. Although point estimates are similar in magnitude to the N-SSATS estimates, they are statistically insignificant at conventional levels.

2.5. CONCLUSION AND DISCUSSION

Despite the best efforts of policy makers and public health initiatives, opioid mortality reached an all-time high in 2020. Since the number of people suffering from opioid use disorders is increasing every day, policymakers need to be able to assess how well the U.S. healthcare system has fared at expanding treatment supply to meet demand. In order to estimate this directly, I exploit the introduction of Medicare Part D as an exogenous increase in the availability of prescription of opioids and addiction. Using the N-SSATS, I find that a 10% increase in Morphine Equivalent Doses, induced by Part D, resulted in a 2.6% increase in the number of residential/inpatient SAT facilities and a 2.3% increase in the number of beds. Furthermore, I find that a 10% increase in MEDs resulted in an 8.4% increase in the number of SAT facilities offering at least one form of MAT.

The effects I estimate incorporate not only the profit-driven responses to demand of private sector actors, but also the mediating effects of any government policy implemented in response to the increase in opioid addiction. Therefore, my results can be viewed as the capability of SAT providers (which includes public sector actors) to respond to population-wide changes in opioid addiction rates. These results are encouraging in that they show that the supply of SAT expanded to meet the increase in need. Furthermore, this effect is particularly pronounced for facilities offering MAT, which is especially important for treating opioid-used disorder. I also find that the increases in residential/hospital inpatient facilities, and possibly outpatient facilities, were driven in large part by increases public facilities. This implies that the government has played a significant role in addressing the shock to addiction induced by Medicare Part D. Therefore, policymakers may consider continuing and expanding the policies that have supported the provision of SAT over the past 15 years in order to combat the new wave of the opioid addiction in the post-COVID era.

2.6. REFERENCES

Alpert, Abby, Darius Lakdawalla, and Neeraj Sood (2015). Prescription drug advertising and drug utilization: The role of Medicare Part D (No. w21714). National Bureau of Economic Research.

Alpert, Abby (2016). The anticipatory effects of Medicare Part D on drug utilization. *Journal of Health Economics, 49,* 28-45.

Alpert, Abby, David Powell, and Rosalie L. Pacula (2018). Supply-side drug policy in the presence of substitutes: Evidence from the introduction of abuse-deterrent opioids. A*merican Economic Journal: Economic Policy*, 10 (4): 1-35.

Bärnighausen, Till and David E. Bloom (2009). Financial incentives for return of service in underserved areas: a systematic review. *BMC Health Services Research* 9, 86.

Blume-Kohout, Margaret E. and Neeraj Sood (2013). Market size and innovation: Effects of medicare part D on pharmaceutical research and development. *Journal of Public Economics*, 97: 327-336.

Caulkins, J. P., Anna Kasunic, and Michael A. C. Lee (2014). Societal burden of substance abuse. *International Public Health Journal*, 6 (3): 269-282.

Comer, Sandra D., Maria A. Sullivan, Elmer Yu, Jami L. Rothenberg, Herbert D. Kleber, Kyle Kampman, Charles Dackis, and Charles P. O'Brien (2006). Injectable, sustained-release naltrexone for the treatment of opioid dependence: a randomized, placebo-controlled trial. *Archives of General Psychiatry*, 63(2): 210-218.

Congressional Research Service (2020). Medicare Coverage of Medication Assisted Treatment (MAT) for Opioid addiction. Retrieved from: https://crsreports.congress.gov/product/pdf/IF/IF10875.

Cubanski, Juliette, Anthony Damico, and Tricia Neuman (2019). *10 Things to know about Medicare Part D coverage and costs in 2019*. Kaiser Family Foundation Data Note. Retrieved from: http://files.kff.org/attachment/Data-Note-10-Things-to-Know-about-Medicare-Part-D-Coverage-and-Costs-in-2019.

Dick, Andrew W., Rosalie L. Pacula, Adam J. Gordon, Mark Sorbero, Rachel M. Burns, Douglas L. Leslie, and Bradley D. Stein (2015). Increasing Potential Access to Opioid Agonist Treatment in U.S. Treatment Shortage Areas. *Health Affairs (Millwood),* 34 (6): 1028-1034.

Duggan, Mark and Fiona S. Morton (2010). The effect of medicare part D on pharmaceutical prices and utilization. *American Economics Review*, 100 (1): 590-607.

Dranove, David, Craig Garthwaite, and Manuel I. Hermosilla (2020). Expected profits and the scientific novelty of innovation. NBER Working Paper 27093.

Dranove, David, Craig Garthwaite, and Manuel I. Hermosilla (2014). Pharmaceutical Profits and the Social Value of Innovation. NBER Work Paper 20212.

Gowrisankaran, Gautam and Robert J. Town (1997). Dynamic equilibrium in the hospital industry. *Journal of Economics and Management Strategy*, 6 (1): 45-74.

Hu, Tianyan, Sandra L. Decker, and Shin-Yi Chou (2017). The impact of health insurance expansion on physician treatment choice: Medicare Part D and physician prescribing. *International Journal of Health Economics and Management*, 17: 333-358.

Jones, Christopher M., Melinda Campopiano, Grant Baldwin, and Elinore McCance-Katz (2015). National and State Treatment Need and Capacity for Opioid Agonist Medication-Assisted Treatment. *American Journal of Public Health*, 105 (8): e55-e63.

Ketcham, Jonathan D. and Kosali I. Simon (2008). Medicare Part D's effects on elderly drug costs and utilization. NBER Working Paper 14326.

Krupitsky, Evgeny, Edwin Zvartau, and George Woody (2010). Use of Naltrexone to treat opioid addiction in a country in which Methadone and Buprenorphine are not available. *Current Psychiatry Report*, 12 (5): 448-53.

Krupitsky, Evgeny, Edward V. Nunes, Walter Ling, Ari Illeperuma, David R. Gastfriend, and Bernard L. Silverman (2011). Injectable extended-release naltrexone for opioid dependence: a double-blind, placebo-controlled, multicentre randomised trial. *The Lancet*, 377 (9776): 1506-1513.

Lichtenberg, Frank R. and Shawn X. Sun (2007). The impact of Medicare Part D on prescription drug use by the elderly. *Health Affairs*, 26 (6): 1735-1744.

Maclean, Johanna C., Ioana Popovici, and Elisheva R. Stern (2018). Health insurance expansions and providers' behavior: Evidence from substance-use-disorder treatment providers. *Journal of Law and Economics*, 61 (2): 279-310.

Maclean, Johanna C., & Saloner, Brendan (2018). Substance use treatment provider behavior and healthcare reform: Evidence from Massachusetts. *Health Economics*, 27: 76-101.

National Academies of Sciences, Engineering, and Medicine. (2019). *Medications for opioid use disorder save lives*. Washington, DC: The National Academies Press.

National Institutes of Health, National Institute on Drug Abuse (NIDA). (2018). *Principles of drug addiction treatment: A research-based guide* (3rd ed.). Retrieved from: https://www.drugabuse.gov/publications/principles-drug-addiction-treatment-research-based-guide-third-edition/preface

Powell, David, Rosalie L. Pacula, and Mireille Jacobson (2018). Do medical marijuana laws reduce addictions and deaths related to pain killers?. *Journal of Health Economic*, 58: 29-42.

Powell, D., Rosalie L. Pacula, and Erin Taylor (2020). How increasing medical access to opioids contributes to the opioid epidemic: Evidence from Medicare Part D. *Journal of Health Economics*, 71: 102286.

Sacks, Daniel W., Alex Hollingsworth, Thuy Nguyen, and Kosali Simon (2021). Can policy affect initiation of addictive substance use? Evidence from opioid prescribing. *Journal of Health Economics*, 76: 102397.

Safran, Dana G., Patricia Neuman, Cathy Schoen, Michelle S. Kitchman, Ira B. Wilson, Barbara Cooper, Angela Li, Hong Chang, and William H. Rogers (2005). Prescription drug coverage and seniors: Findings from a 2003 national survey. *Health Affairs, Supplemental Web Exclusives*, W5-152-W155-166.

Substance Abuse and Mental Health Services Administration (SAMHSA). (1997). A guide to substance abuse services for primary care clinicians. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2018a). *National Mental Health Services Survey (N-MHSS): 2000-2017*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2018b). *National Mental Health Services Survey (N-MHSS): 2017. Data on Mental Health Treatment Facilities*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2020a). Medication and counseling treatment. Retrieved from: https://www.samhsa.gov/medication-assisted-treatment/treatment#medications-used-in-mat.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2020b). *National Survey of Substance Abuse Treatment Services (N-SSATS): 2000-2018. Data on Substance Abuse Treatment Facilities*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Substance Abuse and Mental Health Services Administration (SAMHSA). (2020c). *Treatment Episode Data Set (TEDS): 2000-2017*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Surveillance, Epidemiology, and End Results (SEER). (2019). *Program Populations (2000-2018)*. National Cancer Institute, DCCPS, Surveillance Research Program. Retrieved from: www.seer.cancer.gov/popdata

Swensen, Isaac D. (2015). Substance-abuse treatment and mortality. *Journal of Public Economics*, 122: 13-30.

U.S. Department of Health and Human Services (HHS). (2006). Over 38 million people with Medicare now receiving prescription drug coverage. News Release, June 14, 2006. Retrieved from: http://www.hhs.gov/news/press/2006pres/20060614.html.

U.S. Department of Health and Human Services (HHS). (2016). Medicare Coverage of Substance Abuse Services. *MLN Matters*, SE1604.

Wen, Hefei, Jason M. Hockenberry, and Janet R. Cummings (2017). The effect of Medicaid expansion on crime reduction: Evidence from HIFA-waiver expansions. *Journal of Public Economics*, 154: 67-94.

Yin, Wesley, Anirban Basu, James X. Zhang, Atonu Rabbani, David O. Meltzer, and G. Caleb Alexander (2008). The effects of Medicare Part D prescription benefit on drug utilization and expenditures. *Annals of Internal Medicine*, 148 (3): 1-14.

Zhang, Y., Julie M. Donohue, Judith R. Lave, Gerald O'Donnell, and Joseph P. Newhouse (2009). The effect of medicare part D on drug and medical spending. *New England Journal of Medicine*, 361: 52-61.

Figure 1: Fraction of Population Aged 65+ in 2003 (Elderly Share), by State

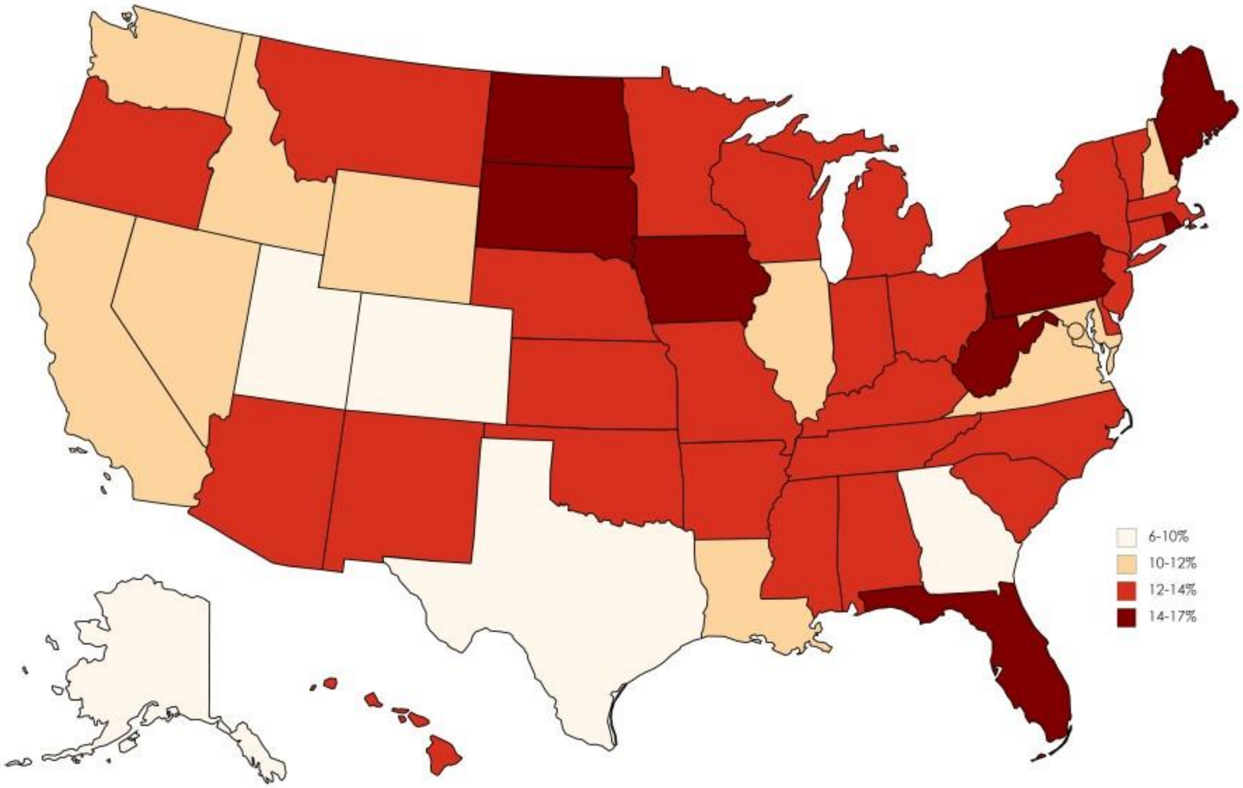
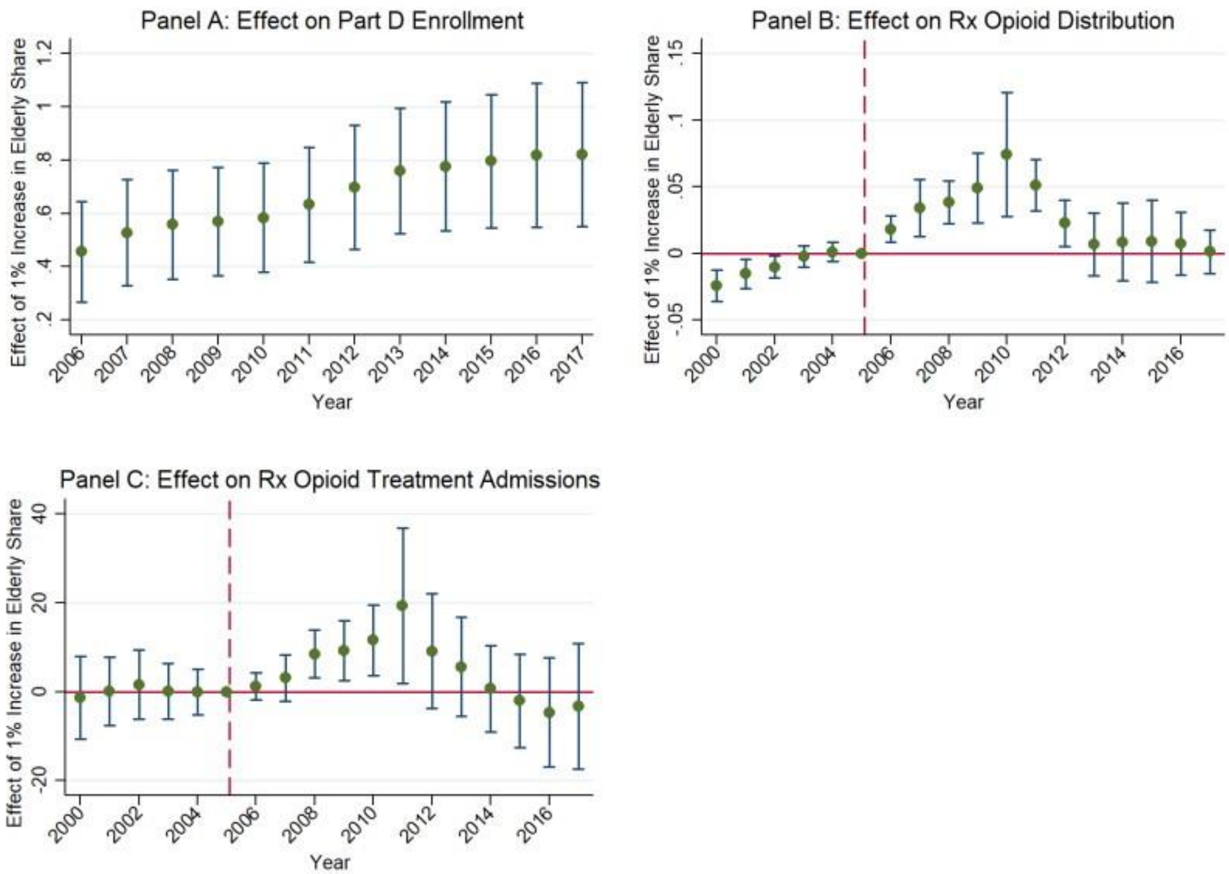
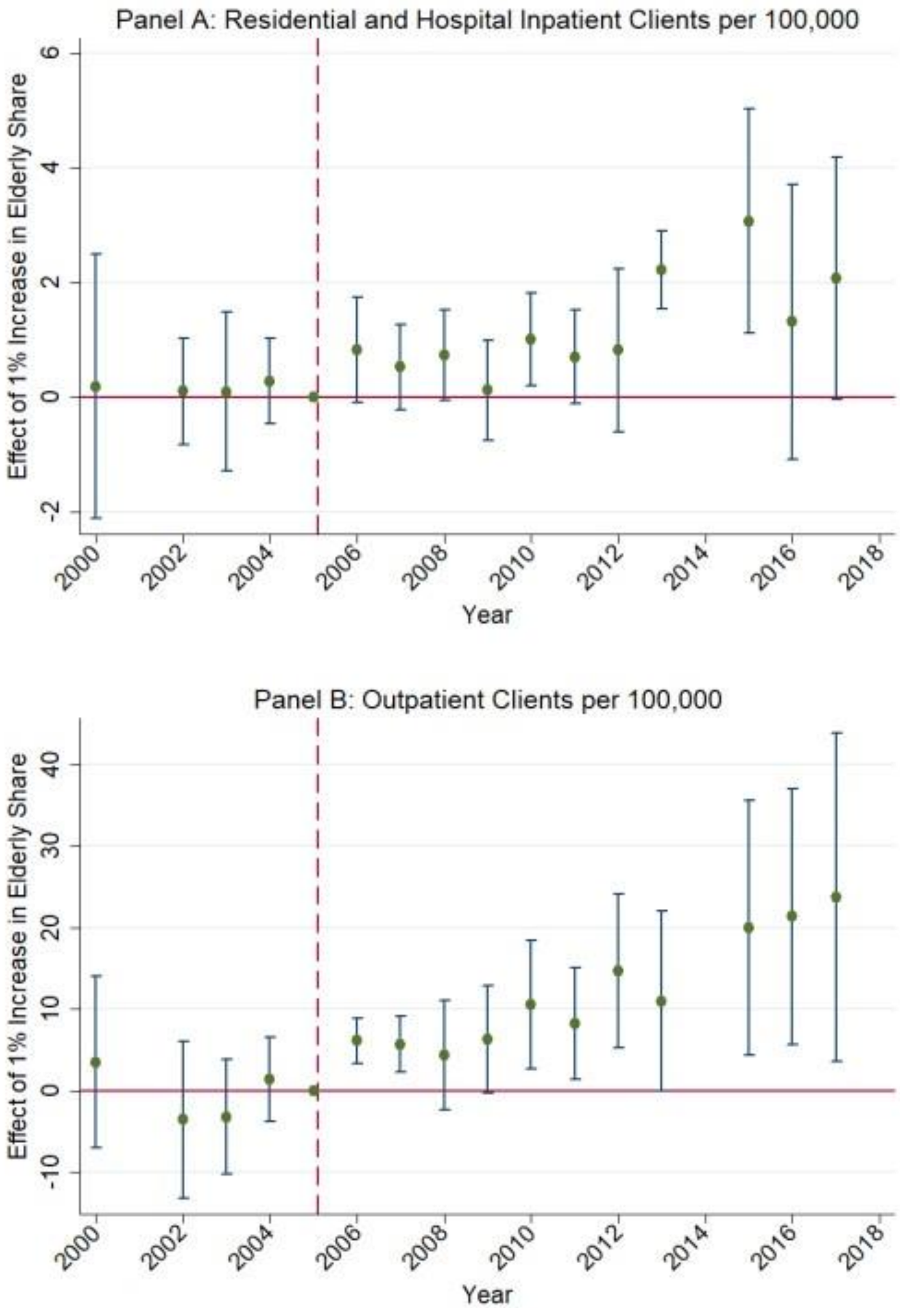


Figure 2: Event Studies of the Effects of Medicare Part D on Demand for SAT



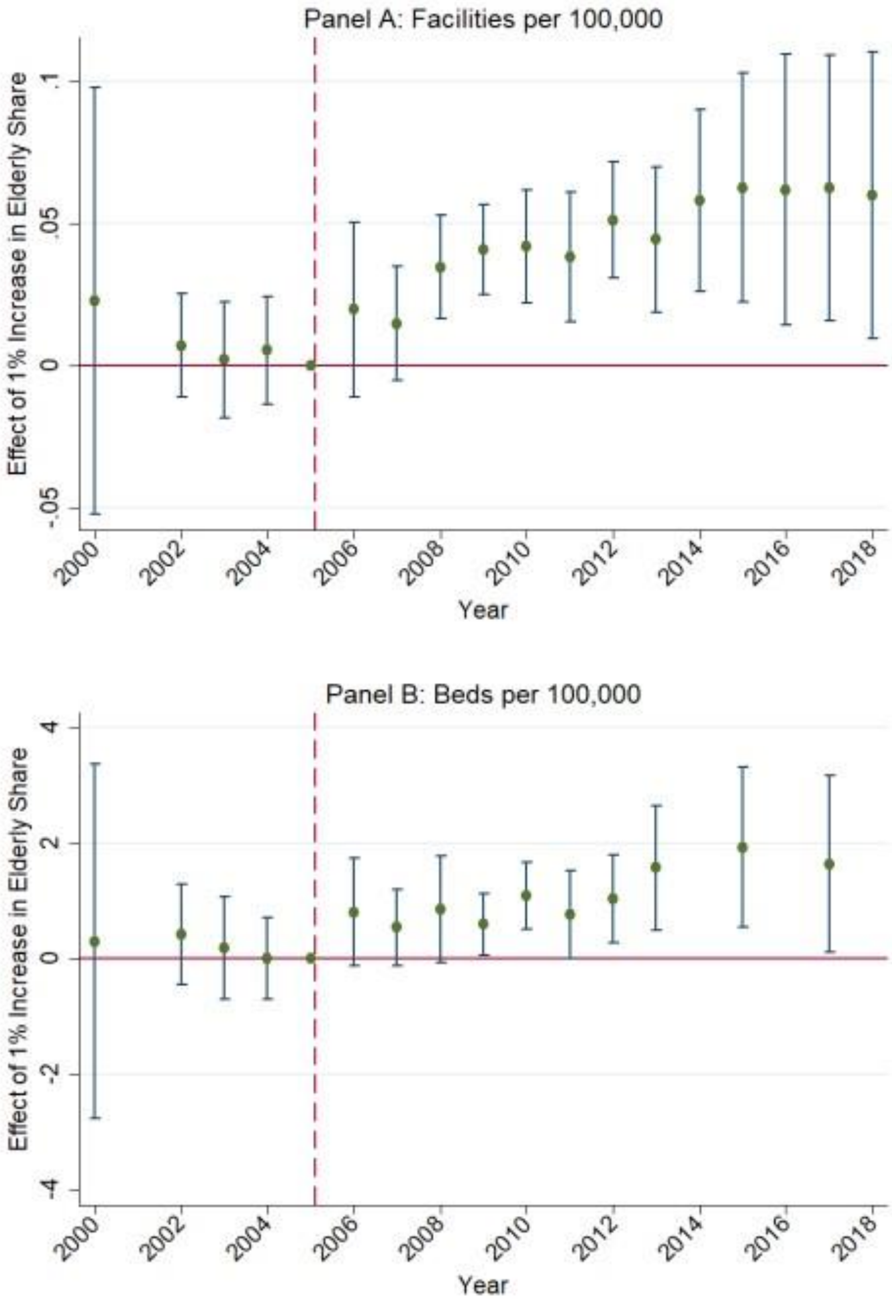
Notes: Panel A plots coefficients from yearly cross-sectional regressions between state-level elderly share and enrollment in Medicare Part D. Panel B plots event study coefficients of from estimating equation (2) with opioids distributions per 100,000 population as the outcome. Panel C plots event study coefficients from estimating equation (2) with SAT admissions for opioid use among individuals aged 12-54 per 100,000 population aged 12-54 as the outcome. Part D enrollment data from CMS, opioid distribution data from ARCOS, SAT admissions data from TEDS. Regressions producing Panels B and C include state and year fixed effects. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. All regressions are weighted by state populations. 95% confidence intervals are clustered at the state level.

