ABSTRACT

Title of Dissertation: Essays on Mental Health, Education, and Parental Labor Force Participation

Rachel Nesbit
Doctor of Philosophy, 2024

Dissertation Directed by: Professors Guido Kuersteiner and Nolan Pope
Department of Economics

This dissertation consists of three chapters in empirical microeconomics. The first chapter focuses on mental health in the criminal justice system. I show that mandated mental health treatment during probation decreases future recidivism and further that paying for these probationers to receive treatment would be a very cost-effective program. The second chapter focuses on the labor supply of same-sex couples. My coauthors and I document the earnings patterns in same-sex couples after the entrance of their first child and contrast them with the earnings patterns in opposite-sex couples. The third chapter evaluates state-level policies to offer a college admissions exam (either the SAT or ACT) free to all high school students. I estimate precise null effects of the policies on future college attendance. The three chapters are described in further detail below.

Chapter 1. Mental health disorders are particularly prevalent among those in the criminal justice system and may be a contributing factor in recidivism. Using North Carolina court cases from 1994 to 2009, this chapter evaluates how mandated mental health treatment as a term

of probation impacts the likelihood that individuals return to the criminal justice system. I use random variation in judge assignment to compare those who were required to seek weekly mental health counseling to those who were not. The main findings are that being assigned to seek mental health treatment decreases the likelihood of three-year recidivism by about 12 percentage points, or 36 percent. This effect persists over time, and is similar among various types of individuals on probation. In addition, I show that mental health treatment operates distinctly from drug addiction interventions in a multiple-treatment framework. I provide evidence that mental health treatment's longer-term effectiveness is strongest among more financially advantaged probationers, consistent with this setting, in which the cost of mandated treatment is shouldered by offenders. Finally, conservative calculations result in a 5:1 benefit-to-cost ratio which suggests that the treatment-induced decrease in future crime would be more than sufficient to offset the costs of treatment.

Chapter 2. Existing work has shown that the entry of a child into a household results in a large and sustained increase in the earnings gap between male and female partners in opposite-sex couples. Potential reasons for this include work-life preferences, comparative advantage over earnings, and gender norms. We expand this analysis of the child penalty to examine earnings of individuals in same-sex couples in the U.S. around the time their first child enters the household. Using linked survey and administrative data and event-study methodology, we confirm earlier work finding a child penalty for women in opposite-sex couples. We find this is true even when the female partner is the primary earner pre-parenthood, lending support to the importance of gender norms in opposite-sex couples. By contrast, in both female and male same-sex couples, earnings changes associated with child entry differ by the relative pre-parenthood earnings of the partners: secondary earners see an increase in earnings, while on average the earnings of primary and equal earners remain relatively constant. While this finding seems supportive of a norm related to equality within same-sex couples, transition analysis suggests a more complicated story.

Chapter 3. Since 2001, more than half of US states have implemented policies that require all public high schools to administer either the ACT or SAT to juniors during the school day free of charge, making that aspect of the college application process less costly in both time and money. I evaluate these policies using American Community Surveys (ACS) from 2000 to 2019. I augment ACS data with the Census Master Address File to precisely identify the state in which individuals took the exam. Exploiting variation in policy implementation across state and time, I find across all specifications that increased access to standardized college entrance exams has no effect on subsequent college attendance. It also does not shift students between public and private colleges or between two- and four-year programs. The results of this chapter suggest that, to the extent that these policies were introduced to encourage college-going among marginal students, they did not accomplish their goal. This provides evidence about the kinds of support necessary to influence educational outcomes for students from disadvantaged families.

ESSAYS ON MENTAL HEALTH, EDUCATION, AND PARENTAL LABOR FORCE PARTICIPATION

by

Rachel Nesbit

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:
Professor Guido Kuersteiner, Co-Chair/Co-Advisor
Professor Nolan Pope, Co-Chair/Co-Advisor
Professor David Blazar
Professor Ethan Kaplan
Professor Melissa Kearney

Preface

This dissertation presents research on three topics in empirical microeconomics, in the areas of labor, health, and public economics. While the topic of each chapter is distinct, together they represent the three focal areas of my research, which is unified by an interest in identifying policy-relevant parameters with a particular focus on understudied or marginalized groups of society.

My research contributions are based in three main areas: mental health, parental labor force participation, and education. In the area of health, my first chapter takes a step toward drawing attention to the importance of mental health in the criminal justice system by studying the impact of judge-mandated mental health treatment on the incidence of future crime. Exploring the impacts of mental health on economically important outcomes such as crime is a main focus of my research agenda. Studying individuals in the criminal justice system also contributes to an understanding of policy interventions that can improve outcomes for a highly disadvantaged group of people.

In the area of labor force participation, my interest is in understanding how individuals make joint decisions about child-rearing and labor supply. My second chapter explores how parenthood impacts the earnings of individuals in same-sex couples. Through comparison with the earnings trajectories of opposite-sex couples, this can shed light on the importance of gender norms in the parental labor supply decision. In addition, the project contributes to increasing knowledge of an understudied population, same-sex couples, who have only recently been reliably identifiable in core US survey data.

In the area of education, my third chapter investigates state-level policies providing increased access to college entrance exams, which both reduced the cost of applying for college and gave students the chance to gain more information about their college readiness. However, I show that the policies were not sufficient to increase college attendance, even among low-income or minority students whom the policies were intended to benefit. This project contributes to an understanding of effective policy design to address education gaps among low-income and minority students.

All three chapters contribute to a greater understanding of understudied groups, especially how public policy can impact educational attainment, employment, and criminal behavior in those groups. Taken together, they summarize the three main strands of my existing research.

The second chapter of my dissertation was written with three excellent collaborators at the US Census Bureau, Barbara Downs, Lucia Foster, and Danielle H. Sandler, who have given their permission for its inclusion. The version in this dissertation is heavily based on our working paper (Center for Economic Studies, CES-23-25). As is standard for work with Census data, the following disclaimer applies:

> Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data (Project No. 7525733, Disclosure Review Board (DRB) approval number: CBDRB-FY22-CES004-013 and CBDRB-FY22-CES004-054).

I also want to take a moment to thank my coauthors for their wisdom, perspective, and support. Working with Barbara, Lucia, and Dani has been such a delight, and I appreciate them not only for their insight into the project, which has resulted in a better paper, but also for their encouragement, mentorship, and friendship.

On the subject of gratitude, it is impossible to properly remember here all the people who have contributed to the completion of this dissertation and my graduate education, but I will now endeavor to express my deepest thanks to the many souls who supported me through this process.

I am immensely grateful for the guidance of my co-chairs, Nolan Pope and Guido Kuersteiner. Their support and assistance has been invaluable. They have consistently challenged me and held me to high standards, and have labored to get me to believe that I am capable of meeting those standards. My work has benefited greatly from their feedback and I am a more thoughtful economist because of their example. Their insightful comments in seminars and advice in one-on-one meetings made me a better and more careful researcher, but their kindness to me is what I will treasure most.

I've learned so much from the faculty in the Department of Economics, and have grown tremendously as an economist and researcher because of their investments of time and energy. Ethan Kaplan, Jessica Goldberg, Judy Hellerstein, and many others have enhanced the quality of my research through their comments in 708 and beyond. A special thanks to Melissa Kearney and John Haltiwanger for, early in the program, providing me with models of the kind of researcher and economist I aspire to be. I have had the great pleasure of meeting many amazing economists throughout my graduate school experience; those economists are amazing not only because of their brilliance, drive, and depth of knowledge, but also because they are kind, patient, inquisitive, and giving of their time.

Our department is built on and supported by a foundation of many wonderful people who do not get nearly the credit and gratitude that they deserve. I appreciate John Shea, Vickie Fletcher, Wanda Hauser, Mark Wilkerson, Brian McLoughlin, and the other members of the department staff for their tireless work facilitating my and others' progress through graduate school. I can always count on Vickie and Mark to greet me with a big smile when I see them in the hallway, and that has made many a tough day better.