Figure 3: Event Studies of the Effects of Medicare Part D on Aggregate SAT Client Counts



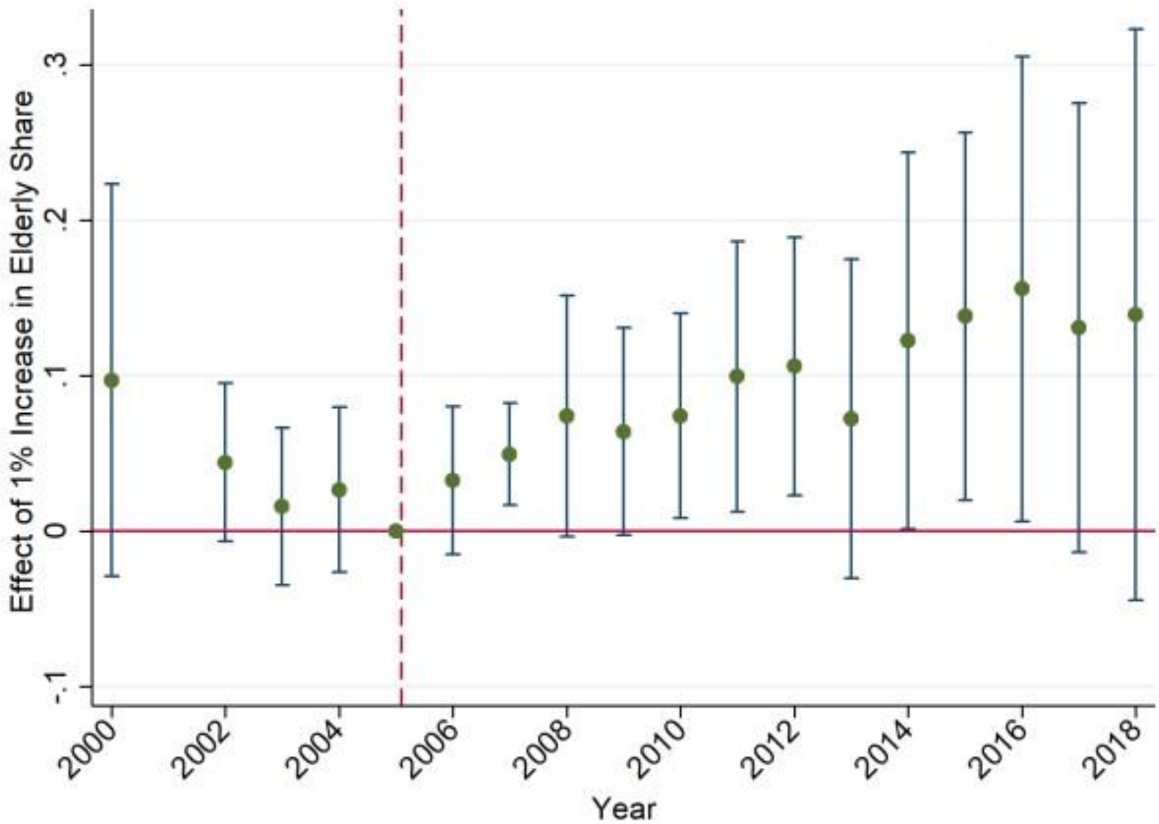
Note: Panels A and B are produced by estimating equation (2) with residential and hospital inpatient or outpatient clients per 100,000 population as the outcomes. Vertical dashed line seperates pre- and post-treatment event study coefficients. Outcome data taken from the N-SSATS. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. All regressions are weighted by state populations. 95% confidence intervals are clustered at the state level.

Figure 4: Event Studies of the Effects of Medicare Part D on Residential and Hospital Inpatient Facilities and Beds



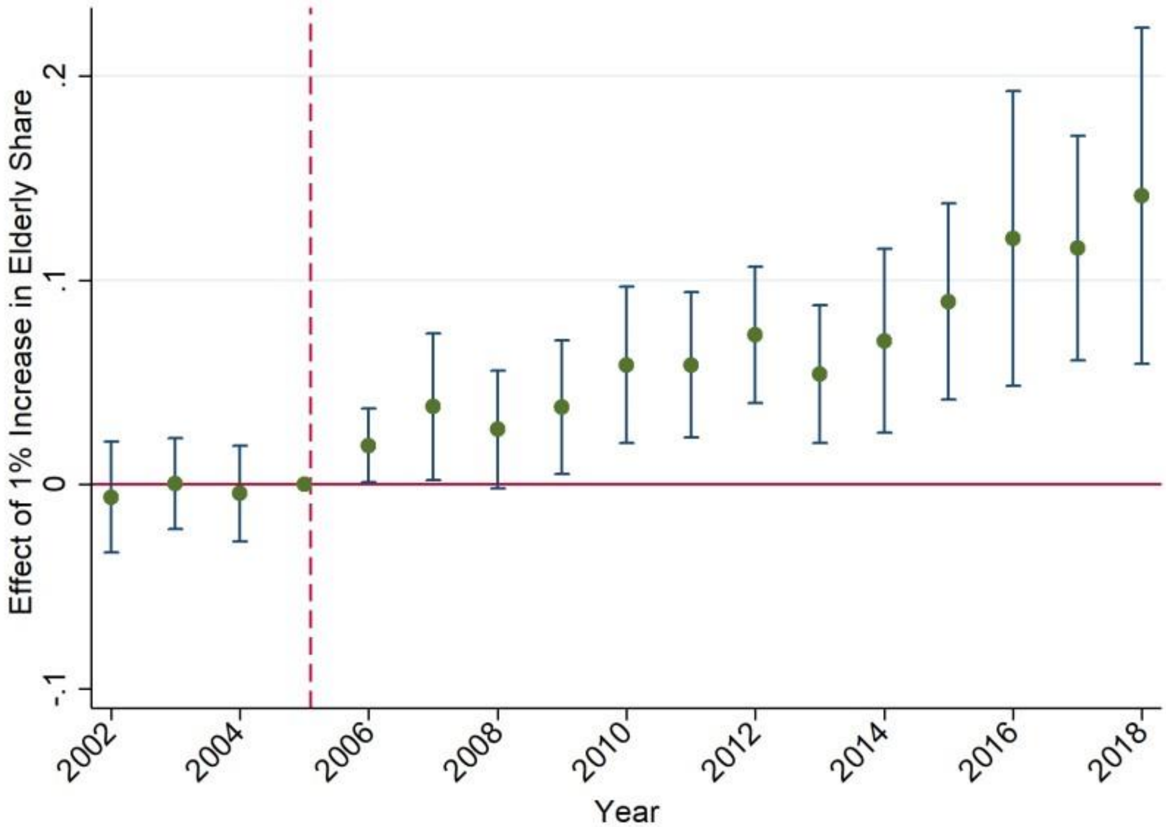
Note: Panels A and B are produced by estimating equation (2) with residential and hospital inpatient facilities or beds per 100,000 population as the outcomes. Vertical dashed line seperates pre- and post-treatment event study coefficients. Outcome data taken from the N-SSATS. Policy controls include ACA Medicaid Expansion, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. All regressions are weighted by state populations. 95% confidence intervals are clustered at the state level.

Figure 5: Event Study of the Effect of Medicare Part D on Outpatient Facilities per 100,000



Note: Figure 5 plots the event study coefficients produced by estimating equation (2) with outpatient facilities per 100,000 population as the outcome. Vertical dashed line seperates pre- and post-treatment event study coefficients. Outcome data taken from the N-SSATS. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. All regressions are weighted by state populations. 95% confidence intervals are clustered at the state level.

Figure 6: Event Study of the Effect of Medicare Part D on Facilities Offering at Least One Form of MAT for OUD per 100,000



Note: Figure 6 plots the event study coefficients produced by estimating equation (2) with facilities offering at least one form of MAT per 100,000 population as the outcome. Vertical dashed line seperates pre- and post-treatment event study coefficients. Outcome data taken from the N-SSATS. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. All regressions are weighted by state populations. 95% confidence intervals are clustered at the state level.

Table 1: Effect of Medicare Part D on Aggregate SAT Client Counts

Outcome:	Clients per 100,000				Client Count
	(1)	(2)	(3)	(4)	(5)
Panel A: Residential and Hospital Inpatients					
% Aged $65+_{2003}$ × Post	1.235**	0.998***	1.039***	1.427**	0.035***
	(0.466)	(0.349)	(0.300)	(0.557)	(0.012)
N	816	816	816	816	816
Mean of Outcome (2000-05)	40.994	40.994	40.994	40.994	2328.19
Outcome:	Clients per 100,000				Client Count
	(1)	(2)	(3)	(4)	(5)
Panel B: Outpatients					
% Aged $65+_{2003}$ × Post	14.155*	13.162**	11.893**	12.128**	0.044**
	(7.552)	(6.102)	(5.146)	(5.849)	(0.018)
N	816	816	816	816	816
Mean of Outcome (2000-05)	344.277	344.277	344.277	344.277	19552.47
Policy Controls		X	X	X	X
Other Controls			X	X	X
Region×Year FEs				X	X
Model	Linear	Linear	Linear	Linear	Poisson

Note: Regressions produced by estimating equation (1) with residential and hospital inpatient or outpatient clients (rates per 100,000 population or counts) as the outcomes. Outcome data taken from the N-SSATS. Each specification includes state and year fixed effects. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Linear regressions are weighted by population. Standard errors clustered at the state level. * p <.1, ** p<.05, *** p<.01.

Table 2: Effect of Medicare Part D on Residential and Hospital Inpatient Treatment Facilities and Beds

Outcome:	Facilities per 100,000				Facility Count
	(1)	(2)	(3)	(4)	(5)
Panel A: Facilities					
% Aged $65+_{2003}$ × Post	0.036**	0.031**	0.037***	0.043**	0.034***
	(0.014)	(0.012)	(0.011)	(0.017)	(0.010)
Mean of Outcome (2000-05)	1.587	1.587	1.587	1.587	82.867
N	918	918	918	918	918
Outcome:	**Beds per 100,000**				**Bed Count**
	(1)	(2)	(3)	(4)	(5)
Panel B: Beds					
% Aged $65+_{2003}$ × Post	1.140**	0.921***	0.877***	1.300**	0.035***
	(0.518)	(0.338)	(0.282)	(0.554)	(0.011)
Mean of Outcome (2000-05)	43.058	43.058	43.058	43.058	2445.379
N	765	765	765	765	765
Outcome:	**Beds per Facility**				
	(1)	(2)	(3)	(4)	
Panel C: Beds per Facility					
% Aged $65+_{2003}$ × Post	0.043	0.042	-0.041	-0.066	
	(0.156)	(0.155)	(0.133)	(0.184)	
Mean of Outcome (2000-05)	30.210	30.210	30.210	30.210	
95% Confidence Interval	[-0.271, 0.357]	[-0.269, 0.353]	[-0.308, 0.227]	[-0.437, 0.304]	
N	712	712	712	712	
Policy Controls		X	X	X	X
Other Controls			X	X	X
Region×Year FEs				X	X
Model	Linear	Linear	Linear	Linear	Poisson

Note: Regressions produced by estimating equation (1). The outcomes in Panel A are the rate of residential and hospital inpatient facilities per 100,000 population or facility count. The outcomes in Panel B are the rate of residential and hospital inpatient beds per 100,000 population or bed count. The outcome in Panel C is the ratio of beds to residential and hospital inpatient facilities. Outcome data are taken from the N-SSATS in Panels A and B. Outcome data for Panel C taken from the N-SSATS Annual Reports (SAMHSA, 2020b), which do not include data on hospital inpatient clients or beds in 2013. Each specification includes state and year fixed effects. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Linear regressions in Panels A and B are weighted by population. Regressions in Panel C weighted by facility count. Standard errors clustered at the state level. * p <.1, ** p<.05, *** p<.01.

Table 3: Effect of Medicare Part D on Outpatient Treatment Facilities

Outcome:	Facilities per 100,000				Facility Count
	(1)	(2)	(3)	(4)	(5)
% Aged $65+_{2003}$ × Post	0.061	0.047	0.058	0.097	0.031*
	(0.044)	(0.034)	(0.38)	(0.058)	(0.016)
Mean of Outcome (2000-05)	3.935	3.935	3.935	3.935	223.483
N	918	918	918	918	918
Policy Controls		X	X	X	X
Other Controls			X	X	X
Region×Year FEs				X	X
Model	Linear	Linear	Linear	Linear	Poisson

Note: Regressions produced by estimating equation (1) with outpatient facilities per 100,000 population as the outcome. Outcome data taken from the N-SSATS. Each specification includes state and year fixed effects. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Linear regressions are weighted by population. Standard errors clustered at the state level. * p <.1, ** p<.05, *** p<.01.

Table 4: Effect of Medicare Part D on the Number of Facilities per 100,000 Offering MATs or OTPs

Outcome:	(1) ≥ 1 MAT	(2) ≥ 1 MAT	(3) ≥ 1 MAT	(4) 0 MAT	(5) ≥ 2 MAT	(6) OTP
% Aged $65+_{2003}$ × Post	0.081**	0.068***	0.053**	0.028	0.022	0.006
	(0.036)	(0.020)	(0.022)	(0.041)	(0.016)	(0.008)
Mean of Outcome (2002-05)	0.984	0.984	0.984	3.757	0.239	0.381
N	867	867	867	867	867	867
Policy Controls		X	X	X	X	X
Other Controls		X	X	X	X	X
Region×Year FEs			X			
Model	Linear	Linear	Linear	Linear	Linear	Linear

Note: Regressions produced by estimating equation (1). The outcomes are the numbers of facilities per 100,000 population with: at least one form of MAT for OUD, zero forms of MAT, at least two forms of MAT, or that offer a OTP. Each specification includes state and year fixed effects. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Linear regressions weighted by population. Standard errors clustered at the state level. * p <.1, ** p<.05, *** p<.01.

Table 5: Effect of Medicare Part D on the Number of Facilities per 100,000 Offering Specific MATs

Outcome:	(1) Naltrexone	(2) Naltrexone	(3) Naltrexone	(4) Buprenorphine	(5) Methadone
% Aged 65+$_{2003}$ × Post	0.072**	0.061***	0.052**	0.025	0.008
	(0.028)	(0.018)	(0.019)	(0.018)	(0.010)
Mean of Outcome	0.579	0.579	0.579	0.336	0.427
N	867	867	867	816	918
Policy Controls		X	X	X	X
Other Controls		X	X	X	X
Region×Year FEs			X		
Model	Linear	Linear	Linear	Linear	Linear

Note: Regressions produced by estimating equation (1). The outcomes are the numbers of facilities per 100,000 population that offer nalterxone, buprenorphine, or methadone. Outcome data taken from the N-SSATS. Estimates use a sample period beginning in 2002 for columns (1)-(3), 2003 for column (4), and 2000 for column (5). Each specification includes state and year fixed effects. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Linear regressions weighted by population. Standard errors clustered at the state level. * p <.1, ** p<.05, *** p<.01.

Table 6: Effect of Medicare Part D on SAT Facilities by Ownership Status

Ownership status:	For-Profit Ownership	Non-Profit Ownership	Public Ownership
	(1)	(2)	(3)
Panel A: Residential and Hospital Inpatient Facilities			
% Aged $65+_{2003}$ × Post	0.006	0.018	0.013**
	(0.008)	(0.013)	(0.006)
Mean of Outcome (2000-05)	0.225	1.157	0.206
N	918	918	918
Ownership status:	**For-Profit Ownership**	**Non-Profit Ownership**	**Public Ownership**
	(1)	(2)	(3)
Panel B: Outpatient Facilities			
% Aged $65+_{2003}$ × Post	0.004	0.027	0.027**
	(0.025)	(0.024)	(0.011)
Mean of Outcome (2000-05)	1.149	2.193	0.593
N	918	918	918
Ownership status:	**For-Profit Ownership**	**Non-Profit Ownership**	**Public Ownership**
	(1)	(2)	(3)
Panel C: Facilities Offering ≥ 1 MAT			
% Aged $65+_{2003}$ × Post	0.015	0.051***	0.002
	(0.010)	(0.019)	(0.005)
Mean of Outcome (2002-05)	0.331	0.481	0.172
N	867	867	867
Policy Controls	X	X	X
Other Controls	X	X	X
Model	Linear	Linear	Linear

Note: Regressions produced by estimating equation (1). The outcomes in Panel A are for-profit, non-profit, or government-run facilities that accept residential and/or hospital inpatient clients. The outcomes in Panel B are similar but for facilities that accept outpatients. The outcomes in Panel C are for facilities that offer at least one form of MAT. Outcome data taken from the N-SSATS. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Regressions are weighted by state population. Standard errors clustered at the state level. * p <.1, ** p<.05, *** p<.01.

Table 7: Effect of Opioid Distribution on SAT Availability: Instrumenting for Opioid Distribution with Medicare Part D

Outcome:	Facilities per 100,000		Beds per 100,000	
	(1)	(2)	(3)	(4)
Panel A: Residential and Hospital Inpatient Facilities				
MEDs per capita	0.273**	1.038***	10.105**	24.975***
	(0.116)	(0.260)	(4.550)	(7.183)
F-Statistic	–	40.81	–	62.28
Mean of Outcome (2000-05)	1.587	1.587	43.058	43.058
N	867	867	765	765
Outcome:	Facilities per 100,000			
	(1)	(2)		
Panel B: Outpatient Facilities				
MEDs per capita	0.385	1.627*		
	(0.312)	(0.883)		
F-Statistic	–	40.81		
Mean of Outcome (2000-05)	3.935	3.935		
N	867	867		
Outcome:	Facilities per 100,000			
	(1)	(2)		
Panel C: Facilities Offering ≥ 1 MAT				
MEDs per capita	0.179	2.178***		
	(0.206)	(0.533)		
F-Statistic	–	32.54		
Mean of Outcome (2002-05)	0.984	0.984		
N	816	816		
Policy Controls	X	X	X	X
Other Controls	X	X	X	X
Estimator	OLS	2SLS	OLS	2SLS

Note: Regressions produced by estimating equation (1). The outcomes in Panel A are residential and hospital inpatient facilities/beds per 100,000 population. The outcomes in Panel B are outpatient facilities per 100,000 population. The outcomes in Panel C are facilities that offer at least one form of MAT per 100,000 population. F-Statistic refers to value associated with excluded instrument. Outcome data taken from the N-SSATS. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Regressions are weighted by state population. Standard errors clustered at the state level. * p <.1, ** p<.05, *** p<.01.

2.7. APPENDIX A: CORRECTING FOR MISREPORTING

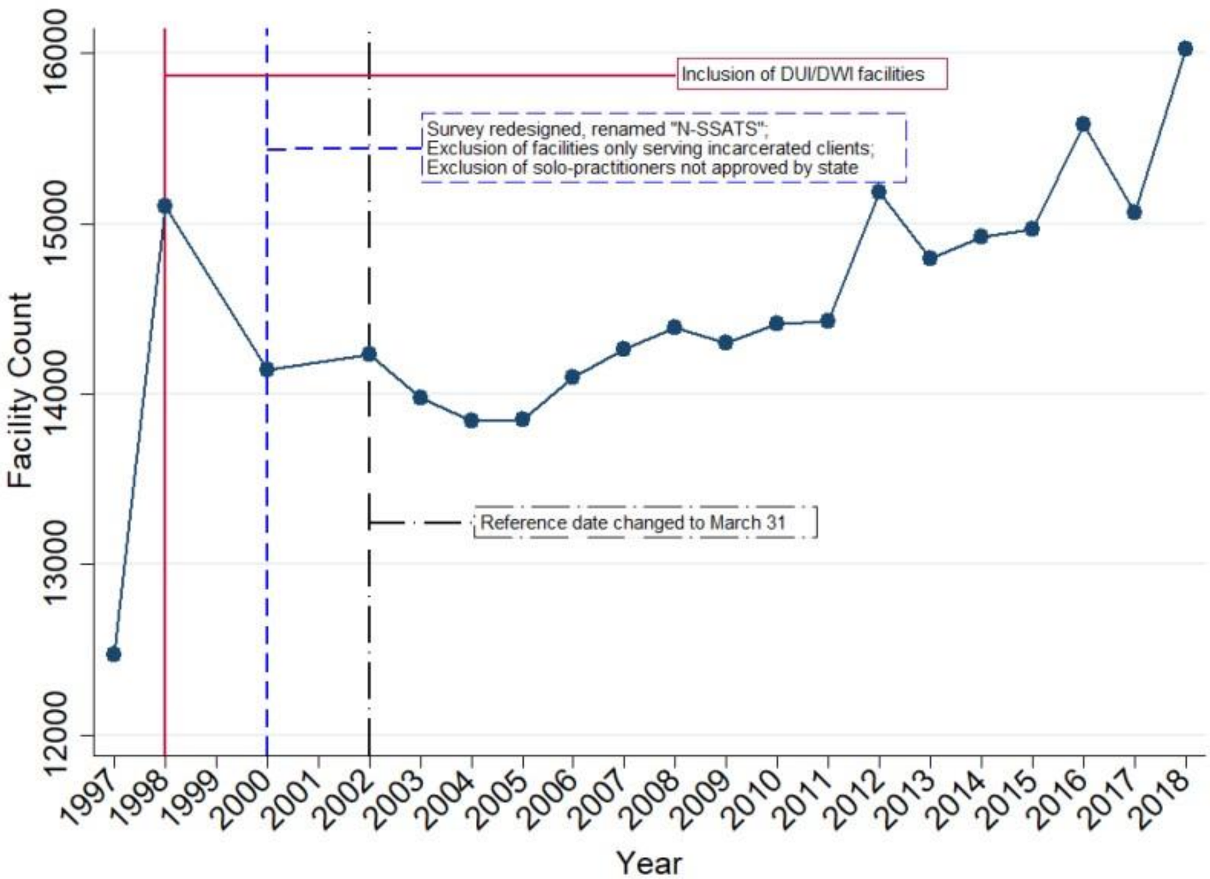
As discussed in section 2.3.1, the N-SSATS data on bed and client counts provided by SAMHSA in state-year aggregates contain several implausibly large observations. Correspondence with SAMHSA has confirmed that these anomalies are driven by single facilities that are misreporting bed and client counts in these individual years. However, since I cannot observe these counts at the facility level, I adopt the following procedure for correcting for outliers using these state-level aggregates. First, I define a state-year observation as an outlier when it is at least two times greater than the values in both adjacent years. I then replace the bed/client counts driving the outlier with a linear interpolation using the client/bed counts in these adjacent years. I provide a list of outliers selected by this criterion, adjacent year values, and result from linear interpolation in Appendix Table A.1. Note that the interpolation results are not exact averages of the adjacent years. This is because I interpolate between the underlying counts (of beds, clients) instead of the rates per 100,000 population. For example, an outlier for inpatient clients may be driven entirely by a state-year observation of residential clients, as opposed to hospital clients. In this case I would interpolate only the count of residential clients, then proceed to construct the total inpatient client rate per 100,000.

2.8. APPENDIX B: CHANGES TO N-SSATS ELIGIBILITY CRITERIA

Prior to a survey re-design in 2000, the N-SSATS was known as the Uniform Facility Data Set (UFDS). Like the N-SSATS, the goal of the UFDS was to survey the universe of SAT facilities in the United States. Unlike the N-SSATS, however, its sample design was altered with each wave as SAMHSA changed their criteria for which facilities were considered eligible. Appendix Figure A.1 plots SAMHSA's known universe of SAT facilities eligible for the N-SSATS/UFDS between 1997-2018. Counts were constructed by dividing the surveys' final sample size each year by the corresponding response rate (the survey underwent re-designs in 1999 and 2001). The large swings in facility counts from 1997 to 2000 reflect the concurrent changes in eligibility criteria and it was not until 2002 that the survey design was finalized with the change in reference date.

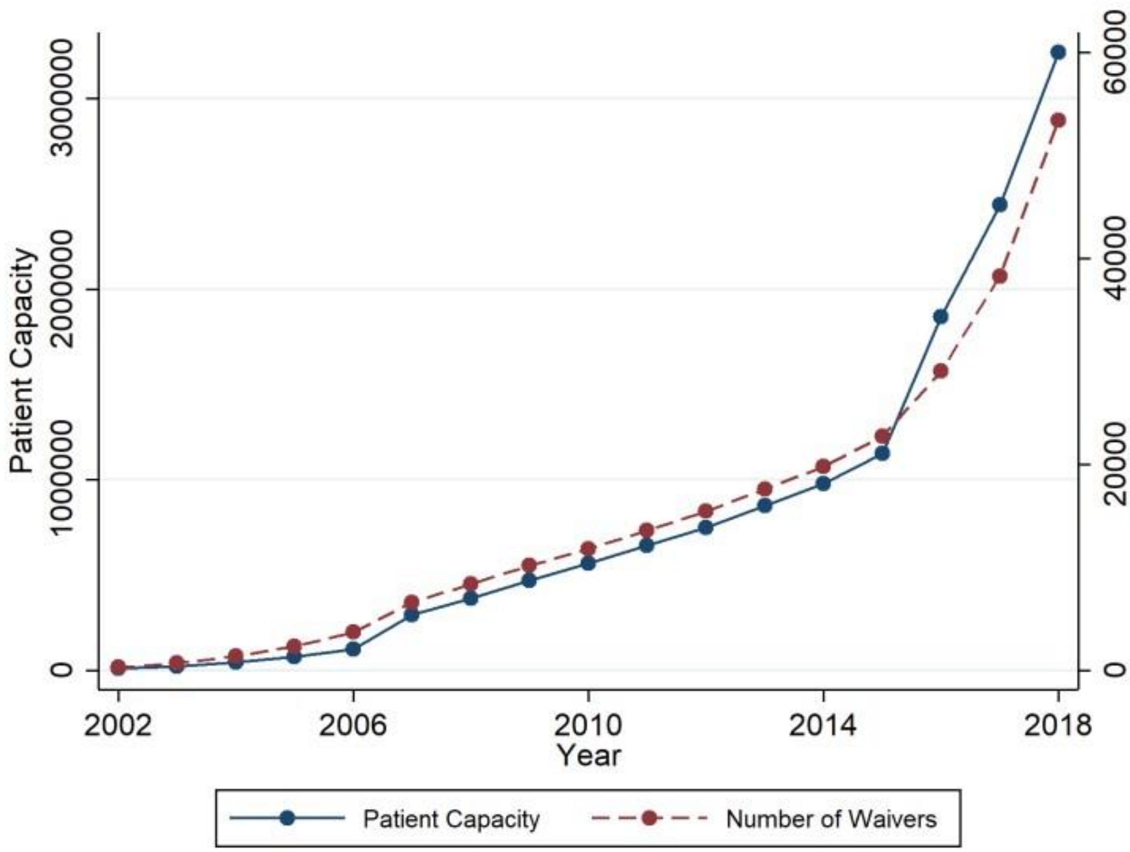
2.9 APPENDIX FIGURES AND TABLES

Figure A.1: National SAT Facility Count Over Time - N-SSATS/UFDS



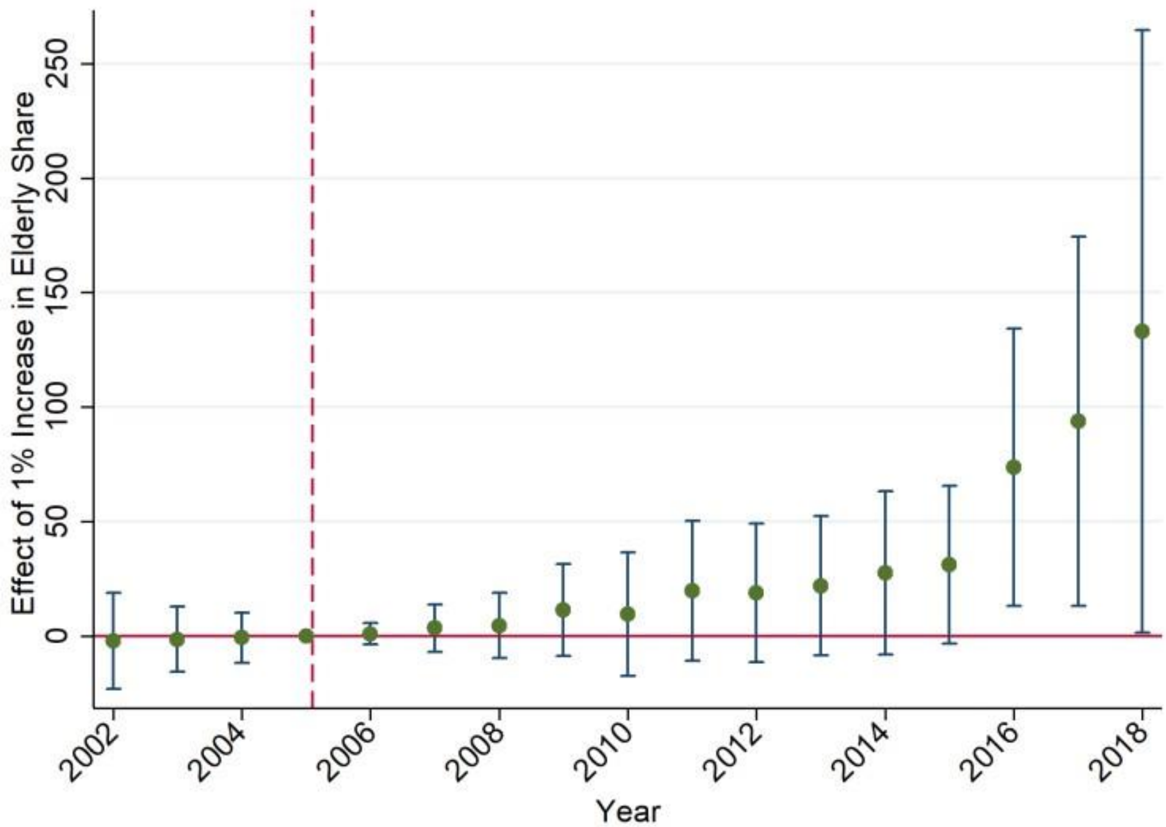
Note: This graph plots the total number of specialty SAT facilities over time. Facility counts are produced by totaling N-SSATS/UFDS respondents by year. Yearly counts are divided by yearly response rate to produce final values.

Figure A.2: Cumulative DATA 2000 Waivers and Patient Capacity Granted to Specialty and Non-specialty Providers



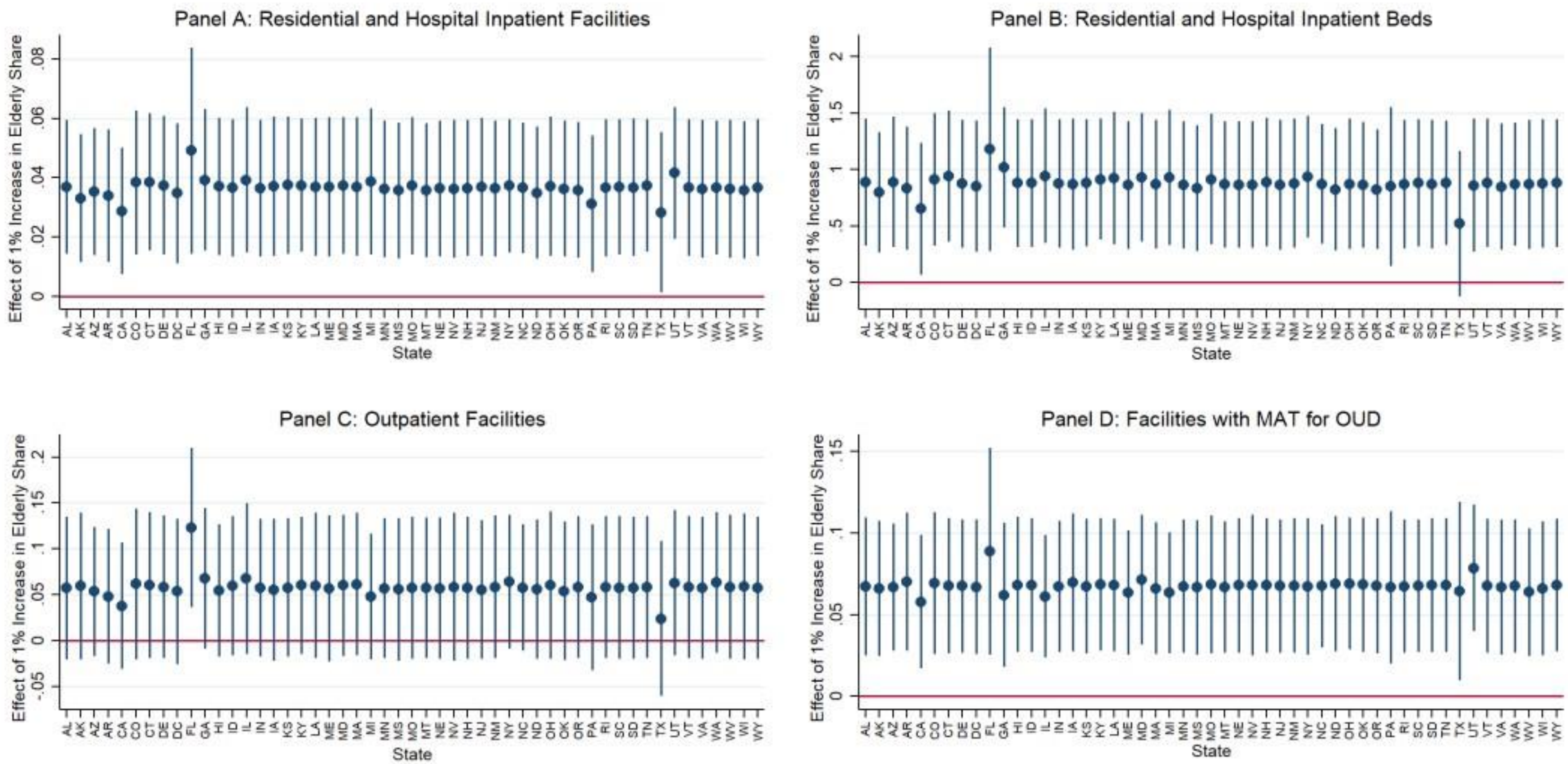
Note: The red dots display the cumulative number of DATA 2000 waivers granted to providers up to and including that year. The blue dots display the cumulative number of patients providers have been granted to treat via DATA 2000 waivers. A year's total DATA 2000 patient capacity granted as of year t can be calculated through the following formula: $CAPACITY_t = 30 \times N30_t + 100 \times N100_t + 275 \times N275_t$ where $N30_t$, $N100_t$, $N275_t$ represent the number of waivers granted for each patient limit as of year t.

Figure A.3: Event Study of the Effect of Medicare Part D on DATA 2000 Patient Capacity per 100,000



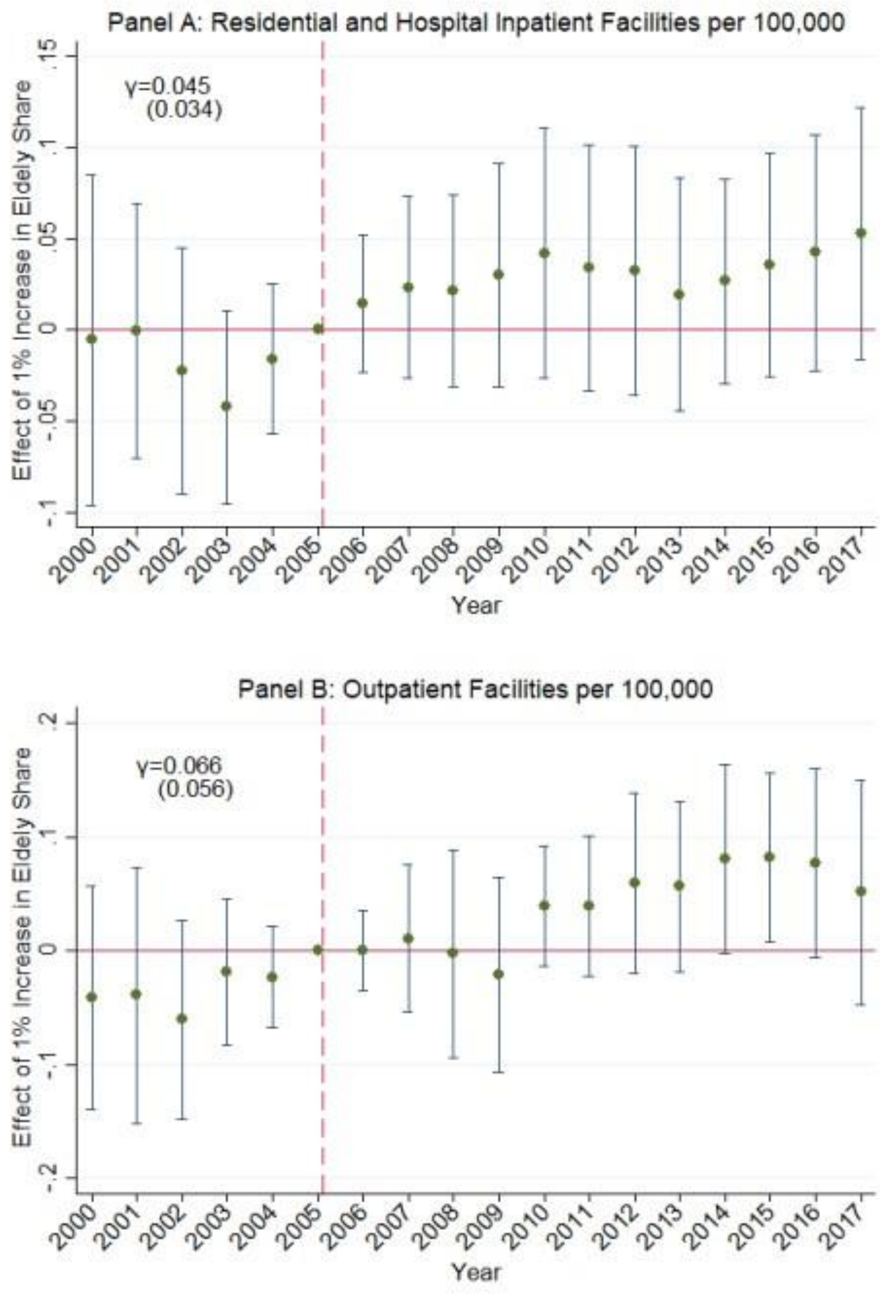
Note: Figure 7 plots the event study coefficients produced by estimating equation (2) with the cumulative amount of DATA 2000 patient treatment capacity per 100,000 population as the outcome. Vertical dashed line seperates pre- and post-treatment event study coefficients. Outcome data received by request from SAMHSA. Policy controls include ACA Medicaid Expansion, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. All regressions are weighted by state populations. 95% confidence intervals are clustered at the state level.

Figure A.4: Leave-One-Out Tests



Note: Panels A-D contain the results of re-estimating equation (1) for each of the following outcomes, while sequentially removing one state a time: residential and hospital inpatient facilities, residential and hospital inpatient beds, outpatient facilities, and facilities with MAT for OUD, per 100,000 population. Outcome data taken from the N-SSATS. Each point estimate is the coefficient on the share of the population aged 65+. 95% percent confidence intervals are cluster-robust at the state level.

Figure A.5: Event Studies Using the County Business Patterns



Note: The graphs in Panels A and B are produced by estimating equation (2) with residential and hospital inpatient facilities or beds per 100,000 population as the outcomes. Point estimates and standard errors from DD models obtained by estimating equation (1). Vertical dashed line seperates pre- and post-treatment event study coefficients. Outcome data taken from the County Business Patterns 2000-2017. Policy controls include ACA Medicaid Expansion, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. All regressions are weighted by state populations. 95% confidence intervals and standard errors are clustered at the state level.

Table A.1: Correcting for Misreporting

State	Year	Outlier	Adjacent Year (left)	Adjacent Year (right)	Interpolation
Panel A: Residential and Hospital Inpatient Clients per 100,000					
Arkansas	2015	98.062	29.215	23.288	27.658
Delaware	2015	448.508	33.391	23.281	25.223
Rhode Island	2009	394.184	38.894	42.073	37.337
Rhode Island	2015	571.215	36.333	32.920	37.654
South Carolina	2015	39.320	19.607	17.253	19.437
Washington	2015	101.249	34.309	34.719	34.270
Panel B: Outpatient Clients per 100,000					
Montana	2011	1020.747	302.392	460.645	377.107
Panel C: Residential and Hospital Inpatient Beds per 100,000					
Rhode Island	2015	571.215	38.921	36.224	39.770
Washington	2015	86.770	38.703	40.872	39.253

Note: This table displays outliers in the bed and client data according to the criterion laid-out in the Appendix. It also shows the corrected values for the outliers, calculated by linearly interpolating between the bed and clients counts for the state-year observations adjacent to the outlier, then using the interpolated count to construct the final rate per 100,000.

Table A.2: Impact of Part D on Controls and N-SSATS Response Rates

Outcome:	Policy Indicator					
	(1) ACA Medicaid Expansions	(2) HIFA Waivers	(3) SAT Parity Laws	(4) Medical Marijuana Laws	(5) Pain Clinic Laws	(6) Strong PDMPs
Panel A: State Policies						
% Aged $65+_{2003}$ × Post	-0.016	0.005	-0.037	0.013	0.010	0.003
	(0.035)	(0.010)	(0.039)	(0.024)	(0.055)	(0.023)
Mean of Outcome (2000-05)	0.000	0.037	0.055	0.048	0.000	0.000
95% Confidence Interval	[-0.085, 0.054]	[-0.015, 0.025]	[-0.115, 0.042]	[-0.034, 0.061]	[-0.100, 0.120]	[-0.044, 0.050]
N	969	969	969	969	969	969
Outcome:	UI Rate	ln(Population)	% White	N-SSATS Response Rate		
	(1)	(2)	(3)	(4)		
Panel B: Other Controls and N-SSATS Response Rate						
% Aged $65+_{2003}$ × Post	0.073	-0.009	0.001	0.001		
	(0.092)	(0.010)	(0.001)	(0.001)		
Mean of Outcome (2000-05)	5.209	16.005	0.811	0.953		
95% Confidence Interval	[-0.112, 0.257]	[-0.029, 0.011]	[-0.000, 0.002]	[-0.002, 0.004]		
N	969	969	969	969		
Model	Linear	Linear	Linear	Linear	Linear	Linear

Note: Regressions produced by estimating equation (1) with each control variable as the outcome. Each specification includes state and year fixed effects. Linear regressions are weighted by population. Standard errors clustered at the state level. * p <.1, ** p<.05, *** p<.01.