When I was a first-year graduate student, several people advised me to start a study group so that we could take turns talking each other out of leaving the program. As it turns out, that was excellent advice, and I am indebted to Ben Layton and Victoria Perez-Zetune for the many times they convinced me to keep going when I was considering giving up. A colossal thank you to many other fellow PhD students, including but not limited to Lea Rendell, Keaton Ellis, Sueyoul Kim, Maranna Yoder, Taylor Landon, Erica Ryan, and Anu Sivaram. Y'all have offered me friendship, guidance, and have always been willing to talk, whether about research or very much not research (I'm thinking especially of my favorite third floor office, the Fun Office). Friends outside of graduate school have also kept me sane and encouraged me, and I owe a special thanks to my dearest Elizabeth Chase, Emily Nobles, and Jean Thomas Rowan.

I would be nowhere without the unconditional love, support, and encouragement of my family. My father, mother, sister, brother, and grandmother have been my biggest cheerleaders and their unwavering belief in me has motivated me to continue my studies. Weekly phone calls with my grandmother about her ukulele club and my latest baking experiments reminded me what was truly important and helped me persevere through the most stressful of times. My sister, who knows me better than I know myself, has helped me stay true to myself when making difficult

decisions. My brother has kept me humble and supplied with never-ending puns. My mother has always been ready to help, whether that's shopping for interview attire, listening to countless anxious phone calls, or gifting me plants to cheer me up. And to my father: all of my hard work is in pursuit of the kind of work ethic and integrity that you have modeled for me. I look up to you more than I can ever express.

Finally, just like I always set aside the best bite of a good meal for last: Colten. Your support and confidence in me has meant the world to me. I couldn't have asked for a more patient, long-suffering partner to be my rock in the turbulent waters of graduate school. I will treasure you always.

Dedication

To my family, the best of all blessings.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACIS	Automated Criminal Infractions System
ACS	American Community Survey
ACT	American College Testing
AMI	Any Mental Illness
AOC	Administrative Office of the Courts
ATUS	American Time Use Survey
BEA	Bureau of Economic Analysis
BLS	Bureau of Labor Statistics
BRFSS	Behavioral Risk Factor Surveillance System
CBO	Congressional Budget Office
CBT	Cognitive Behavioral Therapy
CCD	Common Core of Data
CDC	Centers for Disease Control and Prevention
CES	Center for Economic Studies
CHIP	Children's Health Insurance Program
CPI	Consumer Price Index
CPI-U	Consumer Price Index for All Urban Consumers
DHHS	Department of Health & Human Services
DRB	Disclosure Review Board
DSM-IV	Diagnostic & Statistical Manual of Mental Disorders, 4th Edition
ESSA	Every Student Succeeds Act
FBI	Federal Bureau of Investigation
GAO	Government Accountability Office
GDP	Gross Domestic Product
HHS	Department of Health & Human Services
HS	High School

IPUMS	Integrated Public Use Microdata Series
IRS	Internal Revenue Service
IV	Instrumental Variables
IZA	Institute of Labor Economics
LATE	Local Average Treatment Effect
LEHD	Longitudinal Employer-Household Dynamics
LPM	Linear Probability Model
MAF	Master Address File
MF	Male-Female couple
MH	Mental Health
MHT	Mental Health Treatment
MM	Male-Male couple
MVPF	Marginal Value of Public Funds
NCES	National Center of Education Statistics
NHIS	National Health Interview Survey
NIMH	National Institute of Mental Health
NPC	Northwest Professional Consortium
NSDUH	National Survey on Drug Use and Health
OLS	Ordinary Least Squares
PAA	Population Association of America
PIK	Personal Identification Key
PSAT	Preliminary SAT
QCEW	Quarterly Census of Employment & Wages
RCT	Randomized Control Trial
RSF	Russell Sage Foundation
SAMHSA	Substance Abuse and Mental Health Services Administration
SAT	Scholastic Aptitude Test
SNAP	Supplemental Nutrition Assistance Program
SSA	Social Security Administration
SSRN	Social Science Research Network
STAR	Student/Teacher Achievement Ratio
SUD	Substance Use Disorder