Table A.3: Effect of Medicare Part D on SAT Facility Utilization Rates

Outcome:	Clients-per-Facility			Clients-per-Bed		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Residential and Hospital Inpatient						
% Aged $65+_{2003}$ × Post	0.200	0.179	0.118	0.005	0.007*	0.008
	(0.208)	(0.151)	(0.229)	(0.005)	(0.004)	(0.006)
Mean of Outcome (2000-05)	27.244	27.244	27.244	0.906	0.906	0.906
95% Confidence Interval	[-0.217, 0.618]	[-0.125, 0.482]	[-0.342, 0.577]	[-0.004, 0.015]	[-0.001, 0.015]	[-0.005, 0.020]
N	714	714	714	714	714	714
Outcome:	Clients-per-Facility					
	(1)	(2)	(3)			
Panel B: Outpatient						
% Aged $65+_{2003}$ × Post	3.064*	2.192**	0.874			
	(1.599)	(1.035)	(0.908)			
Mean of Outcome (2000-05)	87.591	87.591	87.591			
95% Confidence Interval	[-0.148, 6.277]	[0.113, 4.271]	[-0.950, 2.698]			
N	816	816	816			
Policy Controls		X	X		X	X
Other Controls		X	X		X	X
Region×Year FEs			X			X
Model	Linear	Linear	Linear	Linear	Linear	Linear

Note: Regressions produced by estimating equation (1). The outcome in Panel A, columns (1)-(3) is the ratio of residential and hospital inpatient clients to facilities of the same type. The outcome in Panel A, columns (4)-(6) is the ratio of residential and hospital inpatient clients to beds. The outcome in Panel B, columns (1)-(3) is the ratio of outpatient clients to outpatient facilities. Outcome data taken from N-SSATS and the N-SSATS Annual Reports (SAMHSA, 2020b). The models producing Panel A restrict the sample only to facilities that report both client and bed counts in the same year. Each specification includes state and year fixed effects. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Linear regressions are weighted by the outcomes' denominators. Standard errors clustered at the state level. * p<.1, ** p<.05, *** p<.01.

Table A.4: Effect of Medicare Part D on Cumulative DATA 2000 Patient Capacity

Outcome:	Patient Capacity per 100,000			
	(1)	(2)	(3)	(4)
Panel A: DATA 2000 Waivers Ever Granted				
% Aged $65+_{2003}$ × Post	54.461**	50.207**	31.835**	18.048
	(26.547)	(20.162)	(15.799)	(11.430)
Mean of Outcome (2002-05)	43.011	43.011	43.011	43.011
N	867	867	867	867
Outcome:	Patient Capacity per 100,000			
	(1)	(2)	(3)	(4)
Panel B: DATA 2000 Waivers Granted to Practitioners Operating in 2020				
% Aged $65+_{2003}$ × Post	26.989*	24.262**	15.026*	6.844
	(14.921)	(9.732)	(8.898)	(6.330)
Mean of Outcome (2002-05)	13.463	13.463	13.463	13.463
N	867	867	867	867
Policy Controls		X	X	X
Other Controls			X	X
Region× Year FEs				X
Model	Linear	Linear	Linear	Linear

Note: Regressions produced by estimating equation (1). The outcome in Panel A is the cumulative amount of DATA 2000 capacity granted to providers in a given state, per 100,000 population. The outcome in Panel B is the cumulative amount of DATA 2000 capacity granted to providers in a given state who are stillp practicing, per 100,000 population. Outcome data taken from the N-SSATS Annual Reports. Panel B uses a sample period beginning in 2002 for columns (1)-(3), 2003 for column (4), and 2000 for column (5). Each specification includes state and year fixed effects. Policy controls include ACA Medicaid Expansion, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Linear regressions weighted by population. Standard errors clustered at the state level. * p <.1, ** p<.05, *** p<.01.

Table A.5: Effect of Medicare Part D on Medicare Admissions and Acceptance at SAT Facilities

Outcome:	Admissions per 100,000	Residential/Hospital Inpatient Facilities per 100,000	Outpatient Facilities per 100,000	Facilities with MAT per 100,000
	(1)	(2)	(3)	(4)
% Aged $65+_{2003}$ × Post	2.675	0.008	0.003	0.021
	(1.934)	(0.005)	(0.014)	(0.019)
Mean of Outcome (2000-05)	24.078	0.475	1.492	0.458
N	898	918	918	918
Policy Controls	X	X	X	X
Other Controls	X	X	X	X
Model	Linear	Linear	Linear	Linear

Note: Regressions produced by estimating equation (1). The outcome in column (1) is the number of SAT admissions per 100,000 population, where each admission reports having public insurance that is not Medicaid (i.e. Medicare, CHAMPUS, etc.). Column (1) outcome data are taken from the TEDS. The outcomes in columns (2)-(4) are the numbers of residential/hospital inpatient facilities, outpatient facilities, and facilities with MAT per 100,000 that accept Medicare. Columns (2)-(4) outcome data are taken from the N-SSATS. Each specification includes state and year fixed effects. Policy controls include ACA Medicaid Expansion, SAT Parity, HIFA Waiver, Pain Clinic Regulation, Medical Marijuana, and Strong PDMP dummies. Other controls include state unemployment rates, natural log of the population, and percent of the population white. Linear regressions are weighted by population. Standard errors clustered at the state level. * p<.1, ** p<.05, *** p<.01.

CHAPTER 3

A FINE PREDICAMENT: CONDITIONING, COMPLIANCE, AND CONSEQUENCES IN A LABELED CASH TRANSFER PROGRAM

3.1. INTRODUCTION

Cash transfers are one of the most popular forms of aid interventions directed toward reducing poverty and the intergenerational transmission of poverty. More than a fifth of all countries have implemented a conditional cash transfer (CCT) program, including about one-third of developing and middle-income countries (Morais de Sá e Silva, 2017). Although most of the inaugural cash transfers programs and many subsequent program efforts have imposed conditions on households' receipt of cash transfers that prescribe how the monies should be used (Baird et al., 2013), unconditional cash transfer (UCT) programs are proliferating as well and are among some of the largest cash transfer programs today (e.g., China's dibao program with about 75 million beneficiaries) (Golan et al., 2015). In fact, because the implementation and enforcement of conditions requires substantial infrastructure and administrative capacity, the implementation of UCTs has become more commonplace in very low-income countries, and “labeled” cash transfer programs (LCTs), where guidance for spending the transfer is articulated but not monitored or enforced, have also been introduced (Benhassine et al., 2013).

In this research, we focus on an under-explored consequence of complying with conditions for households–the costs to them when financial penalties are incurred because of failure to comply with conditions. We undertake this analysis in the context of the Kenya Cash Transfer Programme for Orphans and Vulnerable Children (CT-OVC), a LCT that was noteworthy in its random assignment of health and schooling conditions *with penalties* (CCTs) to a subset of locations in the treatment group. We exploit the random assignment to the conditional treatment arm in the Kenya CT-OVC to explore the implications of penalty fines on household outcomes, given that the "labeling" of the cash transfers resulted in households in both treatment arms having similar beliefs regarding program rules and expectations. Our research, which shows how conditioning with penalties can unintentionally harm those most in need of assistance, has clear policy implications for the design and evolution of cash transfer programs and our understanding of how households respond to income shocks.

Although there is a very large literature on CCTs and UCTs, we identified only one prior study that compared a CCT version of a cash transfer program with an LCT, an evaluation by Benhassine et al. (2013) of the Tayssir cash transfer program in rural Morocco. The Tayssir program is distinct from the Kenya CT-OVC, in that it was a pilot program focused on school-aged children (6-15 years), with receipt of the cash transfers tied to a specific education goal (reductions in school absences), and it had a lower transfer amount as a fraction of baseline average household consumption than the CT-OVC (5% versus 23%).

Our analysis of the Kenya CT-OVC program produces four main findings. First, over a third of households in the conditional treatment arm were ever subjected to penalty fines, and the likelihood of being penalized was greatest for households with the lowest consumption at baseline. Second, we find that despite the high frequency of penalization, perceptions about the rules and requirements for receiving the transfer differed very little between treatment arms. Third, our results indicate that the conditioned-upon outcomes did not differ significantly between CCT and LCT treatment arms at follow-up. More specifically, although limits to our statistical power do not allow us to completely rule out the potential for meaningful differences in conditioned-upon outcomes, such as fewer days missed from school, we did not detect any statistically significant effects of assignment to the CCT arm (versus the LCT arm). Finally, we find that assignment to the CCT arm (versus the LCT arm) resulted in large decreases in non-food consumption at follow-up among households in the bottom quartile of baseline consumption, presumably as a result of the penalty fines. These findings affirm the conventional wisdom that penalties in cash transfer programs disproportionately harm those who are least able to respond to them.

In the following section 3.2, we review the literature on conditional, unconditional and labeled cash transfer programs (including the Tayssir program), focusing on the types of conditions or guidance embodied in the programs, how they were implemented, and evidence on the relationship of conditions to program outcomes. In section 3.3, we present background information on the Kenya CT-OVC program and the nature of the conditions, penalties and labeling of the cash transfers. We also describe the design of the experimental evaluation, data collected and measures, and the checks we perform for covariate balance and attrition. We then present our approach to the empirical analysis and the findings of our main analyses comparing the CCT and LCT treatment arms in section 3.4. We conclude with a discussion of the results in section 3.5.

3.2. LITERATURE REVIEW

One global estimate of the number of beneficiaries of cash transfer programs (Fiszbein et al., 2014) suggests that close to one billion people worldwide are receiving cash transfers as a form of social protection (i.e., social assistance for poor households). The implementation of many cash transfer programs has also been accompanied by rigorous evaluation efforts to identify their impacts, which has contributed to a growing evidence base on a wide range of potential program effects in education, health, labor, consumption, food security, asset building, risky behaviors and more (see: https://transfer.cpc.unc.edu/; Hidrobo et al., 2018; Ralston et al., 2017). In fact, after observing the positive findings of cash transfer programs on communities and households, some governments in poor countries are now implementing them as regular components of their economic development and social protection efforts (Bastagli et al., 2016).

As cash transfer programs have expanded to all regions of the world, variation in their implementation has spread as well, with tinkering typically around the designation and administration of conditions or rules of cash transfer receipt. Among the most common conditions are school enrollment and minimum attendance requirements for the child beneficiaries; regular health and wellness checks and immunizations for infants and young children, and health and nutrition training and information sessions for parents or caregivers of the beneficiaries. For example, two of the earliest and largest CCT programs, Mexico's Prospera program (previously named PROGRESA and Oportunidades), and Brazil's Bolsa Familia program, require households to enroll their children in school and the children to maintain 85 percent attendance rates, ensure that they get preventative healthcare (check-ups) and vaccinations, and participate in educational activities offered by health teams or attend monthly meetings to access health and education information, to receive the transfer (Levy, 2006; Fiszbein et al., 2009). While the marked success of these two CCT programs–including permanent increases in food consumption, reductions in chronic malnutrition, and increased school enrollment rates–galvanized the replication of this CCT model throughout Latin America and beyond (Fernald et al., 2008; Handa et al., 2018), the transmission of the conditionalities to other contexts has hit constraints.

The implementation and enforcement of conditions requires substantial infrastructure and administrative capacity. In Brazil, for example, local education departments are responsible for checking and reporting the school attendance rates of beneficiaries every two months through the (computerized) School Attendance Surveillance System, and principals are required to report the reasons for absences and take appropriate actions when the student attendance report is returned to the school. A separate computer system managed by the Ministry of Health, Sistema de Vigilância Alimentar e Nutricional is used by municipalities for reporting compliance with the health conditions, and municipalities are also required to verify access to quality health services for program beneficiaries. Furthermore, the direct costs of complying with conditions can be burdensome for beneficiaries and may also open the door for corruption in situations where those verifying conditions charge fees or demand payments for certifying compliance (de Brauw and Hoddinott, 2011; Heinrich and Brill, 2015).

3.2.1. Why condition?

Numerous works have articulated the arguments for and against the imposition of conditions (Ferreira, 2008; Fiszbein et al., 2009; Baird et al., 2011; de Brauw and Hoddinott, 2011), which we briefly review here. As Fiszbein et al. and Baird et al. point out, in ideal circumstances—where individuals are well-informed and make rational choices, governments are benevolent and operate efficiently, and markets function perfectly—unconditional cash transfers should be the preferred policy design from both public and private perspectives. However, if we are concerned that individuals lack information to make the most appropriate decisions for use of the transfers, the government can play a role in helping them to overcome these informational problems, e.g., conditioning receipt on uses that are believed to increase their net positive impacts. In other words, the conditions can induce a substitution effect (in spending) that enhances the overall effect of the cash transfers. Another set of arguments pertains to the political feasibility (or political benefits) of offering cash transfers, where public spending on the programs may be viewed as more palatable or popular if the cash transfers are conditioned on “good behavior” or if they are delivered as part of a “social contract” with the state that defines “co-responsibilities” (Fiszbein et al., 2009; Lindert et al., 2007). In addition, de Brauw and Hoddinott (2011) note that if the conditions serve as a mechanism for increasing the effectiveness of the transfers and politicians and policy makers can take credit for the results, the conditions may be a useful tool for helping them to stay in office as well. Lastly, a third prevailing argument in support of CCTs is that the investments in human capital encouraged through conditioning generate positive externalities for the public, such as the benefits associated with immunization, which caregivers would not fully consider in their own decision making (contributing to under investments from a societal perspective).

These potential benefits have to be weighed, however, against the (public and private) costs of administering and complying with the conditions (Baird et al., 2011). There is very limited information available on the costs associated with implementing and monitoring compliance with conditions, largely because it is difficult to distinguish these costs from other administrative costs or to identify those that are imposed on health, education sector and other social welfare staff involved in delivering services. In a study comparing program costs across three Latin American CCTs, Caldes et al. (2006) estimated the costs of conditions–distributing, collecting, and processing registration, attendance, and performance forms to schools and healthcare providers (distinguishing them from overall program monitoring and evaluation costs)–and found that the conditions constituted nearly one quarter of the administrative costs in PROGRESA (in 2000). It is also challenging to fully account for the costs of meeting conditions that are imposed on the program beneficiaries—such as transportation and other transaction costs associated with accessing required services—and to assess who bears those burdens in the household. Of course, there are also direct costs to households of any fines or penalties imposed if they are found not to be in compliance. The research base generally finds that CCTs increase total household consumption and disproportionately affect food consumption in poor households, and that increases in food expenditures are typically directed at increasing quality (e.g., items rich in protein and fruits and vegetables) (Fiszbein et al., 2009; Hoddinott, Skoufias, and Washburn 2000; Macours, Schady, and Vakis, 2012; Maluccio and Flores 2005). If the households who find it most challenging to satisfy the conditions are among the poorest of program eligible, this could unduly penalize household consumption among those most in need (de Brauw and Hoddinott, 2011; Heinrich & Brill, 2015; Rodríguez-Castelán, 2017). Indeed, research summarized by Handa et al. (2016) suggests that UCTs (including Kenya's CT-OVC program) likewise have strong, positive effects on household consumption, and hence, any penalties associated with noncompliance in CCTs may be unjustifiably punitive.

3.2.2. Nature, role and effects of conditions in cash transfer programs

In the growing evidence base on CCTs, UCTs, and their program variants, researchers have sought to characterize the nature and role of conditions in implementation and to understand how they relate to program effectiveness (Morais de Sá e Silva, 2017). In their 2013 meta-analytic review of 35 studies of cash transfer programs focused on CCTs with at least one condition tied to schooling, Baird et al. conceded that the binary classification of CCTs vs. UCTs disregarded considerable variation in the nature and intensity of the conditions. In their analysis, they further categorized the cash transfer programs as having: (i) no schooling conditions, (ii) some schooling conditions with no enforcement or monitoring, and (iii) explicit schooling conditions that were monitored and enforced; within each of these categories, they attempted to capture variation in nature and intensity of the conditions. For example, Baird et al. describe both Bolsa Familia and

PROGRESA as having "explicit conditions," but with imperfect monitoring and minimal enforcement. Other research similarly suggests that the distinction between the second and third categories may not always be precise; that is, there may be more of a gradation from monitoring and enforcement to no monitoring and enforcement in many programs, where the degree of "softness" is realized in implementation of the cash transfer programs (Fizbein et al., 2009; Ralston et al., 2017; Hidrobo et al., 2018). Silva (2007), for instance, describes the Bolsa Familia conditions as a "soft type of conditionalities," where the sanctions imposed for not complying with conditions are moderate and implemented at different levels, ranging from a simple warning to temporary suspension of payments or definitive removal (following a progression of non-compliance), and take into consideration the reasons for non-compliance.

The more flexible approach to the implementation of conditions in Bolsa Familia reflects concerns that some families with a greater likelihood of non-compliance may be more economically vulnerable (and harmed by a financial penalty), and that weaknesses in infrastructure, such as resources and staff for meeting demand for education and health services (as well as in the administrative and financial capacities for managing the program), may limit the support families receive in attempting to meet the conditions. Prospera (in Mexico) likewise applies a multi-stage approach to fines or sanctions, with suspension of payments as a first step, indefinite suspension with the option of re-admittance as a second step, followed by permanent suspension. Other programs also allow exceptions or exemptions to the conditions and sanctions they impose, such as forgiving absences on grounds of illness, or in the case of Jamaica, granting waivers from attendance requirements for disabled children (Fiszbein & Shady, 2009; Mont, 2006). In contrast, the Chile Solidario program does not begin paying cash transfers until families have complied with the first criterion, and noncompliance results in an immediate termination of the transfers (Palma & Urzúa, 2005).

Somewhat distinct from cash transfer programs with a continuum of hard to soft conditions is the LCT, where the cash transfer is distributed to households with a "nudge" or "label" indicating its intended use, in contrast to a monetary carrot or stick to ensure compliance with specified uses (Benhassine et al., 2013). For example, if an LCT is to be spent exclusively on more nutritious food, program administrators would convey this through "loose guidance" to recipients when the cash transfer is received. Like Baird et al.'s first category (conditions with no enforcement or monitoring), no monitoring takes place to determine whether the recipients are following the guidance on how the money is to be spent. Benhassine et al.'s (2013) evaluation of the Tayssir (pilot) cash transfer program in rural Morocco compared a CCT version of the program with an LCT arm that portrayed the cash transfers as an educational intervention. Monitoring of the enrollment of children ages 6-15 years was conducted at schools by headmasters, with receipt of the cash transfers tied to reductions in school absences, albeit without formal requirements for attendance or enrollment. Both the CCT and LCT had two variants: in one, the cash was transferred to the father, and in the other, the cash transfer went to the mother. More than 320 school sectors (with at least two communities in each) were randomly assigned to either a control group or one of these four program variants.

Benhassine et al.'s (2013) analysis of over 44,000 children in more than 4,000 households found significant impacts of the Tayssir cash transfers on school participation for each program variant they tested, and that these impacts did not differ significantly between the CCT and LCT. Interestingly, they also saw little difference between the LCT and CCT in how the program's intended uses were perceived, and parents' beliefs about the returns to education increased in both the LCT and CCT treatment arms. Benhassine et al. (2013) suggested that this is consistent with parents interpreting the intervention as a pro-education government program, regardless of whether they formally required regular school participation (through conditioning). They also found that dropouts related to the "child not wanting to attend school" and to "poor school quality" declined significantly in the LCT and CCT.

Similarly, Baird et al. (2013) found in their analysis–including 26 CCTs, five UCTs, and four studies that compared CCTs to UCTs–that both CCTs and UCTs significantly increased school enrollment, with the odds of a child being enrolled in school 41 percent higher in the CCTs and 23 percent higher in the UCTs (compared to no cash transfers). These differences in effects between the CCTs and UCTs were not statistically significant. However, they also compared cash transfer program effects across the three categories that included the middle design alternative (some schooling conditions with no enforcement or monitoring). When distinguishing between whether or not the schooling conditions were monitored and enforced, they did find that programs where the conditions were monitored and enforced had significantly higher odds of increasing children's enrollment than those with no conditions. At the same time, their own randomized controlled trial comparing conditional vs. unconditional cash transfers in Malawi (Baird et al., 2011) found that the largest effects of cash transfers on teenage pregnancy and marriage rates were among adolescent girls who had dropped out of school but continued to receive *un*conditional cash transfers; there were no statistically significant effects in the CCT arm of the experiment on teenage fertility or marriage. More generally, the implementation of program conditions (i.e., intensity of conditions) was the only measured design feature of the 35 cash transfer programs that significantly moderated the overall effect sizes of the programs.

We expand on this research in our analysis of the Kenya CT-OVC program, in which cash transfers were explicitly earmarked or "labeled" for spending on education and healthcare for orphans and vulnerable children in the household, but conditions with monitoring and penalties for noncompliance were assigned randomly to some districts and a sub-location within the treatment group (Hurrell, Ward & Merttens, 2008). While as noted above, there are many studies in the literature assessing outcomes of CCTs and a few comparing CCTs and UCTs, the Benhassine et al. study is the only other we are aware of that employed a random assignment design to compare the outcomes between an LCT and CCT program[1]. In addition, the Benhassine et al. study focused on rural areas and school-aged children, with program conditions based only on school absences, whereas the Kenya CT-OVC program covered infants and preschool-aged children as well and included more geographic variation and a wider set of program expectations or conditions (i.e., program rules). Like Benhassine et al., we use detailed information on cash transfer recipients' understanding of the program rules, guidance, and consequences of failure to comply with conditions to understand the extent to which the imposition of conditions with penalties (vs. labeling *only* of cash transfers) influenced household responses and program outcomes. Based on existing research evidence (discussed above), we expect the costs of the CCT monetary penalties to be felt most immediately in terms of household consumption. Thus, our comparison of the CCT and LCT treatment arms focuses on households' total, food and non-food consumption, as well as the health and education outcomes conditioned upon by the program.

3.3. PROGRAM BACKGROUND, STUDY DESIGN, DATA AND MEASURES

[1] In some research publications on the Kenya CT-OVC, the program is described as a UCT or "social cash transfer" program, while at the same time acknowledging that it involves "social messaging" (Asfaw et al., 2014, p. 1175).

The CT-OVC program is the Kenyan government's primary intervention for social protection. The program provides a flat transfer equal to approximately 20 USD per month (in 2007 dollars, exchange rate: US$1: KSh 75) that is paid bi-monthly to the caregiver for the care and support of orphans and vulnerable children (OVCs) in the household (Handa et al., 2014). In terms of the average (per adult equivalent) consumption levels at baseline (2007), the monthly cash transfers represent about 23 percent of average monthly consumption. The CT-OVC began as a pilot program in 2004, and following a three-year demonstration period, the government formally approved its integration into the national budget and began rapidly expanding the program in 2007. By the end of the impact evaluation in 2011, the CT-OVC program was providing cash transfers to more than 130,000 households and 250,000 OVCs, with the aim to scale up coverage to 300,000 households (900,000 OVCs). As of fiscal year 2015-2016, approximately 246,000 households and nearly half a million children were benefitting from the cash transfer[2].

We use data from an experimental evaluation of the Kenya CT-OVC program, mandated by the Government of Kenya, Department of Children's Services (in the Ministry of Gender, Children and Social Development), and undertaken by Oxford Policy Management with financial assistance from UNICEF. The baseline quantitative survey was conducted between March and August 2007 using questionnaires in Swahili, Luo and Somali, and follow-up surveys were administered in 2009 and 2011. The surveys collected information on household consumption expenditures, education and employment of adults, assets owned, housing conditions and other socio-economic characteristics, as well as information on child welfare measures such as anthropometric status, immunizations, illness, health-care seeking behavior, school enrollment and attendance, child work and birth registration. As many of the outcome indicators of interest for the

[2] See the Kenyan government website: https://www.socialprotection.or.ke/social-protection-components/social-assistance/national-safety-net-program/cash-transfer-for-orphans-and-vulnerable-children-ct-ovc.

children are only available in the 2007 and 2009 data collections, we restrict our analysis to these two years. A total of 2,759 households were included in the 2007 baseline sample, and of these, 2,255 were interviewed at follow-up in 2009. As Handa et al. (2014) explain, the 17 percent attrition between baseline and the first follow-up was concentrated in Kisumu and Nairobi, where the turmoil of the disputed national elections in December 2007 caused the most unrest.

The evaluation of the Kenya CT-OVC was designed as a clustered randomized controlled trial (RCT) and took place in seven districts in the country (see Figure 1 that illustrates the design)[3]. Within each of the seven districts, two sub-locations out of four were randomly assigned to be treatment locations and two were randomly assigned to the control state (no cash transfer distribution). Households in treatment locations were eligible to receive cash transfers if at least one OVC resided in them, they met the designated poverty criteria, and the OVC(s) were not benefitting from any other cash transfer program. In treatment locations, a list was compiled containing the households eligible to receive the cash transfer, and households on the list were reportedly prioritized for treatment by several "vulnerability" criteria (Hurrell, Ward & Merttens, 2008). These included the age of the caretakers of the OVCs, and the number of OVCs and chronically ill living in the household (in that order). Thus, within treatment locations, there was an intent to prioritize more "vulnerable" households for cash transfer receipt. We include these three prioritization criteria in all regressions to account for this selection. However, it is important to note that since our study focuses on comparing the two treatment arms to *one another*, this prioritization of vulnerable households into the treatment group has no effect on our main results.

[3] During the time of the CT-OVC evaluation and prior to the new constitution in Kenya that became effective in 2013, Kenya was divided into eight provinces, which were further subdivided into 46 districts (excluding Nairobi) and are today recognized as semi-autonomous counties.

In every treatment location, beneficiary households were expected to comply with program guidance or expectations for how the cash transfers would be used. These included visits to health facilities for immunizations, growth monitoring and nutrition supplements, school enrollment and attendance, and caregiver "awareness" session (see Appendix A, Table A.1), although attendance requirements were waived for children deemed to be without access to schools or clinics (Government of Kenya, 2006). In half of these locations—all treatment locations in Homa Bay, Kisumu and Kwale districts and one sub-location in Nairobi (Kirigu)—households were randomly assigned to the CCT treatment arm, where the expected penalty for not following the program conditions was a deduction of KSh 500 from the transfer amount per infraction, and multiple infractions could result in ejection from the program. Treatment locations in the other districts and one sub-location—Garissa, Migori, Suba and the other Nairobi location (Dandora B)—were assigned to the labeling only (LCT) arm where non-compliance was not supposed to be penalized.

Centrally, the CT-OVC program was coordinated through the Department of Children's Services in the Ministry of Gender, Children and Social Development (MGCSD), but its implementation and monitoring was managed locally through District Children's Offices (DCO). The DCO, in turn, collaborated with committees of voluntary members, typically composed of community leaders. These "Beneficiary Welfare Committees (BWCs)" were charged with the responsibilities of general program operations, including promoting awareness, monitoring and supporting implementation, and addressing grievances. As a labeled cash transfer program, it was intended that all beneficiaries would be made aware of the expectations that the cash transfers should be spent on visits to health facilities and expenses associated with children's enrollment and attendance in school. In fact, the final operational and impact evaluation report (Ward et al., 2010) indicated that 84 percent of cash transfer recipients believed that they had to follow some sort of rules to continue receiving the cash transfers, although the report also noted that most beneficiaries were not aware of the full set of conditions with which they were expected to comply.

Qualitative research on the program's implementation revealed that largely because of the decentralized nature of administration and reliance on volunteers for its execution, monitoring of the conditions (and enforcement of the penalties in the CCT arm) lacked structure and was uneven across and within locations (FAO, 2014). In addition, monitoring and enforcement were hindered by onerous forms and logistical challenges. The community representatives responsible for communicating and checking on conditions were often informally appointed, and implementation of that role was highly dependent on a given community representative's knowledge, interpretation of their obligations, and activism. Two years after baseline, many beneficiaries in the CCT arm had not been reached with communications about the penalties (Ward et al., 2010; FAO, 2014). The literature on CCTs suggests that these types of challenges in implementing conditions are relatively common, and that they can delay actions to sanction noncompliance, which can weaken the "positive quid pro quo" effects of the conditions on program outcomes (Fiszbein et al., 2009).

3.3.1 Measures of treatment implementation

Following the baseline data collection and implementation of the cash transfer program, household surveys were conducted in 2009 to assess the receipt of cash transfers and how households used them. For all households that received the transfer, household members were asked about their perceptions of any conditions or obligations they faced in receiving the cash transfers and about any consequences they faced for noncompliance, as well as how they used the cash transfers. In addition, the household members were asked if they "have to follow any rules in order to continue receiving the program," and they were prompted to list the rules that they thought they had to follow “in order to receive the full payment from the OVC program.” Furthermore, household members were asked if they knew which members of the household the rules applied to, if they knew what would happen if they did not follow the rules, and if they believed that anyone was checking on the conditions.

In regard to identifying the penalties that were applied in association with the conditional treatment arm, the 2009 household survey asked respondents if they had ever gone to the Post Office to collect their payment and “received less than 3000KSh for the payment cycle[4] ”. The interviewer was instructed to look at all of the receipts the respondent provided and to identify cash transfer amounts of less than KSh 3000 to determine if a monetary penalty had been applied. Household respondents identified as having been fined were also asked if they knew why the payment was less than the full amount, and if they were aware of an appeal/complaints process they could pursue if they received less than 3000 KSh in a payment cycle. Appendix A, Table A.2 and A.3 shows the survey questions that were used in constructing measures of program perceptions and implementation.

Because the implementation of conditions in the CCT arm was intended to impose concrete expectations for how households would spend the cash transfers and penalties for violations thereof, we hypothesize that households in districts and sub-locations randomly assigned to the conditional arm might differ in their perceptions, responses to, and uses of the cash transfer from those randomly assigned to the labeling only arm. Furthermore, because it is well-documented that taking a “hard line” on compliance with CCT conditions is likely to impose higher costs on the poorest and most vulnerable among those targeted for cash transfers—who, because of their greater need, also have less budgetary capacity to absorb the monetary loss—we expect there may

[4] Payment cycles were two months in length. Since households were to receive 1500 KSh per month (if no fines had been applied), this translates to a transfer of 3000 Ksh each cycle.

be differential consequences of being penalized or fined for noncompliance by household baseline wealth.

3.3.2. Outcome measures

We evaluate the difference between CCT and LCT arms in the Kenya CT-OVC program on the following dimensions of household and child wellbeing: consumption (food and non-food), health, i.e., vaccinations (total doses and sequences completed) and receipt of vitamin A supplements, and schooling (enrollment and absences from school). Most of these outcomes are linked with the program conditions shown in Appendix A Table A.1, which are intended to promote children's nutrition, growth and immunizations through increased consumption and health facility visits and their enrollment and attendance of school. We include consumption outcomes in our analysis as proxies for overall household well-being and wealth[5]. The sample sizes in our regressions vary by outcome, primarily because the outcomes we focus on are measured for distinct groups receiving the cash transfers: households for consumption, children 0-7 years for health outcomes, and school-aged children (6-17 years) for education outcomes.

We follow the Kenya CT-OVC Evaluation Team (2012) in adjusting consumption (reported at baseline in 2007) for household adult equivalents; children under age 15 were counted as three-quarters of an adult, and individuals aged 15 and over were counted as one adult. Consumption measured at follow-up (in 2009) was deflated to 2007 Kenya Shillings (KSh), following Ward et al. (2010), with separate price deflators for food and non-food items. These price adjustments were critical, given that the Kenyan post-election violence and world food crisis

[5] Deaton and Zaidi (2002) consider consumption data to be the "gold standard" for proxying wealth for several reasons. First, since consumption is presumed to be smoothed for households over periods of time, it provides a more accurate measure of wealth than income in short reference periods. Second, levels of income are often more difficult to assess in developing countries due to self- and informal sector employment.

that occurred between baseline and follow-up each engendered upward pressures on the relative price of food and increased poverty among the beneficiary population as a whole (Kenya CT-OVC Evaluation Team, 2012). Household expenditures (by broad household item groups) were combined into three main categories for our analysis: total household consumption, food consumption, and non-food consumption. Analyses by the Kenya CT-OVC Evaluation Team showed that none of the nine separate categories of household (food and non-food) expenditures were significantly different at baseline between CT-OVC treatment and control households, in spending levels, shares, or proportion of households reporting positive spending.

Children in the Kenya CT-OVC program (LCT and CCT arms) were expected to visit a health facility every two months and to receive vaccinations, vitamin A supplements and growth monitoring. According to the final operational and impact evaluation report, children 0-7 years were considered fully vaccinated if they had received (at a minimum) the following vaccinations: three DPT (diphtheria, pertussis, and tetanus) doses, three oral polio (OPV) doses, one BCG (bacille Calmette-Guerin, a vaccine for tuberculosis) and one measles (Ward et al., 2010). The household survey inquired about four OPV doses, which is recommended by the World Health Organization, thus, we consider an OPV sequence complete if four doses were received. The outcome measures we constructed to assess the impact of receiving a LCT or CCT on children's vaccinations included the total number of doses received (of all vaccinations recommended) and the number of vaccine sequences completed. For vitamin A supplements, the household survey recorded whether the child had received the supplement from a health worker within the last 6 months.

The third primary outcome we investigate, school attendance, was one element of the Kenya CT-OVC program's explicit goal to increase schooling (enrollment, attendance and retention) of children aged 6 to 17 years. At baseline (2007), about 95 percent of children aged 6-17 years in both treated and control households were enrolled in school, and the final impact evaluation report (Ward et al., 2010) did not find statistically significant impacts of the cash transfers on enrollment or attendance of *basic* schooling (although it did report statistically significant increases of 6-7 percentage points in enrollment in *secondary* schooling). The baseline (2007) data also show that children in our sample missed an average of 1.5 days of school in last month, and 10 percent of these children missed over five days in one month. We therefore focus our analysis on school attendance, which we measure as days missed from school during the school year (in 2007 and 2009). The education literature has also increasingly looked to attendance as a more informative measure of children's progress in schooling. Attendance rates have been linked to the development of important sociobehavioral skills such as motivation and self-discipline (Gernshenson, 2016; Heckman, Stixrud & Urzua, 2006) and to improved cognitive development (Gottfried, 2009), as well as to retention rates and increased educational attainment (Gershenson et al., 2017; Nield & Balfanz, 2006; Rumberger & Thomas, 2000). In addition, existing research finds that the harm of absences, in terms of reduced academic achievement, is greater among low-income students (Gershenson et al., 2017; Gottfried, 2011), and that non-school factors, such as poverty, family emergencies and work obligations, are the primary determinants of attendance rates (Balfanz & Byrnes, 2012). If being fined reduces resources for poor families that enable them to overcome these non-school barriers to school attendance, we would expect assignment to the CCT arm to diminish the cash transfer program's impact on reducing student absences compared to the LCT arm.

3.3.3. Balance checks and attrition

To estimate the unbiased difference between the CCT and LCT treatment arms, we make the identifying assumption that assignment to either arm was independent of potential outcomes. To phrase this another way, we are assuming that randomization produced two statistically equivalent groups at the onset of the experiment. We verify that randomization was successfully implemented through a series of balance tests below. Furthermore, we also check for differential attrition by treatment status to verify that our results are not driven by changes in sample composition.

One methodological challenge to evaluating these data is the small number of randomization clusters in the experimental design. As described in Figure 1, the districts Homa Bay, Kisumu, and Kwale and the sub-location Kirigu (in Nairobi district) were randomly assigned to administer a CCT to their transfer households. The remaining districts (Garissa, Migori, and Suba) and transfer sub-location in Nairobi (Dandora B) were assigned to the LCT. This produces eight randomization clusters in total. The traditional formula for consistently estimating clustered standard errors relies on the assumption that the number of clusters is sufficiently large to approximate asymptotic results, the minimum for which is 30-50 clusters. However, multiple methods now exist to produce consistent clustered standard errors when clusters are fewer than 30. The first is the wild cluster bootstrap, which, in Cameron, Gelbach, and Miller (2008), is shown to produce standard errors that are robust when the number of clusters is as few as six (as long as Webb weights are used). The second method is randomization inference. The main advantage to randomization inference in our context is that it allows us to conduct valid hypothesis tests even in the presence of small sample sizes, regardless of error structure (Young, 2019). Another advantage is that randomization inference acts as its own "placebo test". As the method consists of correlating "placebo" treatments with outcome values from the actual experiment, it verifies that treatment effects do not exist when they, in fact, should not (i.e., experimental outcomes are uncorrelated with re-randomized treatments). We report p-values produced by both of these methods in our analyses.

3.3.3.1. CCT versus LCT balance and attrition

In this section, we test our identifying assumption by assessing the comparability of the CCT and LCT treatment arms at baseline[6]. Accordingly, we present in Table 1 the results of our balance tests for the two arms by estimating equation (1) on the sample of households assigned to receive the cash transfer. Here, x_{ijk} refers to a baseline characteristic of household i. CCT_{jk} is a binary variable indicating if district k or sub-location j was assigned to the CCT (versus the LCT). We include $carerIndex_{ijk}$, $totalOVC_{ijk}$, and $totalChronicallyIll_{ijk}$ to adjust for the transfer prioritization criteria. We also test if assignment to either treatment arm is predicted jointly by a vector of baseline characteristics, X_{ijk}, by conducting an F-test for joint orthogonality using the wild cluster bootstrap after estimating equation (2), where X_{ijk} contains all variables in Table 1 except for those with multicollinearity issues[7]. When running the joint test, replace missing observations of the regressors in X_{ijk} with the variables' sample means. Additionally, we include a dummy for each regressor in X_{ijk} that equals 1 when the observation is missing and 0 otherwise. These dummies for specified as D_{ijk} in equation (2). The results in Table 1 do not indicate any statistically significant differences between households in the LCT and CCT arms across all t-test and the F-test, implying that the randomization within the transfer (treated) group was successfully executed.

[6] The household characteristics we test are based on the balance test in Annex F of Ward et al. (2010).
[7] We exclude HH Owns Livestock, HH Food Consumption, and People Aged 0-5 in HH from X_{ijk} to avoid perfect multicollinearity.

$$x_{ijk} = \propto + \delta CCT_{jk} + \gamma_1 carerIndex_{ijk} + \gamma_2 totalOVC_{ijk} + \gamma_3 totalChronicallyIll_{ijk} + e_{ijk} \quad (1)$$

$$CCT_{ijk} = \propto + X'_{ijk}\beta_1 + D'_{ijk}\beta_2 + \gamma_1 carerIndex_{ijk}$$
$$+\gamma_2 totalOVC_{ijk} + \gamma_3 totalChronicallyIll_{ijk} + e_{ijk} \quad (2)$$

The existing evidence base suggests that imposing conditions on cash transfer receipt, accompanied by fines, could potentially change household responses to and use of the cash transfers. The literature also suggests that we should pay special attention to the heterogeneous effects of cash transfer receipt by baseline levels of household wealth (proxied by per adult-equivalent consumption in our study) (de Brauw and Hoddinott, 2011; Rodríguez-Castelán, 2017). Since we are interested in whether differences in outcomes between treatment arms vary by baseline household consumption, we must also show that the treatment arms are balanced on baseline characteristics across the consumption distribution. We do this by grouping households into bins by quintile of baseline consumption, estimating equation (1) separately within each bin, and conducting inference using both the wild cluster bootstrap and randomization inference. The full results from this analysis are available upon request. Testing the 26 baseline characteristics from Table 1 within each of 5 quintile groups results in 26 × 5 = 130 estimated differences and 130 × 2 = 260 separate hypothesis tests. Of these 260 hypothesis tests, only 5 result in p-values less than 0.05[8]. Since this number of significant differences is no greater than what one would expect from chance, this provides more evidence that the treatment arms are also balanced at baseline across the consumption distribution.

Attrition would also be a concern in estimating the effects of the CCT versus LCT treatment arms of the program if the likelihood of attritting varied by CCT vs. LCT status. This would imply

[8] The wild cluster bootstrap produces two p-values less than 0.05: in the 40th to 60th percentile bin for "Years of Edu. of HH Head" and in the 80th to 100th percentile bin for "HH Receives Outside Transfer". Randomization inference produces three p-values less than 0.05: in the 80th to 100th percentile bins for "Poor Quality Floors" and "Rural", and in the 40th to 60th percentile bin for "HH Food Consumption".

that the estimated parameter of interest would represent not only the effect of the treatment, but also differences in sample composition induced by treatment arm assignment. In our sample, attrition within the transfer group is about 19 percent. In Table B.2 Panel B, we report the results of a test to determine if households assigned to the CCT arm within the transfer group experienced a differential rate of attrition compared to the LCT arm. Differential attrition is low between the two treatment arms, at only slightly more than 3 percent. This difference is statistically insignificant according to both the wild cluster bootstrap and randomization inference. We also split the sample into bins by baseline consumption quintile, as we do when checking for balance, and test for differential attrition within each bin. None of the coefficients are significant at the 5% level for any of the bins, which leads us to conclude that attrition should not distort our comparison of outcomes between the CCT and LCT groups.

3.4. RESULTS: CCT VERSUS LCT IMPLEMENTATION AND IMPACTS

We are primarily interested in how assignment to the conditional arm (versus labeling only) of the Kenya CT-OVC program affected household and children's outcomes, as well as how the effects varied based on the households' baseline wealth. In Appendix B, we present an analysis of how assignment to receive cash transfers in the CT-OVC program affected outcomes as a whole, comparing outcomes of households in sub-locations randomized to receive the transfer (CCT and LCT arms pooled together) to the outcomes of households in sub-locations randomized to the control group. Consistent with the findings of the Kenya CT-OVC Evaluation Team (2012), we found that cash transfer receipt increased both food and non-food consumption in households, although the only conditioned-upon outcome that was affected by the CT-OVC program was school attendance conditional on enrollment. We keep these results in mind as we compare program outcomes across the CCT and LCT treatment arms by estimating equation (3). y_{lijk} represents the outcome of interest and which may vary at the level of the child l or the household i. X'_{1ijk} is a vector of household characteristics at baseline, which include: the gender of the household head, whether someone in the household earns wages from an outside job, total consumption, an indicator for owning livestock, acres of agricultural land owned, an indicator for being in a rural location, the number of households members, and the baseline level of the outcome (if the outcome varies at the household level). X'_{2ijk} is a vector that contains child-level controls, consisting of the child's age, gender, OVC status, and the baseline level of individual-varying outcomes. The vectors of controls also include a dummy that equals 1 if the baseline value of the outcome is missing and 0 otherwise[9]. This dummy could vary at the household- or individual-level, depending on the outcome.

$$
\begin{aligned}
y_{lijk} = \propto + \delta CCT_{jk} + \gamma_1 carerIndex_{ijk} + \gamma_2 totalOVC_{ijk} \\
+ \gamma_3 totalChronicallyIll_{ijk} + X'_{1ijk}\beta_1 + X'_{2ijk}\beta_2 + e_{lijk} \qquad (3)
\end{aligned}
$$

3.4.1. Enforcement and salience of conditions

We now show that the conditions and penalties were meaningfully implemented on the ground and that households were indeed at risk for being penalized. The estimation sample is the group of households assigned to receive the transfer (in either the CCT or LCT arm). The outcome is set as an indicator for whether the households reported ever receiving less than their full transfer amount for at least one payment cycle by the time of the follow-up survey (two years later). We view this as an important test of the first stage that assignment to the CCT group was a meaningful

[9] In order to retain observations with missing baseline values of the outcome, we employ a method described in McKenzie (2012). This method entails coding the missing values of baseline outcome variables as 0, and adding the dummy described in the text to the specification.

treatment for households. Table 2 contains the results of estimating equation (3) with and without controls. Assignment to the CCT group increased the likelihood that a household was ever fined by about 34 percentage points (with and without controls), compared to a control mean of about 0.8 percent. The control mean is not zero because it appears as though a few households in the LCT arm were either fined by mistake, or misreported that they had experienced a transfer deduction. We interpret these results as evidence of a strong first stage, which in our context means that assignment to the CCT substantially increased households' likelihood of ever being fined. The results also provide some assurance that the survey data on fining do not suffer from substantial error or overreporting (particularly for the LCT arm). Lastly, the results imply that CCT households received less money in transfers overall than the LCT households due to the imposition of penalty fines. In section 3.4.3.2, we will show that the magnitude of this effect varied by baseline household wealth and subsequently impacted downstream outcomes.