SUDT	Substance Use Disorder Treatment
UI	Unemployment Insurance
UPM	Unordered Partial Monotonicity
WP	Working Paper

Chapter 1: The Role of Mandated Mental Health Treatment in the Criminal Justice System

1.1 Introduction

Poor mental health is widely prevalent and has been growing over time, with around 58 million adults in the United States suffering from a mental illness in 2021 (SAMHSA, 2022).[1] Poor mental health is also widely impactful; it has direct negative effects on physical and social health and is highly intertwined with other aspects of life. For example, it contributes to upwards of 40 percent of all illnesses under the age of 65 (Layard, 2013). Beyond (or as a consequence of) these direct negative effects on health, mental illness has been associated with behaviors like absenteeism at work, decreased educational attainment, and crime (Breslau et al., 2008; Bubonya et al., 2017; Burton et al., 2008; Mojtabai et al., 2015).

Poor mental health is particularly prevalent in the criminal justice system. Among the approximately five million men on probation in 2009, 33 percent met the threshold for any mental illness - nearly twice the prevalence in the general population at the time (Feucht and Gfroerer, 2011).[2] Individuals with a mental illness can struggle to make rational, welfare-maximizing

[1]Among them, approximately 22% had a serious mental illness, 37% had a moderate mental illness, and 40% had a mild mental illness (Bagalman and Cornell, 2018).

[2]Any Mental Illness (AMI) is defined by SAMHSA (2022) as having "any mental, behavioral, or emotional disorder in the past year of sufficient duration to meet criteria from the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV), excluding developmental disorders and SUDs [Substance Use Disorders]"

decisions, which may play a critical role in their interactions with the criminal justice system. Since the medical field has shown that therapy and medication can reduce symptoms of mental illness (Cronin et al., 2020; Paykel et al., 1999), it may be the case that mental health treatment can reduce behavioral outcomes such as crime. That reduction could have large benefits for both treated individuals and society. However, though meaningful correlational evidence exists, there is little direct causal evidence on the effect of therapy and psychiatric medication on outcomes such as criminal behavior, or whether those interventions could be carried out in a cost-effective way.[3]

This paper evaluates the causal impact of mandated mental health treatment on the likelihood of committing a future crime. Specifically, it focuses on how being assigned mental health treatment at the time of probation impacts recidivism over the next five years. To do this I exploit judge variation in court-mandated mental health treatment in North Carolina, using the universe of criminal court cases from 1994 to 2009. The requirement that the defendant seek mental health treatment is a possible condition of probation (supervised release without serving time in prison). While the type of mental health treatment can vary, it typically includes a psychological evaluation, weekly therapy sessions for the duration of the sentence, and potentially a referral to a psychiatrist for medication. Simply comparing outcomes of probationers who are and are not mandated mental health treatment could result in a biased estimate if judges make their sentencing decisions using additional information that is unobserved to the researcher. In particular, while the data provide information on demographics and criminal history, they do not provide information about mental illness, which is likely correlated with both the judge's sen-

[3] See, for example, Fazel et al. (2010), Grady and Gough (2014), Hiday et al. (2002), and Wüsthoff et al. (2014) for treatment as a protective factor in risk of violence; Ziemek (2017) for a review of evidence on treatment in correctional facilities; Green (2020) for a review of the link between mental illness and criminal behavior.

tence and recidivism risk. To address this, I use the randomly-assigned judge's propensity to mandate mental health treatment among other cases as an instrument for actually being assigned mental health treatment as a term of probation. This research design has been used in various applications in which a judge, case worker, or other type of administrator has discretion over a randomly assigned caseload.[4] In order for this instrument to provide exogenous variation, the judges must be assigned to cases in a method that is unrelated to factors impacting sentencing and future criminal behavior. In North Carolina, judges are assigned in a rotating pattern which is independent of the individuals in their jurisdiction (Silveira, 2017; Sloan et al., 2016). To verify random assignment, I conduct balance tests which show that demographic and criminal background characteristics, while predictive of receiving therapy, are not predictive of the assigned judge's residualized propensity measure.