3.4.2. Perceptions of conditions and penalties

Within the transfer group, assignment to the CCT arm may have affected outcomes through two primary channels. The first is in how the penalty fines (deductions to transfers) directly reduced household income. The second is in how the potential for penalties (associated with conditions) might have affected household decision-making. Assignment to the CCT arm (versus the LCT arm) was only likely to have affected household decisions if it produced a different understanding of program rules and consequences (potential penalties) between treatment arms. In the Kenya CT-OVC, over one third of households in the CCT arm were fined at some point, which stands in contrast to the Tayssir program, in which CCT households rarely received penalty fines due to high rates of compliance with the program's single condition. Benhassine et al. (2015) also found that understanding of program conditions was somewhat poor among households overall and that the differences in knowledge between LCT and CCT groups were small, which the authors attributed to the infrequency of penalties. Furthermore, the Tayssir program's transfer was a smaller percentage of mean baseline consumption than the Kenya CT-OVC (5 percent versus 23 percent). Together, these factors suggest that a stronger feedback loop or greater incentive for CCT households to internalize the conditions in the Kenya CT-OVC relative to Tayssir may have been present. We can investigate this, and the degree to which perceptions of program rules and consequences differed across treatment arms, using the large battery of survey questions available in the Kenya CT-OVC evaluation of households' perceptions of conditions.

We once again estimate equation (3), but set the outcome variable as an indicator that equals 1 if the household purports to understand or perceive that a particular rule or operational detail of the program applied to them. Table 3 contains the results from these regressions. The first observation is that labeling alone leads over 73 percent of households to believe that they needed to comply with rules to receive the transfer. Households assigned to the CCT treatment arm were 13 percentage points more likely to believe this, although the difference is statistically insignificant. Households also did not differ significantly in their beliefs about the specific rules (or conditions) that they perceived they had to follow to continue receiving the cash transfers (enrollment/attendance in school, health facility visits, attendance at program awareness sessions)[10]. In the last row of Panel A in Table 3, we show the results for a summary measure or "index of program understanding" that we created by adding together the five dummy variables for the specific rules households were expected to follow. The treatment effect for this index is

[10] If a household answered that they did not have to follow rules to receive the transfer, it was a "logical skip" in the survey that they did not have to answer questions about specific rules. Thus, we coded these households as not believing they needed to follow any of the specific rules.

small in magnitude and statistically insignificant. This suggests that assignment to the CCT arm did not affect the likelihood that households believed they had to follow rules to receive the transfer *and* to know what those specific rules were.

If households in both arms understood the program perfectly, we would have expected a large difference between treatment arms in their beliefs about having to follow these rules. This does not appear to have been the case. In fact, general understanding of the rules appears to have been low across both treatment arms, consistent with the pattern observed by Benhassine et al. (2015) in the Tayssir program, despite the much larger transfer amount and higher risk of being penalized in the Kenya CT-OVC program. Moreover, households in both the LCT and CCT arms of the Kenya CT-OVC program believed that they could be disbarred from the program if they did not follow the rules or guidance (see Panel B of Table 3), which did not apply to the LCT households. Taking our results with those in Benhassine et al. (2015), it appears as though households have a difficult time distinguishing labeling or guidance from conditions with penalties when program rules are explained to them.

At the same time, we do observe a statistically significant difference between treatment arms in household perceptions about program penalties. As shown in Panel B, CCT households were 17 percentage points more likely to believe they would receive a monetary fine on their transfer for each violation of the perceived rules. This finding is of note for two reasons. First, it provides additional (first stage) evidence that the conditions in the CCT arm were implemented successfully. Second, it indicates that "understanding of the penalties" might be the primary margin on which assignment to the CCT arm influences households' perceptions of program rules differentially from labeling (the LCT arm).

It is also possible that if the CCT group were more likely to believe they could penalized than the LCT group, they may have been more likely to act on their perceptions of the rules and penalties. The results of our exploration of this possibility are reported in Panel C of Table 3, which shows that there was no statistically significant difference between treatment arms in households' beliefs that someone was monitoring them. Lastly, Panel D considers what households believed about the rules or criteria for being ejected entirely from the program. Very few households in either treatment arm knew what the particular criteria were for total disbarment, and rates of understanding did not vary by treatment arm. Because the ejection criteria were more complex than the basic program rules, and ejection appears to have been a rare occurrence, this is not a surprising result. Overall, we conclude that households in both the CCT and LCT treatment arms appear to have had similar perceptions about the program rules and consequences, with the one exception being the beliefs of CCT households regarding monetary fines for violations of program rules.

3.4.3. *Program compliance, outcomes, and heterogeneous effects*

We have shown above that CCT households were no more likely than LCT households to believe they had to follow rules to receive the cash transfer or to believe that they were being monitored. These facts, combined with the observation that one third of CCT households were fined at least once, leads us to hypothesize that CCT households should have experienced similar, if not worse, downstream outcomes than LCT households. We further hypothesize that relatively poorer households (proxied by consumption) in the CCT group should have experienced the worst outcomes relative to similar households in the LCT group. This stands in contrast with the findings of Benhassine et al. (2015), in which very few CCT households in the Tayssir program ever had their transfers penalized, and downstream outcomes between the LCT and CCT did not differ significantly. We explore the average effects of assignment to the CCT in section 3.4.3.1. Then, in section 3.4.3.2, we explore the heterogeneous incidence of being fined across the household consumption distribution and its consequences for outcomes of interest, particularly consumption. Lastly, we draw upon the 2011 wave of the survey to provide some limited evidence on the longer-run differences in outcomes between the CCT and LCT arms in section 3.4.3.3.

3.4.3.1. Average effects of assignment to CCT versus LCT

We now assess whether, on average, random assignment to the CCT versus LCT arm had any impact on the schooling, health, or consumption-related outcomes of interest described above. We do this by estimating equation (3) where the dependent variable is a child-level outcome, and present the results in Table 4.

We do not find any evidence that CCT households experienced different outcomes, on average, than LCT households. These results, and especially those for education outcomes, are consistent with the findings of Benhassine et al. (2015), despite the fines levied on CCT households in our study. It is important to note that the precision of our results do not completely rule out the presence of economically meaningful effects for some outcomes. For example, the 95% confidence interval on the point estimate for Days Missed from School is [-0.366, 0.284] according to the wild cluster bootstrap, which are 33% and 26% of the control mean, respectively. However, the overall takeaway is that even in the presence of a larger cash transfer amount than the Tayssir program, coupled with a higher probability of being penalized for noncompliance, we still cannot detect statistically significant average effects of assignment to the CCT group (versus the LCT arm) on conditioned-upon outcomes.

3.4.3.2. Heterogeneous effects of assignment to CCT versus LCT

One possible explanation for these null effects is that while CCT households may have been more motivated to comply with the conditions, the penalty fines created financial constraints that prevented them from doing so. As we suggest in section 3.4.1, a common concern about CCTs is that they may be least beneficial for vulnerable households that have trouble complying with conditions. This appeared to have been the case in Baird et al. (2011), in which girls who dropped out of school (thus breaking the conditions) suffered worse marital and fertility outcomes in the CCT group than the UCT group. Furthermore, resource-constrained households that cannot comply with conditions are also likely to be the ones who potentially benefit most from cash transfer programs. Thus, there is reason to believe that a household's likelihood of being fined, and its ability to cope with said fines, may vary according to baseline household We analyze the heterogeneous effects of assignment to the CCT arm versus the LCT arm on likelihood of being fined by estimating equation (4) which adds an interaction term between the assignment variable and baseline per capita consumption (our proxy for wealth). The main effect for baseline consumption, $cons_{ijk}$, is contained in the vector X'_{1ijk}, as described above.

$$y_{lijk} = \propto + \delta_1 CCT_{jk} + \delta_2 CCT_{jk} \times cons_{ijk} + \gamma_1 carerIndex_{ijk} + \gamma_2 totalOVC_{ijk}$$
$$+ \gamma_3 totalChronicallyIll_{ijk} + X'_{1ijk}\beta_1 + X'_{2ijk}\beta_2 + e_{lijk} \qquad (3)$$

We plot these results in Figure 2, which includes randomization inference p-values and wild cluster bootstrapped confidence intervals. According to the linear specification, assignment to the CCT arm increased relatively poorer households' likelihood of ever being fined more than other (wealthier) groups. These results suggest that CCT households' burden of fines on transfer income varied by baseline consumption on the extensive margin. These findings motivate us to re-estimate equation (4) with our downstream outcomes of interest on the left-hand side. We report these results in Table 5.

Across most outcomes and consumption percentiles, we find little difference in outcomes between CCT and LCT arms. One exception, however, is that households at and below the 25th percentile of consumption reported significantly lower non-food consumption at follow-up in the CCT arm than in the LCT arm. In particular, according to the wild cluster bootstrap, the 95% confidence intervals for these differences are [-0.309, -0.041] and [-0.308. -0.052] at the 25th and 10th consumption percentiles, respectively. It appears as though the poorest households in the CCT arm may have been substantially affected by the penalty fines, leading them to reduce their non-food (i.e., likely less essential) consumption. Note that these results withstand significance tests using randomization inference p-values, which means that they are likely not to be driven by only one or two of the CCT clusters, thereby increasing our confidence in their validity. Although we attribute this reduction in non-food consumption primarily to the penalty fines, we do not claim that they are the only mechanism at work here. Lastly, we do not want to overstate the precision of the null results in Table 5 because, like in section 3.4.3.1, the confidence intervals on the point estimates are quite wide. For example, the effect of being assigned to the CCT on Days Missed from School for children in the 10th percentile of household consumption has a point estimate of -0.129 and a 95% confidence interval of [-0.643, 0.312]. This interval contains effect sizes that range from 58% to 28% of the control mean, which are of substantial magnitude in both directions. Thus, we cannot claim that being assigned to the CCT versus the LCT had no effect on outcomes apart from non-food consumption, just that they were not statistically detectable in this analysis.

Overall, these findings imply that assignment to the CCT arm within the transfer group of the Kenya CT-OVC did not result in statistically significant improvements in the conditioned-upon outcomes. If anything, it appears as though receiving the CCT instead of the LCT resulted in reductions in non-food consumption among the poorest households, likely due to the burden (cash loss) imposed by the penalty fines. As a robustness check, we estimate equation (4) on downstream outcomes while controlling for the false discovery rate (FDR) to adjust for multiple hypothesis testing[11]. This produces q-values–or the lowest FDR that would allow us to reject the null hypothesis for a given p-value–that are below 0.05 for our main results, non-food consumption for households at the 25th percentile of consumption and below (when using p-values from the wild cluster bootstrap). We present our full re-analysis in Appendix C, accompanied by a more detailed description of the false discovery rate and q-values.

3.4.3.3. Longer-Run Effects

In addition to the follow-up survey that was conducted in 2009, a second set of follow-up data were collected in 2011 that potentially allow us to study the longer-run effects of the Kenya CT-OVC. These data and our analysis of them come with several important caveats. The first is that is that we were unable to find program documentation that addressed whether the conditionality continued to be meaningfully implemented and enforced between 2009 and 2011. Additionally, the second follow-up survey asked few questions about program implementation and did not collect data on whether households experienced penalty fines on their transfers. Moreover, since the Kenya CT-OVC evaluation team did not specify how they calculated the 2009 price deflators, we are unable to use the same methods to create an updated pair of deflators for the food and non-food components of the 2011 consumption variables. Instead, we draw upon the country-wide inflation rate of the KSh, compiled by the International Monetary Fund between 2009 and

[11] We control for the false discovery rate using the publicly available Stata code posted by Michael Anderson, as described in Anderson (2008).

2011, and use that information to deflate the 2011 values to 2007 shillings. Since this only gives us a single value for inflation, we apply it to both food and non-food consumption. Additionally, there are no 2011 data on Vitamin A usage among children, and the data on Days Missed from School are only for a two-week time window. Lastly, there was additional attrition between the 2009 and the 2011 surveys, further reducing the sample size.

With these limitations in mind, we estimate the longer-run impacts of the transfer by estimating appendix equation (B.1) with the 2011 outcome values on the left-hand side. These results are presented in Appendix D, Table D.1[12]. Most of the significant effects from before are either insignificant or marginally significant when using 2011 data. The most notable of these changes are the newly null effects on all of the consumption variables, which were previously quite robust. Next, to obtain the long run effects of being assigned to the CCT arm versus the LCT arm, we estimate equation (4) using the 2011 outcomes and display the results in Table D.2. Given the insignificant effects of the pooled transfer, it is unsurprising that there are no significant differences in outcomes between the CCT and LCT arms. This holds true when we are looking at both average and heterogeneous effects by baseline consumption. One potential explanation for this is that inflation had greatly eroded the value of the transfer. By 2011, the monthly transfer was only worth 991 KSh in 2007 shillings, about two-thirds of the real value from four years earlier. This erosion was acknowledged by the program coordinators and the transfer was increased from 1500 to 2000 KSh in the 2011-12 fiscal year. However, according to transfer receipts collected by the

[12] Since the specification includes the indicator for missing baseline values of the outcome on the right-hand side, we adjust the variable as needed when we substitute the 2011 outcomes for the 2009 outcomes.

enumerators from participants, this increase appears not to have been effective until late in the 2011 data collection period[13].

3.5. POLICY IMPLICATIONS AND CONCLUSION

In a 2013 blog post[14], Berk Ozler characterized efforts to describe or define cash transfer programs as “an unconditional mess,” arguing that the distinctions between CCTs and UCTs were “too blurry” and that interested stakeholders (donors, policymakers) would be better off thinking about them along a “continuum from a pure UCT to a heavy-handed CCT”. Our research further suggests that a particular cash transfer program, such as the Kenya CT-OVC program, may not correspond to a single point along such a continuum. Indeed, our examination of the Kenya CT-OVC program shows that where it fits along a continuum from fully unconditional to “hard” conditions may depend on the implementation of the program as experienced and understood by households. And as Ozler opined and we found in this research, there are tradeoffs for household outcomes in terms of how the conditions (or lack thereof) are implemented. Our findings show that the imposition of conditions in a CCT arm of the Kenya CT-OVC program–i.e., a “heavy-handed” implementation that monetarily penalized families for their failure to comply with program conditions–did not improve children's outcomes relative to the LCT arm and had tradeoffs for household non-food consumption that varied by baseline poverty or wealth. These findings are consistent with prior literature showing that the effects of CCTs on household consumption vary according to household baseline wealth (or depth of poverty) (Fiszbein et al., 2009; Hoddinott, Skoufias, and Washburn 2000; Macours, Schady, and Vakis, 2012; Maluccio and Flores 2005).

[13] In the 2011 wave of data collection, households were asked to hand enumerators their most receipt transfer receipts. Of the households that could supply this information, approximately 10% of them reported having received the increased transfer amount (4000 Ksh per payment cycle versus 3000 KSh).

[14] https://blogs.worldbank.org/impactevaluations/defining-conditional-cash-transfer-programs-unconditional-mess

Indeed, one of the more compelling aspects of our estimates showing that the consumption of poorer households may be harmfully reduced is that they are largely consistent with what development practitioners and researchers have long suspected (even if debate in the literature is ongoing).

Having a program where households face penalties for not complying with expectations to spend cash transfers wisely (or for the benefit of the children) is a potentially promising way to achieve the broader goals of cash transfers programs, that is, to reduce not only poverty but also the intergenerational transmission of poverty. But it also creates more administrative burdens and costs for program implementation in the monitoring of household compliance with program conditions and enforcement of penalties. Furthermore, researchers and practitioners have long been concerned about the undue burdens that conditional cash transfers also place on the poorest of poor households. Not only is complying with rules more challenging for them, but penalizing their transfers may cut them off from purchasing basic necessities that their more meager budgets barely afford. Regrettably, this is what appears to have happened in the case of the Kenya CT-OVC program. These concerns are underscored by the fact that our analysis was not able to detect significant effects of assignment to the CCT on outcomes that were conditioned-upon by the program. However, we have also acknowledged that the few randomization clusters in this experiment prevent us from completely ruling out the possibility that the CCT arm experienced *any* change in outcomes relative to the LCT arm. In fact, the confidence intervals on our estimates are sufficiently wide such that even moderately-sized effects on other outcomes could have occurred. What we can say with confidence, though, is that the negative effects on non-food consumption were sufficiently large that even an analysis with limited power (such as ours) was able to detect them.

If the insignificant differences between the CCT and LCT arms are to be taken at face value (i.e., if one overlooks their precision), then the policy implications of our results for a hypothetical cash transfer program depend on the said program's stated objectives. In a program that is purely concerned with improving conditioned-upon outcomes at the lowest cost, then the choice of CCT versus LCT hinges upon the cost of implementing and enforcing the conditions relative to the forecasted savings in transfer money withheld from households in the form of fines. If the program is concerned about overall household wellbeing (instead of primarily conditioned-upon outcomes), then those who are planning the program should also take into account how the imposition of fines on transfers reduces consumption for households at the lower end of the wealth distribution. Put another way, planners would have to weigh the costs of implementing the CCT with the money saved in the form of withheld transfer payments *and* the fines' negative effects on household consumption, which complicates the analysis. Finally, to add further complexity, these calculations change if one relaxes the assumption that assignment to the CCT arm produces the same conditioned-upon outcomes as assignment to the LCT arm. The width of our confidence intervals suggests that the effects of being assigned to the CCT on these outcomes could have been large enough to change this cost-benefit analysis substantially in either direction. Future work in this area could explore how CCTs and LCTs compare in the context of a well-powered experiment, in which the transfer constitutes a large fraction of households' baseline consumption (as in the Kenya CT-OVC). Only through further research, coupled with detailed data on implementation costs, can the relative benefits of CCTs versus LCTs be more clearly established, and even then, the conclusions may be tempered by contextual factors in implementation.

Surprisingly, given the expansive literature that has emerged over time on CCTs and UCTs (and the nascent literature on LCTs), we found little empirical exploration of the consequences of experiencing financial penalties (or suspension or termination of benefits) for households and children receiving cash transfers. The random assignment between CCT and LCT arms in the Kenya CT-OVC may have allowed us a unique opportunity to examine the consequences of financial penalties in CCTs in terms of household and children's outcomes. That said, while we believe that we have presented compelling evidence on the differential impacts of CCTs and LCTs, our study is not without limitations. As noted above, our data on penalty fines are only for a two-year window of program implementation, and we do not have detailed data to identify the frequency or timing of penalties on households at all. Ideally, we would have had better data to explore a fuller range of impacts of being fined on household and children's well-being, but we are constrained by sample sizes within the CT-OVC treatment group and by the fact that many outcomes were measured only for age-appropriate subgroups. We hope this research will spur further interest in “labeling” or other behavioral nudges in cash transfer programs that can offset the welfare costs inherent in traditional CCTs, as we observed in the Kenya CT-OVC program.

3.6. REFERENCES

Anderson, Michael L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103, no. 484: 1481-1495

Asfaw, Solomon, Benjamin Davis, Josh Dewbre, Sudhanshu Handa, and Paul Winters (2014). Cash Transfer Programme, Productive Activities and Labour Supply: Evidence from a Randomised Experiment in Kenya. *The Journal of Development Studies*, 50 (8): 1172-1196

Baird, Sarah, Francisco H. G. Ferreira, Berk Ozler, and Michael Woolcock (2013). Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systemic Review. *Campbell Systematic Reviews*.

Baird, Sarah, Craig McIntosh, and Berk Ozler (2011). Cash or Condition? Evidence from a Cash Transfer Experiment. *The Quarterly Journal of Economics* 126 (4): 1709-1753.

Balfanz, Robert and Vaughan Byrnes (2012). The importance of being in school: A report on absenteeism in the nation's public schools. Baltimore, MD: Johns Hopkins University.

Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, Tanja Schmidt, and Luca Pellerano (2016). Cash transfers: what does the evidence say? Overseas Development Institute. https://www.odi.org/sites/odi.org.uk/files/resource-documents/10749.pdf

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen (2015). Turning a Shove into a Nudge? A "Labeled Cash Transfer" for Education. *American Economic Journal: Economic Policy*, 7 (3): 86-125.

Benjamini, Yoav and Yosef Hochberg (1995), Controlling the False Discovery Rate. *Journal of the Royal Statistical Society, Ser. B*, 57 (1): 289–300.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli (2006), Adaptive Linear Step-Up Procedures That Control the False Discovery Rate. *Biometrika*,, 93 (3): 491–507.

Box, George E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, 71: 791–799.

Cameron, A.C., Jonah B. Gelbach, and Douglas L. Miller (2008). Boostrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90 (3): 414-427.

Caldes, Natalia, David Coady, and John A. Maluccio (2006). The Cost of Poverty Allevation Transfer Programs: A Comparative Analysis of Three Programs in Latin America. *World Development*, 34 (5): 818-837

Deaton, Angus and Salman Zaidi. (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis. LSMS Working Paper, No. 135. World Bank.

de Brauw, Alan and John Hoddinott (2011). Must conditional cash transfer programs be conditioned to be effective? The impact of conditioning transfers on school enrollment in Mexico. *Journal of Development Economics*, 96 (2): 359–370.

Fernald, Lia C. H., Paul J. Gertler, Lynnette M. Neufeld (2008). The Role of Cash in Conditional Cash Transfer Programmes for Child Health, Growth, and Development: An Analysis of Mexico's Oportunidades. *The Lancet*, 371 (9615): 828–837.

Ferreira, Francisco H. G. (2008). "The Economic Rationale for Conditional Cash Transfers." Unpublished manuscript, World Bank, Washington, DC.

Fisher, Eleanor, Ramlatu Attah, Valentina Barca, Clare O'Brien, Simon Brook, Jeremy Holland, Andrew Kardan, Sara Pavanello, and Pamela Pozarny (2017). The Livelihood Impacts of Cash Transfers in Sub-Saharan Africa: Beneficiary Perspectives from Six Countries. *World Development*, 99: 299-319.

Fiszbein, Ariel, Ravi Kanbur, and Ruslan Yemtsov (2014). Social protection and poverty reduction: Global patterns and some targets. *World Development*, 61, 167–177.

Fiszbein, Ariel, Norbert Schady, Francisco H. G. Ferreira, Margaret Grosh, Niall Keleher, Pedro Olinto, Emmanuel Skoufias (2009). Conditional Cash Transfers: Reducing Present and Future Poverty. Washington, DC: World Bank.

Gershenson, Seth, Alison Jacknowitz, and Andrew Brannegan. (2017). Are student absences worth the worry in U.S. primary schools? *Education Finance and Policy*, 12 (2): 137-165.

Golan, Jennifer, Terry Sicular, and Nithin Umapathi (2015). Unconditional cash transfers in China : an analysis of the rural minimum living standard guarantee program. World Bank Policy Research Working Paper WPS7374.

Gottfried, Michael A. (2009). Excused versus unexcused: How student absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis*, 31 (4): 392–419.

Government of Kenya, Office of the Vice President and Ministry of Home Affairs. (2006). Program Design, Cash Transfer Pilot Project. Nairobi.

Handa, Sudhanshu, David Seidenfeld, Benjamin Davis, and Gelson Tembo (2015). The Social and Productive Impacts of Zambia's Child Grant. *Journal of Policy Analysis and Management*, 35 (2): 357-387.

Handa, Sudhanshu, Carolyn T. Halpern, Audrey Pettifor, and Harsha Thirumurthy. (2014). The Government of Kenya's Cash Transfer Program Reduces the Risk of Sexual Debut among Young People Age 15–25. *PLoS One*, 9 (1): e85473.

Heckman, James J., Jora Stixrud, and Sergio Urzua. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics* 24 (3): 411–482.

Heinrich, Carolyn J. and Robert Brill (2015). Stopped in the Name of the Law: Administrative Burden and its Implications for Cash Transfer Program Effectiveness. *World Development*, 72: 277–295.

Hidrobo, Melissa, John Hoddinott, Nehma Kumar, Meghan Olivier (2018). Social Protection, Food Security, and Asset Formation. *World Development*, 101: 88–103.

Hoddinott, John F., Emmanuel Skoufias, and Ryan Washburn (2000). The Impact of PROGRESA on Consumption: A Final Report. International Food Policy Research Institute, Washington, DC.

Huang, Carolyn, Kavita Singh, Sudhanshu Handa, Carolyn Halpern, Audrey Pettifor, and Harsha Thirumurthy (2017). Investments in children's health and the Kenyan cash transfer for orphans and vulnerable children: evidence from an unconditional cash transfer scheme. *Health Policy and Planning*, 32 (7): 943-955.

Hurrell, Alex, Patrick Ward, Fred Merttens (2008). Kenya CT-OVC Programme Operational and Impact Evaluation Baseline Survey Report: Final Report. Oxford Policy Management, July.

Imbens, Guido W. and Jeffrey M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47 (1): 5-86.

Levy, Santiago. (2006). Progress Against Poverty: Sustaining Mexico's PROGRESA-Oportunidades Program. Washington, DC: Brookings Institution Press.

Lindert, Kathy, Anja Linder, Jason Hobbs, and Bénédicte de la Brière (2007). The Nuts and Bolts of Brazil's Bolsa Família Program: Implementing Conditional Cash Transfers in a Decentralized Context. Social Protection Discussion Paper 0709, World Bank, Washington, DC.

Macours, Karen, Norbert Schady and Renos Vakis (2012). Cash Transfers, Behavioral Changes, and the Cognitive Development of Young Children: Evidence from a Randomized Experiment. *American Economic Journal: Applied Economics*, 4 (2), 247-273.

Maluccio, John A. and Rafael Flores (2005). Impact Evaluation of a Conditional Cash Transfer: The Nicaraguan Red de Protección Social. Research Report 141, International Food Policy Research Institute, Washington, DC.

McKenzie, David (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99 (2): 210-221.

Mont, Daniel (2006). Disability in Conditional Cash Transfer Programs: Drawing on Experience in LAC. Report prepared for the Third International Conference on Conditional Cash Transfers, Istanbul, Turkey, June 26–30.

Morais de Sá e Silva, Michelle (2017). Poverty Reduction, Education, and the Global Diffusion of Conditional Cash Transfers. Palgrave Macmillan.

Nield, Ruth C., and Robert Balfanz. (2006). An extreme degree of difficulty: The educational demographics of urban neighborhood high schools. *Journal of Education for Students Placed at Risk*, 11 (2): 123–141.

Palma, Julieta and Raul Urzua (2005). Anti-poverty Policies and Citizenry: The "Chile Solidario" Experience. United Nations Educational, Scientific and Cultural Organisation, Policy Papers 12.

Ralston, Laura, Colin Andrews, and Allan Jer-Yu Hsiao (2017). The Impacts of Safety Nets in Africa What Are We Learning? Policy Research. World Bank: Social Protection and Labor Global Practice Group & Africa Region Working Paper 8255.

Rodriguez-Castelan, Carlos. 2017. Conditionality as Targeting? : Participation and Distributional Effects of Conditional Cash Transfers. Policy Research working paper,no. WPS 7940;; Policy Research Working Paper;No. 7940. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/25949 License: CC BY 3.0 IGO

Rumberger, Russell W., and Scott L. Thomas. (2000). The distribution of dropout and turnover rates among urban and suburban high schools. *Sociology of Education*, 73 (1): 39–67.

Silva, Maria Ozanira da Silva e. (2007). O Bolsa Família: problematizando questões centrais na política de transferência de renda no Brasil. *Ciência & Saúde Coletiva*, 12 (6): 1429-1439.

The Kenya CT-OVC Evaluation Team (2012). The impact of the Kenya Cash Transfer Program for Orphans and Vulnerable Children on household spending. *Journal of Development Effectiveness*, 4 (1): 9-37.

Ward, Patrick, Alex Hurrell, Aly Visram, Nils Riemenschneider, Lucas Pellerano, Clare O'Brien, Ian MacAuslan, and Jack Willis (2010). Cash Transfer Programme for Orphans and Vulnerable Children (CT-OVC), Kenya: Operational and Impact Evaluation, 2007-2009: Final report. Oxford Policy Management.

World Bank. Cash transfers. https://www.unicef.org/esaro/5483_cash_transfers.html

Young, Alwyn (2019). Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. *The Quarterly Journal of Economics*, 134 (2): 557-598.

Figure 1: RCT Design

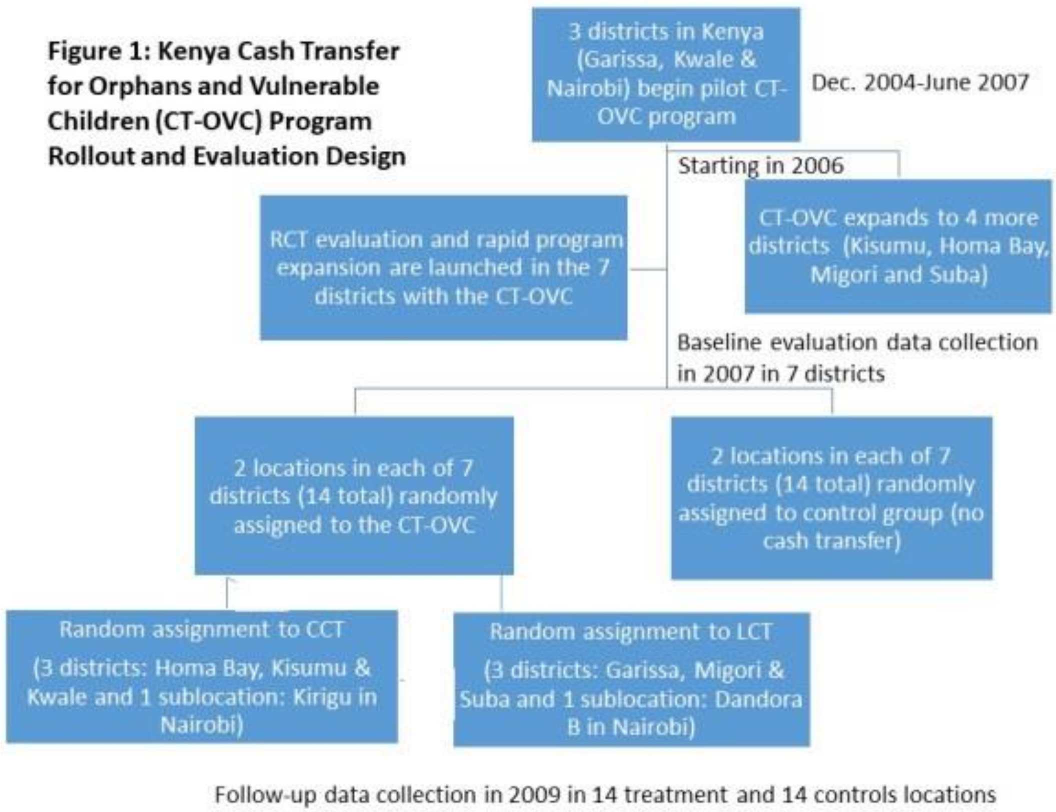
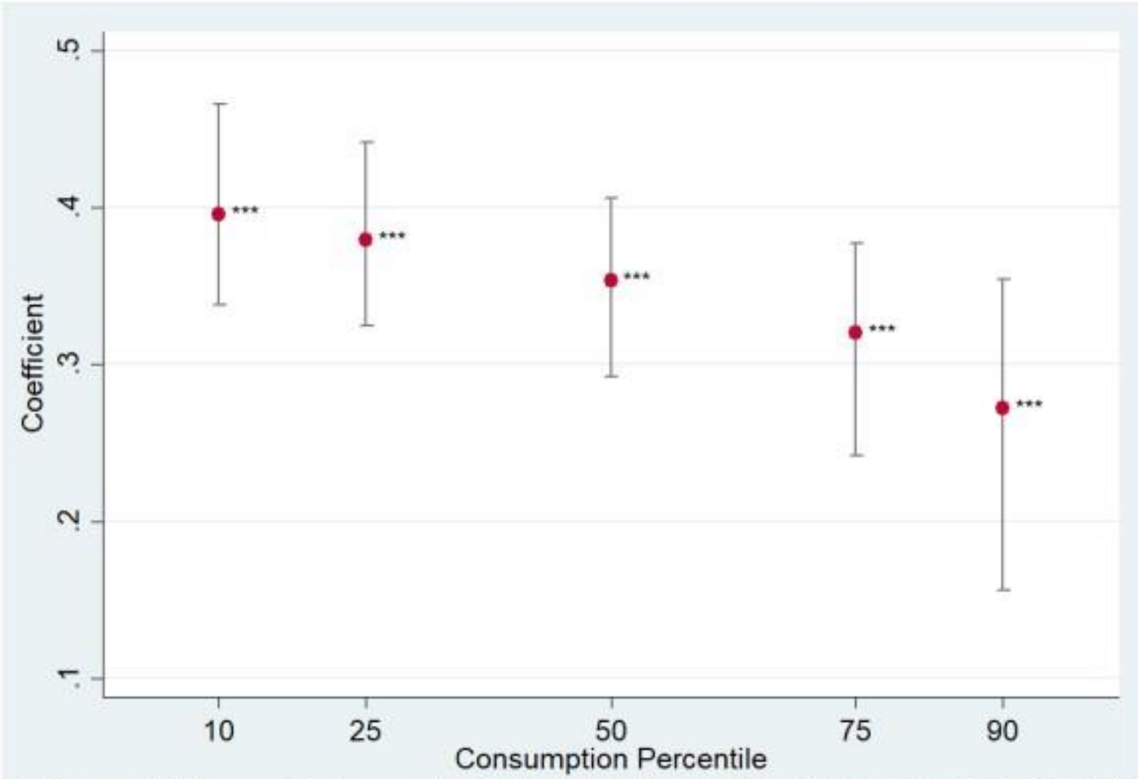


Figure 2: Impact of Assignment to CCT on Ever Being Fined by Consumption Percentile



Note: P-values calculated by randomization inference, CIs by wild cluster bootstrap, both clustered at district level. * p<.1, ** p<.05, *** p<.01. N=1090

Table 1: Balance Table: LCT vs CCT

	(1) LCT	(2) CCT	(3) Bootstrap P-Value	(4) RI P-Value
Years of Edu. of HH Head	5.788	6.034	0.235	0.249
Sex of HH Head	0.355	0.348	0.877	1.000
HH Receives Labor Wages	0.038	0.029	0.855	0.901
HH Receives Outside Transfer	0.355	0.217	0.294	0.325
Poor Quality Walls	0.683	0.781	0.655	0.843
Poor Quality Floor	0.724	0.737	0.949	0.849
HH Owns Livestock	0.831	0.758	0.389	0.240
Cattle Owned	1.186	1.432	0.428	0.406
Poultry Owned	3.899	5.123	0.251	0.427
Owns Telephone	0.105	0.106	0.612	0.833
Owns Blanket	0.823	0.865	0.900	0.822
Owns Mosquito Net	0.647	0.551	0.184	0.293
Acres of Land Owned	1.387	1.866	0.373	0.383
Household in Rural Location	0.885	0.766	0.311	0.306
HH Total Consumption	1.668	1.521	0.485	0.568
HH Food Consumption	0.972	0.926	0.611	0.659
HH Non-food Consumption	0.695	0.595	0.478	0.350
Dietary Diversity Score	4.972	5.29	0.437	0.617
Size of the HH	5.409	5.489	0.689	0.738
Age of HH Head	56.498	60.046	0.860	0.843
People Aged 0-5 in HH	0.649	0.702	0.549	0.465
People Aged 6-11 in HH	1.261	1.180	0.875	0.955
People Aged 12-17 in HH	1.325	1.418	0.161	0.231
People Aged 18-45 in HH	1.136	1.120	0.709	0.792
People Aged 46-64 in HH	0.665	0.636	0.740	0.803
People Aged 65+ in HH	0.373	0.433	0.428	0.507
N:	609	483		
F-test :	N: 1092		0.644	

Note: The p-values in column (3) are calculated with the wild cluster bootstrap procedure, and the p-values in column (4) are calculated with randomization inference. Clustering is done at the district level. All estimates are calculated conditioning on the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The F-test for Joint Orthogonality regresses CCT assignment (versus LCT) on a vector containing baseline characteristics (excluding HH Owns Livestock, HH Food Consumption, and People Aged 0-5 in HH to avoid perfect multicolinearity) and the transfer prioritization criteria. The following variables have fewer observations than reported beside "N: t-tests" due to missing responses: Years of Edu. of HH Head, Cattle Owned, Poultry Owned, and Age of HH Head.

Table 2: Impact of Assignment to CCT on Ever Being Fined

	No Controls			Controls		
	(1) Differential Effect	(2) Bootstrap P-value	(3) RI P-value	(4) Differential Effect	(5) Bootstrap P-value	(6) RI P-value
Assigned to CCT	0.341	0.002	0.000	0.344	0.002	0.000
LCT Mean	0.008			0.008		
N	1090			1090		

Note: The p-values in columns (2) and (5) are calculated with the wild cluster bootstrap procedure, and the p-values in columns (3) and (6) are calculated with randomization inference. Clustering is done at the district level. All estimates are calculated conditioning on the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent.

Table 3: Impact of Assignment to CCT on Perceptions of Conditions and Penalties

	(1) LCT Mean	(2) CCT Differential Effect	(3) Bootstrap P-value	(4) RI P-Value	(5) N
Panel A: Understanding of Program Rules					
Believes HH Must Follow Rules to Receive Payments	0.733	0.132	0.324	0.612	1086
Enrollment/Attendance in Primary or Secondary School	0.294	0.004	0.936	0.946	1092
Visit Health Facility for Immunizations	0.154	0.054	0.528	0.463	1092
Visit Health Facility for Growth Monitoring	0.090	0.057	0.134	0.296	1092
Visit Health Facility for Vitamin A Supplement	0.059	-0.003	0.905	0.889	1092
Attendance at Program Awareness Sessions	0.043	0.036	0.201	0.269	1092
Index of Program Understanding	0.640	0.148	0.246	0.428	1092
Panel B: Understanding of Penalty for Violation					
Monetary Fine on Transfer	0.048	0.170	0.009	0.000	1092
Total Disbarment from Program	0.437	-0.007	0.938	0.942	1086
Panel C: Perceived Likelihood of Being Monitored					
Believes Someone is Checking if HHs are Following Rules	0.421	0.085	0.518	0.582	876
Panel D: Understanding of Ejection Criteria					
Claims to Know Specific Criteria for Ejection from Program	0.437	-0.007	0.865	0.946	1086
HH has no OVCs Below 18 Years Old	0.146	-0.025	0.558	0.645	1092
At Least One Program Rule is Ignored for Three Consecutive Pay Periods	0.189	0.098	0.522	0.505	1092
HH Moves to Non-Program District	0.018	-0.014	0.442	0.912	1092
HH Does Not Collect Transfer for Three Consecutive Pay Periods	0.011	0.000	0.999	0.946	1092

Note: The p-values in column (3) are calculated with the wild cluster bootstrap procedure, and the p-values in column (4) are calculated with randomization inference. Clustering is done at the district level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. Note that if a household did not believe it needed to follow any rules to receive the transfer, it is also coded not to believe it had to follow the specific rules and not to believe it knew the penalty for violations. Similarly, a household that did not claim to know the criteria for ejection from the program is coded not to know the individual ejection criteria either.

Table 4: Impact of Assignment to CCT versus LCT: Average Effects

	(1) LCT Mean	(2) Assigned to CCT	(3) Bootstrap P-value	(4) Bootstrap CI	(5) RI P-Value	(6) N
Enrolled in School	0.937	0.000	0.964	[-0.017, 0.023]	0.943	2549
Days Missed from School	1.109	-0.067	0.510	[-0.366, 0.284]	0.761	2242
Total Doses of Vaccinations	7.429	-0.507	0.349	[-1.249, 0.954]	0.456	235
Number of Vacc. Sequences Completed	3.010	-0.235	0.303	[-0.707, 0.466]	0.395	235
Received Vitamin A Supplement Six Months	0.510	0.004	0.971	[-0.182, 0.364]	0.951	561
HH Total Consumption	2.107	-0.055	0.619	[-0.531, 0.227]	0.705	1092
HH Food Consumption	1.275	0.030	0.679	[-0.155, 0.180]	0.751	1093
HH Non-food Consumption	0.835	-0.083	0.175	[-0.355, 0.074]	0.251	1095

Note: The p-values in column (3) and confidence intervals in column (4) are calculated with the wild cluster bootstrap procedure. The p-values in column (5) are calculated with randomization inference. Clustering is done at the district level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The sample for both education variables consists only of children that were aged at 2 to 17 at baseline. Only children enrolled in school at baseline and follow-up are included in the sample for Days Missed from School, and the reference period is the past 2 months. Vaccinations include BCG (1 dose sequence), DPT (3 doses), OPV (4 doses), and measles (1 dose). Only children aged seven or under at follow-up are included in the sample for the vaccination variables and the vitamin A supplement, the reference period for the latter is the past 6 months.

Table 5(A): Impacts of Assignment to CCT versus LCT: Heterogeneous Effects

	(1) Enrolled in School	(2) Days Absent from School	(3) Total Doses of Vaccinations	(4) Number of Vacc. Sequences Completed
Effects by Baseline Consumption Percentile				
Percentile: 10	0.013	-0.129	-0.586	-0.289
	(0.312)	(0.507)	(0.443)	(0.387)
	[0.357]	[0.568]	[0.549]	[0.351]
Percentile: 25	0.009	-0.109	-0.561	-0.272
	(0.325)	(0.443)	(0.404)	(0.348)
	[0.367]	[0.618]	[0.501]	[0.351]
Percentile: 50	0.002	-0.076	-0.521	-0.244
	(0.782)	(0.460)	(0.341)	(0.303)
	[0.763]	[0.719]	[0.456]	[0.403]
Percentile: 75	-0.007	-0.035	-0.469	-0.209
	(0.652)	(0.792)	(0.328)	(0.317)
	[0.745]	[0.793]	[0.509]	[0.443]
Percentile: 90	-0.019	0.024	-0.395	-0.158
	(0.587)	(0.937)	(0.278)	(0.445)
	[0.560]	[0.952]	[0.550]	[0.591]
LCT Mean	0.937	1.109	7.429	3.010
N	2549	2242	235	235

Note: P-values calculated with the wild cluster bootstrap procedure are in parentheses, and p-values calculated with randomization inference are in brackets. Clustering is done at the district level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The sample for both education variables consists only of children that were aged at 2 to 17 at baseline. Only children enrolled in school at baseline and follow-up are included in the sample for Days Missed from School, and the reference period is the past 2 months. Vaccinations include BCG (1 dose sequence), DPT (3 doses), OPV (4 doses), and measles (1 dose). Only children aged seven or under at follow-up are included in the sample for the vaccination variables and the vitamin A supplement, the reference period for the latter is the past 6 months.