The analysis consists of two main parts, first evaluating the impact of mandated mental health treatment on recidivism and then evaluating the cost-effectiveness of mandating mental health treatment for judge-chosen individuals on probation. Using the judge instrument, I find that mandated mental health treatment as a term of probation decreases future crime in both the short and longer term. Three years after conviction, individuals sentenced to mental health treatment are 12.1 percentage points less likely to recidivate. That effect corresponds to a 36 percent decrease in three-year recidivism relative to the mean, and is robust to various specifications of the instrument and treatment. The IV estimate is more than twice as large as the OLS estimate, consistent with judges being more likely to assign treatment to those with underlying mental health issues, who are at a higher risk of recidivism.

[4] See, for example: Kling (2006), Dobbie et al. (2018), Leslie and Pope (2017) for other applications in the criminal justice system; Kalish (2023) for the foster care system; Belloni et al. (2012) for eminent domain cases; Autor and Houseman (2010) for evaluation of a welfare program.

In addition, I find that the effect of mental health treatment persists over time, driven by reductions in the first three years. By five years after conviction, offenders are still about 11 percentage points (25 percent) less likely to recidivate. That suggests that mental health treatment is not just shifting when recidivism occurs. Because the average length of probation is about six months, the persistence of the effect means the treatment is effective well beyond the period during which it is mandated. Lacking information on treatment after probation, persistence could be driven by the effectiveness of treatment even after it ends, or instead by the offenders continuing to seek treatment after the mandated period is over. Nevertheless, the persistence stands in contrast with evaluations of cognitive behavioral therapy (CBT) programs in more targeted populations, which find that effects quickly dissipate after the treatment ends (Arbour, 2021; Blattman et al., 2017; Heller et al., 2017).

To explore heterogeneity in the treatment effect, I estimate two-stage least squares regressions among various subsamples. The results from this exercise suggest that treatment decreases recidivism similarly among different types of crime, first-time and repeat offenders, men and women, offenders of different races and ethnicities, and offenders with different financial resources. The effects range from a 10 percentage point decline (among offenders with a private attorney) to a 20 percentage point decline (among repeat offenders), but are not statistically distinguishable from one another. This analysis shows that offenders with many different characteristics are similarly benefited by mental health treatment. I also test whether treatment has heterogeneous effects on the type of future crime. I find that being mandated to seek mental health treatment is most effective at reducing financial crimes, and is also highly effective at reducing violent and property crimes.

A main identifying assumption underlying this research design is that the sentencing judge only affects recidivism through assignment of mental health treatment. In other contexts the judge has control over more aspects of the sentence, but in North Carolina during this time period, a series of structured sentencing laws mechanically determined the offender's sentence class based on the severity of their offense and the number of prior convictions. Using that information, the analysis sample can be restricted to those who are only eligible for probation. On the other hand, judges do have control over other special conditions of probation. Along with mandating mental health treatment, judges can also require individuals to seek substance use disorder (SUD) treatment. Building on Humphries et al. (2023), Mogstad et al. (2021), Vytlacil (2002), Mountjoy (2022), and others, I construct a model of judge decision-making that is consistent with the North Carolina sentencing context and with unordered partial monotonicity. Several tests show that the observed patterns in the data are consistent with that monotonicity assumption. The empirical strategy in this paper identifies a margin-specific local average treatment effect, holding the judge's propensity toward SUD treatment constant while varying the judge's propensity toward mental health treatment. This allows me to isolate the effect of mental health treatment from that of SUD treatment.

The instrument variable strategy identifies the local average treatment effect (LATE), which is the causal effect of mental health treatment for those on the margin of being sentenced to treatment as a term of probation. However, as Blandhol et al. (2022) note, the result of two-stage least squares with covariates is often not interpretable as the LATE. It produces an average of covariate-specific LATEs only in the case where the specification is saturated in the covariates; accordingly, I fully interact all covariates as a robustness check of my main results. A related concern is the validity of the monotonicity assumption, which requires that if one judge is ranked as more likely to assign mental health treatment than another judge, that ranking will hold true for all offenders regardless of their individual characteristics. In my main results, I rely on a relaxed version of that assumption by creating my instrument separately for different types of offenses. In addition, as a robustness check I follow Mueller-Smith (2023) and others by creating my instrument separately for many groups defined by a range of demographic and criminal history characteristics and their interactions. The results of both corrections to address covariate bias are qualitatively similar to my main results.

In the second part of my analysis, I turn to quantifying the costs and benefits of mental health treatment among probationers. I carry out the analysis within five broad offense groups, using separate estimates of the effect of treatment and the costs of recidivism for each group. This exercise focuses on the direct effect of mandated treatment on recidivism, but mental health treatment could also improve other aspects of life. Therefore this exercise can be thought of as an underestimate of the total benefit of the treatment. Currently North Carolina expects probationers to cover the costs of mental health treatment themselves; I consider a policy in which the government instead takes on those costs. Such a policy would provide benefits through the reduction of direct and indirect costs of crime at a rate of 5 to 1. I relate the individual's willingness to pay to the net costs to the government via the marginal value of public funds (MVPF) of the policy. Incorporating various aspects of uncertainty from the estimates, I estimate an MVPF of 19.3. This suggests that among probationers, mental health treatment for the marginal individual is highly beneficial for society. Paired with the evidence that treatment effects are fairly homogeneous, the results of this analysis suggest treatment could be similarly cost-effective in a broader context.

The idea that treatment of mental illness could improve more than just symptoms of that illness has been documented by many observational studies in sociology, psychology, and criminology (Breslau et al., 2008; Bubonya et al., 2017; Burton et al., 2008; Hakulinen et al., 2021;

Mojtabai et al., 2015; Patel and Kleinman, 2003). However, causal evidence on the treatment of individuals with poor mental health who have had contact with the United States criminal justice system is mostly limited to evaluations of mental health courts and assisted outpatient treatment. Both programs are focused on individuals who have a prior diagnosis of severe mental illness and are at a crisis point. Though most studies have found reduced criminal behavior after these programs, the group being treated is likely to improve over time regardless, since they are committed at "rock bottom" (Frank and McGuire, 2010; Mitton et al., 2007). In addition, most analysis focuses on short term outcomes measured in the months after contact with mental health courts or assisted outpatient treatment. And few studies provide a discussion of the cost-effectiveness of the programs. I contribute to this literature by presenting evidence on a treatment that is less targeted; offenders are not required to meet thresholds of mental illness severity or even to be diagnosed with a mental illness in order to be sentenced to seek treatment. I evaluate short-term and longer-term effects of mental health treatment, finding that the effects of judge-mandated mental health treatment persist at least three years after the sentence. Importantly, the results of this study also show that bringing about a reduction in crime through mandated mental health treatment is a cost-effective strategy.

My findings are consistent with recent studies of other behavioral interventions in the criminal justice system, but I expand on those studies by presenting evidence on a treatment that explicitly focuses on addressing mental health. A 2002 review by Pearson et al. of behavioral programs administered in prisons identified decreases in recidivism, but concentrated among programs that included a cognitive component rather than just behavioral modifications. That suggests scope for improving mental functioning in order to bring about behavioral change. Most of the interventions they review are carried out in the prison system; in contrast, Heller et al. (2017)

evaluate RCTs among low income youth in Chicago and estimate a decrease in recidivism of similar magnitude, between 20 and 30 percent. Importantly, the effect they estimate disappears after the program ends, whereas my estimate persists to at least three years after sentencing. Similarly, Arbour (2021) studies a cognitive behavioral therapy (CBT) program in a Canadian prison that teaches prisoners increased awareness of their thought patterns; he finds that the program decreases recidivism, but the effectiveness dissipates quickly after release for repeat offenders. In contrast, I find that mandated mental health treatment is similarly effective and persistent for both first-time and repeat offenders. I also study adults on probation, an important group that is more than twice the size of the incarcerated population in the United States. Prendergast et al. (2015), Hall et al. (2017), Guydish et al. (2011), and Scott and Dennis (2012) show that attempts to incentivize voluntary participation in programs to treat mental health have been largely unsuccessful. My evidence suggests that mandating participation can be an effective strategy to increase use of mental health treatment.