Table 5(B): Impacts of Assignment to CCT versus LCT: Heterogeneous Effects

	(1) Received Vitamin A Supplement	(2) HH Total Consumption	(3) HH Food Consumption	(4) HH Non-food Consumption
Effects by Baseline Consumption Percentile				
Percentile: 10	0.025	-0.215	-0.050	-0.161
	(0.870)	(0.176)	(0.526)	(0.005)
	[0.854]	[0.201]	[0.582]	[0.000]
Percentile: 25	0.018	-0.165	-0.025	-0.137
	(0.901)	(0.230)	(0.738)	(0.004)
	[0.951]	[0.251]	[0.690]	[0.042]
Percentile: 50	0.007	-0.085	0.015	-0.098
	(0.949)	(0.422)	(0.833)	(0.070)
	[0.951]	[0.600]	[0.794]	[0.204]
Percentile: 75	-0.007	0.018	0.066	-0.047
	(0.959)	(0.862)	(0.338)	(0.467)
	[1.000]	[0.900]	[0.443]	[0.508]
Percentile: 90	-0.028	0.165	0.139	0.025
	(0.840)	(0.393)	(0.141)	(0.804)
	[0.850]	[0.395]	[0.245]	[0.796]
LCT Mean	0.510	2.107	1.269	0.838
N	561	1092	1093	1095

Note: P-values calculated with the wild cluster bootstrap procedure are in parentheses, and p-values calculated with randomization inference are in brackets. Clustering is done at the district level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The sample for both education variables consists only of children that were aged at 2 to 17 at baseline. Only children enrolled in school at baseline and follow-up are included in the sample for Days Missed from School, and the reference period is the past 2 months. Vaccinations include BCG (1 dose sequence), DPT (3 doses), OPV (4 doses), and measles (1 dose). Only children aged seven or under at follow-up are included in the sample for the vaccination variables and the vitamin A supplement, the reference period for the latter is the past 6 months.

3.7 APPENDIX A: CT-OVC PROGRAM GUIDANCE AND CONDITIONS

Table A.1(A)

Kenya CT-OVC Program Conditions and Compliance Monitoring

Children aged one year and under should:

- Attend the health facility for immunizations, growth monitoring and vitamin A supplement
    - Frequency of required compliance: six times per year
    - Frequency of compliance monitoring: every two months.

Children aged between one and five years should:

- Attend the health facility for growth monitoring and vitamin A supplement
    - Frequency of required compliance: twice per year
    - Frequency of compliance monitoring: every six months.

Children aged between six and 17 years should:

- Enroll in school
    - Frequency of required compliance: once per academic year
    - Frequency of compliance monitoring: every 12 months.

- Attend basic education institutions
    - Frequency of required compliance: 80 per cent attendance of effective days
    - Frequency of compliance monitoring: every two months.

One adult parent or caregiver should:

- Attend awareness sessions
    - Frequency of required compliance: once per year
    - Frequency of compliance monitoring: every 12 months.

Table A.1(B): Household Survey Questions on Knowledge of Program Rules and Conditions

Survey questions
Do families participating in the OVC cash transfer programme have to follow any rules in order to continue receiving payments? 1 = Yes 2 = No 98 = Don't Know
Can you please list the rules that you think cash transfer families have to follow in order to receive the full payment from the OVC programme? A = Enrolment / attendance in primary school only B = Enrolment / attendance in primary and secondary schools C = Attendance to health facility for immunizations D = Attendance to health facility for growth monitoring E = Attendance to health facility for vitamin A supplement F = Adequate food and nutrition for children G = Clean and appropriate clothing for children H = Attendance at OVC Programme community awareness sessions I = Birth certificate for children J = Other, specify ____________________ 98 = Don't Know
Which household members do these rules apply to? 1 = All children in the household 2 = Only to orphans and vulnerable children 3 = Other, specify____________________ 98 = Don't know
Do you know what will happen if cash transfer families do not follow the rules? 1 = Yes 2 = No
What will happen to a cash transfer family if they do not follow all of the rules? 1 = Nothing 2 = Kicked out of the programme 3 = Go to jail 4 = A penalty fine will be deducted from the next payment – but do not know the amount 5 = A penalty fine will be deducted from the next payment – 500KS for every rule that is not followed 6 = Other ____________________
Is anyone checking to see if cash transfer families are following the rules? 1 = Yes 2 = No 98 = Don't know

Table A.1(C): Household Survey Questions on Knowledge of Program Rules and Conditions

Survey questions
Can you please list the reasons why a cash transfer family would be asked to leave the OVC cash transfer programme? A = After being in the programme for 5 years B = The household no longer has orphans or vulnerable children below 18 years old C = Household members do not follow all of the rules of the OVC Programme for 3 consecutive periods D = The household moves to another district where the OVC Programme is not operating E = The household caregiver has presented false information related to the eligibility for the Programme F = The household does not collect the payment for 3 consecutive collections G=Misuse of the money, specify ___________________ H=Neglect of the OVC, specify ___________________ I = Other, *specify* ___________________________ 98 = Don't know
Have you ever gone to the Post Office to collect your payment and received less than 3000KS for the payment cycle? 1 = Yes 2 = No **Interviewer:** Look at all of the receipts provided the respondent and look for cash transfer amounts of less than KS 3000.
For the last time you received less than 3000KS for your payment, do you know why you received less? 1 = Yes 2 = No
Do you know if there is an appeal/complaints process if you ever receive less than 3000 KS in a payment cycle? 1 = Yes 2 = No

3.8. APPENDIX B: CT-OVC OVERALL PROGRAM IMPACTS

In order to examine how cash transfer receipt in the CT-OVC program affected outcomes as a whole, we first assess the comparability of the transfer and control groups at baseline as shown in Appendix Table B.1. Instead of using traditional clustered standard errors, we conduct inference using p-values generated from the wild cluster bootstrap and randomization inference (discussed in more detail in section 3.3.3). While most of the differences in treatment versus control group means were statistically insignificant at the 5 percent level according to the wild cluster bootstrap, there were additional statistically significant differences when using randomization inference. These differences appear to be driven by the prioritization of households with very old caregivers of OVCs, as evidenced by the imbalance in household composition of older members. Although we controlled linearly for the prioritization criteria when comparing means, these differences persist. We also looked at attrition by treatment status between the cash transfer and control groups (overall about 24 percent), and there did appear to be some differential attrition between these groups, as households that received transfers were 10 percentage points less likely to attrit than control households (see Appendix Table B.2). However, since the main focus of this study is on comparing the randomly assigned treatment arms to one another (CCT versus LCT), we were not overly concerned about the imperfect balance and presence of differential attrition between the pooled transfer and control groups.

We estimate the overall impact of the CT-OVC cash transfer program using equation (B.1) below. The variable y_{iljk} refers to the outcome measures for child l (school enrollment/attendance, immunizations and vitamin A supplementation) and household i (consumption outcomes) in sub-location j and district k. $transfer_{jk}$ indicates random assignment of sub-location j to receive the cash transfer. The variables X'_{1ijk}, X'_{2ijk}, $carerIndex_{ijk}$, $totalOVC_{ijk}$, and $totalChronicallyIll_{ijk}$ have the same meanings as described in the main text.

$$y_{lijk} = \propto + \delta transfer_{jk} + \gamma_1 carerIndex_{ijk} + \gamma_2 totalOVC_{ijk}$$
$$+ \gamma_3 totalChronicallyIll_{ijk} + X'_{1ijk}\beta_1 + X'_{2ijk}\beta_2 + e_{lijk} \quad (3)$$

The results from these estimates are given in Appendix Table B.3 and suggest that assignment to the transfer increased both food and non-food consumption. These results are consistent with the findings of the Kenya CT-OVC Evaluation Team (2012) in their differences-in-differences impact analysis, which indicated that CT-OVC cash transfer receipt was associated with increases in household consumption of both food and non-food items and a reduction in poverty levels by about 13 percentage points. However, it appears as though the only conditioned-upon outcome that was affected by the cash transfer was school attendance conditioned on enrollment. The Kenya CT-OVC Evaluation Team reported impacts on secondary school enrollment, but similar to what we find in our analysis, they found no overall impacts on child health indicators. One exception to this pattern of findings on child health was reported by Huang et al. (2017), who identified a reduction in the incidence of illness (fever and hot body symptoms) among children 0-7 years in the CT-OVC.

Table B.1: Balance Table: Cash Transfer vs. Control Group

	(1) Control	(2) Transfer	(3) Bootstrap P-Value	(4) RI P-Value
Years of Edu. of HH Head	6.748	5.902	0.682	0.535
Sex of HH Head	0.411	0.352	0.448	0.386
HH Receives Labor Wages	0.080	0.034	0.402	0.130
HH Receives Outside Transfer	0.199	0.294	0.411	0.205
Poor Quality Walls	0.849	0.726	0.123	0.044
Poor Quality Floor	0.824	0.730	0.182	0.041
HH Owns Livestock	0.797	0.799	0.757	0.640
Cattle Owned	1.606	1.289	0.174	0.017
Poultry Owned	6.080	4.413	0.054	0.009
Owns Telephone	0.164	0.105	0.570	0.386
Owns Blanket	0.872	0.842	0.500	0.229
Owns Mosquito Net	0.703	0.604	0.254	0.157
Acres of Land Owned	2.32	1.599	0.089	0.010
Household in Rural Location	0.700	0.832	0.955	0.945
HH Total Consumption	1.640	1.603	0.590	0.372
HH Food Consumption	0.925	0.952	0.404	0.198
HH Non-food Consumption	0.715	0.651	0.876	0.808
Dietary Diversity Score	5.513	5.114	0.150	0.021
Size of the HH	5.703	5.444	0.736	0.709
Age of HH Head	48.249	58.067	0.513	0.366
People Aged 0-5 in HH	0.833	0.672	0.814	0.752
People Aged 6-11 in HH	1.320	1.225	0.455	0.245
People Aged 12-17 in HH	1.374	1.366	0.130	0.045
People Aged 18-45 in HH	1.514	1.129	0.574	0.563
People Aged 46-64 in HH	0.427	0.652	0.005	0.000
People Aged 65+ in HH	0.235	0.399	0.067	0.005
N	438	1092		

Note: The p-values in column (3) are calculated with the wild cluster bootstrap procedure, and the p-values in column (4) are calculated with randomization inference. Clustering is done at the district level. All estimates are calculated conditioning on the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. Some variables have fewer observations than given in the final row due to missing responses.

Table B.2: Differential Attrition by Treatment Status

Panel A: Cash Transfer vs. Control Group	(1) Control Mean	(2) Transfer Differential Effect	(3) Bootstrap P-Value	(4) RI P-Value	(5) N
Assigned to Transfer	0.298	-0.107	0.003	0.000	1978

Panel B: CCT vs. LCT	(1) LCT Mean	(2) CCT Differential Effect	(3) Bootstrap P-Value	(4) RI P-Value	(5) N
Assigned to CCT	0.171	0.038	0.266	0.183	1351
Percentile: 1-20	0.222	0.040	0.064	0.108	310
Percentile: 20-40	0.149	0.017	0.643	0.637	291
Percentile: 40-60	0.156	-0.008	0.886	0.960	259
Percentile: 60-80	0.143	0.074	0.147	0.099	251
Percentile: 80-99	0.176	0.072	0.177	0.266	240

Note: The p-values in column (3) are calculated with the wild cluster bootstrap procedure, and the p-values in column (4) are calculated with randomization inference. Clustering is done at the sub-location level in Panel A and distict level in Panel B. All estimates are calculated conditioning on the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. Percentiles refer to households' places in the baseline consumption distribution.

Table B.3: Impact of Assignment to Cash Transfer versus Control

	(1) Control Mean	(2) Transfer Differential Effect	(3) Bootstrap P-value	(4) RI P-Value	(5) N
Enrolled in School	0.935	0.006	0.570	0.392	3716
Days Missed from School	1.195	-0.417	0.024	0.008	3294
Total Doses of Vaccinations	7.375	-0.366	0.361	0.378	371
Number of Vacc. Sequences Completed	3.000	-0.219	0.198	0.216	371
Received Vitamin A Supplement	0.477	0.099	0.213	0.146	895
HH Total Consumption	2.021	0.342	0.006	0.002	1530
HH Food Consumption	1.211	0.207	0.003	0.002	1531
HH Non-food Consumption	0.810	0.136	0.034	0.014	1535

Note: The p-values in column (3) are calculated with the wild cluster bootstrap procedure, and the p-values in column (4) are calculated with randomization inference. Clustering is done at the sub-location level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The sample for both education variables consists only of children that were aged at 2 to 17 at baseline. Only children enrolled in school at baseline and follow-up are included in the sample for Days Missed from School, and the reference period is the past 2 months. Vaccinations include BCG (1 dose sequence), DPT (3 doses), OPV (4 doses), and measles (1 dose). Only children aged seven or under at follow-up are included in the sample for the vaccination variables and the vitamin A supplement, and the reference period is the past 6 months.

3.9. APPENDIX C: CORRECTING FOR MULTIPLE HYPOTHESIS TESTING

When testing for the significance of many coefficients, as we do with this paper, it is possible to find significant treatment effects purely by chance even when the true effects are zero. In order correct for this, we re-analyze the data while controlling for the false discovery rate (FDR) (Benjamini and Hochberg, 1995). The FDR is the expected proportion of null hypothesis rejections that are Type I errors (false rejections). By implementing the procedure detailed in Benjamini, Krieger, and Yekutieli (2006), each p-value from our previous analysis is assigned a "q-value", or the lowest FDR that would allow us to reject the null hypothesis for the given p-value. These q-values are separately calculated for each family of outcomes, or a group of outcomes for which p-values are expected to be positively correlated. Our study defines four families of outcomes: educational outcomes (Days Missed from School, Enrollment), vaccination outcomes (Total Doses of Vaccines, Completed Vaccination Sequences), Vitamin A Supplement Received, and consumption outcomes (HH Total Consumption, HH Food Consumption, HH Non-food Consumption). We calculate two sets of q-values for each family, one for each method of conducting inference (wild cluster bootstrap and randomization inference), and report them in Appendix Table C.1.

Table C.1(A): Impacts of Assignment to CCT versus LCT: Heterogeneous Effects (Controlling for FDR)

	(1) Enrolled in School	(2) Days Absent from School	(3) Total Doses of Vaccinations	(4) Number of Vacc. Sequences Completed
Effects by Baseline Consumption Percentile				
Percentile: 10	0.014	-0.174	-0.586	-0.290
	(1.000)	(1.000)	(0.802)	(0.802)
	[1.000]	[1.000]	[1.000]	[1.000]
Percentile: 25	0.010	-0.150	-0.561	-0.273
	(1.000)	(1.000)	(0.802)	(0.802)
	[1.000]	[1.000]	[1.000]	[1.000]
Percentile: 50	0.003	-0.112	-0.520	-0.245
	(1.000)	(1.000)	(0.802)	(0.802)
	[1.000]	[1.000]	[1.000]	[1.000]
Percentile: 75	-0.006	-0.064	-0.469	-0.210
	(1.000)	(1.000)	(0.802)	(0.802)
	[1.000]	[1.000]	[1.000]	[1.000]
Percentile: 90	-0.019	0.006	-0.394	-0.160
	(1.000)	(1.000)	(0.832)	(0.832)
	[1.000]	[1.000]	[1.000]	[1.000]
LCT Mean	0.937	1.109	7.429	3.010
N	2549	2242	235	235

Note: The FDR q-values contained in parentheses are based on wild cluster bootstrap p-values, and those in brackets are based on randomization inference p-values. Clustering is done at the district level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The sample for both education variables consists only of children that were aged at 2 to 17 at baseline. Only children enrolled in school at baseline and follow-up are included in the sample for Days Missed from School, and the reference period is the past 2 months. Vaccinations include BCG (1 dose sequence), DPT (3 doses), OPV (4 doses), and measles (1 dose). Only children aged seven or under at follow-up are included in the sample for the vaccination variables and the vitamin A supplement, the reference period for the latter is the past 6 months.

Table C.1(B): Impacts of Assignment to CCT versus LCT: Heterogeneous Effects (Controlling for FDR)

	(1) Received Vitamin A Supplement	(2) HH Total Consumption	(3) HH Food Consumption	(4) HH Non-food Consumption
Effects by Baseline Consumption Percentile				
Percentile: 10	0.022	-0.206	-0.041	-0.163
	(1.000)	(0.732)	(1.000)	(0.039)
	[1.000]	[1.000]	[1.000]	[0.001]
Percentile: 25	0.016	-0.156	-0.015	-0.139
	(1.000)	(0.852)	(1.000)	(0.039)
	[1.000]	[1.000]	[1.000]	[0.417]
Percentile: 50	0.005	-0.075	0.026	-0.100
	(1.000)	(1.000)	(1.000)	(0.436)
	[1.000]	[1.000]	[1.000]	[1.000]
Percentile: 75	-0.009	0.028	0.079	-0.050
	(1.000)	(1.000)	(1.000)	(1.000)
	[1.000]	[1.000]	[1.000]	[1.000]
Percentile: 90	-0.029	0.177	0.154	0.022
	(1.000)	(1.000)	(0.732)	(1.000)
	[1.000]	[1.000]	[1.000]	[1.000]
LCT Mean	0.510	2.107	1.269	0.838
N	561	1092	1093	1095

Note: The FDR q-values contained in parentheses are based on wild cluster bootstrap p-values, and those in brackets are based on randomization inference p-values. Clustering is done at the district level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The sample for both education variables consists only of children that were aged at 2 to 17 at baseline. Only children enrolled in school at baseline and follow-up are included in the sample for Days Missed from School, and the reference period is the past 2 months. Vaccinations include BCG (1 dose sequence), DPT (3 doses), OPV (4 doses), and measles (1 dose). Only children aged seven or under at follow-up are included in the sample for the vaccination variables and the vitamin A supplement, the reference period for the latter is the past 6 months.

3.10 APPENDIX D: LONGER-RUN CT-OVC PROGRAM IMPACTS

Table D.1: Impact of Assignment to Cash Transfer versus Control: Longer-Run Effects

	(1) Control Mean	(2) Transfer Differential Effect	(3) Bootstrap P-value	(4) RI P-Value	(5) N
Enrolled in School	0.930	0.018	0.207	0.088	3447
Days Missed from School	0.655	-0.152	0.421	0.058	1681
Total Doses of Vaccinations	7.759	0.325	0.479	0.582	109
Number of Vacc. Sequences Completed	3.070	0.150	0.638	0.691	104
HH Total Consumption	2.459	0.036	0.841	0.819	1380
HH Food Consumption	1.512	0.083	0.543	0.467	1384
HH Non-food Consumption	0.957	-0.048	0.531	0.425	1382

Note: The p-values in column (3) are calculated with the wild cluster bootstrap procedure, and the p-values in column (4) are calculated with randomization inference. Clustering is done at the sub-location level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The sample for both education variables consists only of children that were aged at 2 to 17 at baseline. Only children enrolled in school at baseline and follow-up are included in the sample for Days Missed from School, and the reference period is the past 2 weeks in the 2011 data. Vaccinations include BCG (1 dose sequence), DPT (3 doses), OPV (4 doses), and measles (1 dose). Vitamin A supplament data were not available for the 2011 follow-up period.

Table D.2(A): Impacts of Assignment to CCT versus LCT: Longer-Run Effects

	(1) Enrolled in School	(2) Days Absent from School	(3) Total Doses of Vaccinations	(4) Number of Vacc. Sequences Completed
Panel A: Average Effects				
Assigned to CCT	-0.007	0.011	-0.333	-0.083
	(0.591)	(0.878)	(0.641)	(0.846)
	[0.619]	[1.000]	[0.649]	[0.792]
Panel B: Effects by Baseline Consumption Percentile				
Percentile: 10	-0.008	-0.049	-0.513	-0.135
	(0.606)	(0.619)	(0.615)	(0.830)
	[0.633]	[0.558]	[0.844]	[0.848]
Percentile: 25	-0.008	-0.029	-0.425	-0.110
	(0.578)	(0.717)	(0.606)	(0.828)
	[0.586]	[0.850]	[0.795]	[0.794]
Percentile: 50	-0.007	0.004	-0.283	-0.070
	(0.586)	(0.956)	(0.618)	(0.856)
	[0.660]	[1.000]	[0.740]	[0.844]
Percentile: 75	-0.006	0.047	-0.100	-0.018
	(0.658)	(0.577)	(0.784)	(0.933)
	[0.674]	[0.948]	[0.901]	[0.907]
Percentile: 90	-0.005	0.107	0.162	0.056
	(0.775)	(0.457)	(0.758)	(0.858)
	[0.841]	[0.542]	[0.904]	[0.851]
LCT Mean	0.936	0.612	7.773	3.047
N	2361	1137	79	74

Note: P-values calculated with the wild cluster bootstrap procedure are in parentheses, and p-values calculated with randomization inference are in brackets. Clustering is done at the district level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The sample for both education variables consists only of children that were aged at 2 to 17 at baseline. Only children enrolled in school at baseline and follow-up are included in the sample for Days Missed from School, and the reference period is the past 2 weeks in the 2011 data. Vaccinations include BCG (1 dose sequence), DPT (3 doses), OPV (4 doses), and measles (1 dose). Vitamin A supplement data were not available for the 2011 follow-up period.

Table D.2(B): Impacts of Assignment to CCT versus LCT: Longer-Run Effects

	(1) HH Total Consumption	(2) HH Food Consumption	(3) HH Non-food Consumption
Panel A: Average Effects			
Assigned to CCT	0.092	0.062	0.024
	(0.641)	(0.643)	(0.781)
	[0.698]	[0.691]	[0.734]
Panel B: Effects by Baseline Consumption Percentile			
Percentile: 10	-0.033	0.038	-0.081
	(0.871)	(0.774)	(0.296)
	[0.903]	[0.799]	[0.354]
Percentile: 25	0.006	0.046	-0.048
	(0.974)	(0.732)	(0.502)
	[0.948]	[0.799]	[0.538]
Percentile: 50	0.069	0.058	0.005
	(0.711)	(0.668)	(0.931)
	[0.856]	[0.792]	[0.948]
Percentile: 75	0.150	0.074	0.073
	(0.500)	(0.599)	(0.411)
	[0.450]	[0.582]	[0.426]
Percentile: 90	0.266	0.097	0.171
	(0.313)	(0.560)	(0.117)
	[0.284]	[0.595]	[0.189]
LCT Mean	2.066	1.321	0.745
N	985	987	985

Note: P-values calculated with the wild cluster bootstrap procedure are in parentheses, and p-values calculated with randomization inference are in brackets. Clustering is done at the district level. All estimates are calculated conditioning on control variables and the transfer prioritization criteria: Carer Age Index, Chronically Ill Members of HH, and Number of OVCs in HH. Consumption variables are in terms of 1000 KSh per adult-equivalent. The sample for both education variables consists only of children that were aged at 2 to 17 at baseline. Only children enrolled in school at baseline and follow-up are included in the sample for Days Missed from School, and the reference period is the past 2 weeks in the 2011 data. Vaccinations include BCG (1 dose sequence), DPT (3 doses), OPV (4 doses), and measles (1 dose). Vitamin A supplement data were not available for the 2011 follow-up period.