In addition, development economists have used RCTs to study factors related to mental health in many countries within South America, South and East Asia, and Africa. Most of those studies focus on the benefits to economic productivity, finding that antipoverty programs improve mental health which in turn induces economic gains (Ridley et al., 2020). Blattman et al. (2022) specifically identify a decrease in antisocial behaviors like robbery and selling drugs, which persists as far out as ten years. This literature focuses on policies that relax income constraints. Relatedly, Jácome (2020) studies the relationship between losing Medicaid eligibility and crime in the United States, and shows that the detrimental effects are concentrated among individuals with a prior history of mental illness. Consistent with this, Burns and Dague (2023) show that enrollment in Medicaid reduces the likelihood of recidivism and that the effect operates both through financial and health channels. In contrast, I show that even when faced with additional costs, being sentenced to get mental health treatment decreases recidivism. Because in North Carolina the cost of therapy is usually not covered by the court system, mandating mental health treatment introduces a competing force which might make offenders more likely to recidivate. They could violate their probation if they aren't able to attend treatment due to financial constraints. That suggests the estimated effect in this study is likely an underestimate of the effect of mental health treatment, and might be larger in an environment where the responsibility to pay is not on the offender. On the other hand, being required to pay could create deterrence effects that make the treatment more effective for some groups. In this paper I provide suggestive evidence that both are true among different subsets of the probationers studied.

1.2 Context

1.2.1 Probation in the United States

Probation makes up a large part of the criminal justice system. The Bureau of Justice Statistics estimates that 4.2 million adults were on probation in 2009, compared to 1.6 million adults in prison in the United States (Glaze and Bonczar, 2010; West et al., 2010). Probation is a broad term for sentences that do not include incarceration.[5] The probationer is expected to obey all laws and any additional requirements imposed by their judge or probation officer. For example, if probation is the result of a drug crime, submitting to drug testing will likely be a requirement of probation. They are also expected to report regularly to their probation officer, and the frequency and terms of those meetings can depend on both the severity of the offense and the

[5] A sentence of incarceration is often referred to as an "active" sentence. In this paper, references to a "non-active" sentence mean a sentence that does not involve incarceration.

individual judgment of the officer. Breaking the law or not following any additional requirements imposed is considered a violation of the probation, though whether a violation is reported to the court is at the discretion of the probation officer. Once a violation has been reported, a probation violation hearing will be held and the judge will determine whether the terms were indeed violated. Consequences for a violation span from an extension of the probation sentence to the introduction of an active prison sentence, and may also involve paying additional fines.

Among those on probation, the prevalence of any diagnosable mental illness was 33 percent, twice as high as in the general population. Probationers also reported twice as much unmet need for mental health services – 8.4 percent compared to 4.4 percent in the general population (Feucht and Gfroerer, 2011). Poor mental health can vary widely in intensity, from severe mental illness to occasional symptoms that do not meet the criteria for a full diagnosis. Because Feucht and Gfroerer's estimate of mental illness among probationers does not include individuals who experience poor mental health without meeting the criteria for a mental illness, it may represent an underestimate of the amount of probationers who could benefit from improvements to their mental health. Other studies of mental health interventions in the criminal justice system have focused on high-risk situations, such as creating separate court systems for individuals with serious mental illness; in comparison, a study of probationers likely includes individuals with a larger variation in the degree of poor mental health. By evaluating a more diverse population in terms of their underlying mental health, this paper provides evidence about the effectiveness of treatment beyond the most severe cases.

Mental illness may be related to criminal activity by reducing the ability to make rational, welfare-maximizing decisions. In the context of the criminal justice system, it is correlated with higher risk of committing future crimes and with higher risk of probation violations (Louden and

Skeem, 2011; Ziemek, 2017). Mandating mental health treatment for likely-in-need individuals on probation has the potential to mitigate the higher risk of recidivism by treating the underlying mental illness, teaching better decision-making, and supporting longer term behavioral change. In addition, it is difficult to get people to voluntarily participate in effective programs to treat mental health. Prendergast et al. (2015), Hall et al. (2017), Guydish et al. (2011), and Scott and Dennis (2012) show that attempts to incentivize participation in community treatment (using financial incentives and intensive case management programs) have not worked. Because it is a requirement of probation, mandated mental health treatment does not suffer from those same implementation difficulties.

When mental health treatment is mandated as a term of probation, it can take many forms. The judge might specify the detailed conditions of treatment, but often they are left to the probation officer to decide. The required components of the treatment commonly consist of an evaluation, counseling sessions, and possible referral to a psychiatrist for medication. Probation officers will often help connect probationers to therapists who work with patients in the criminal justice system. Those therapists use a range of treatment modalities, including but not limited to motivational interviewing, stages of change, cognitive behavioral therapy, reality therapy, Gorski's model of relapse prevention, and moral reconation therapy (Melissa Enoch, personal communication, May 26, 2023). Following the initial evaluation with a licensed practitioner, counseling sessions are generally weekly for the duration of probation, which lasts between six months and a year on average. About half are taking depression, anxiety, or bipolar medications (Tremayne Butler, personal communication, June 16, 2023).

Judges mandate mental health treatment for about five percent of probationers in my sample. Importantly, offenders required to seek mental health treatment are also required to pay for that treatment themselves. Without Medicaid coverage, individuals would typically pay $100 to $300 per session (Melissa Enoch, personal communication, May 26, 2023). The financial pressure this applies could cause individuals to be more likely to violate their probation if they are unable to pay for treatment. If so, mandated treatment might be more effective in an environment in which individuals are not required to cover the cost of treatment by themselves.

The effect of mandated mental health treatment is ambiguous because of these contextual details. Countervailing forces, including the expectation that individuals cover the costs of the treatment and the threat of violating probation if they do not comply, could mechanically result in increased recidivism in response to mental health treatment mandates. In addition, some individuals may respond to mandated mental health treatment with more hesitance than if they were to seek out that treatment freely. On the other hand, the financial cost of mandated treatment could offer an additional deterrence against future crime among those who are able to pay, making the treatment more effective.

1.2.2 Mapping of North Carolina Court System to Empirical Strategy

This paper's empirical strategy exploits variation in assigned sentencing judges' propensities to mandate mental health treatment. There are several characteristics of the criminal court system within North Carolina that make it an appropriate setting for this research design. I first briefly describe the sentencing protocol North Carolina follows, which allows me to limit the analysis sample to only those offenders who will receive a non-prison sentence. I then discuss the two court levels and how judges are randomized within them.

During the time period of this analysis, North Carolina followed a protocol known as structured sentencing. Appendix Figures A.1 and A.2 depict how, under structured sentencing, different combinations of offense severity and prior convictions determine the class of punishment and the minimum and maximum sentence length. Using this protocol, I restrict the analysis sample to only those who, based on their charged offense, would have been assigned to punishments other than an active prison sentence. The restriction is based on the charged offense rather than the convicted offense to avoid selection on an outcome, since judges could potentially manipulate the specific convicted offense in order to achieve a certain sentence level. By making the restriction to non-active sentences, I can identify the effect of mental health treatment during probation compared to being on probation but not receiving mental health treatment. The estimated effect does not include comparisons between offenders sentenced to probation and those sentenced to prison.[6]

Judge randomization in the North Carolina criminal court system has been used for identification in previous work by Sloan et al. (2013), Sloan et al. (2016), and Silveira (2017). Because the randomization differs between the two court levels, I first provide a brief overview of the different courts. The North Carolina court system is broken into District and Superior Courts, which differ in the severity of crimes they oversee. The District Court hears cases involving misdemeanors and infractions, whereas the Superior Court hears cases involving felonies and appeals. District Court trials are before a judge with no jury, whereas Superior Court trials are before a judge and a jury of twelve. Counties are formed into groups called districts in both courts, and all cases in a district are tried in the county seat. As Appendix Figure A.3 shows, by 2009, North Carolina was divided into 42 District Court districts (served by 270 judges) and 50

[6]There is a third type of sentence in the United States, parole, which differs from probation in that it is assigned to offenders after they have served a prison sentence, often in conjunction with an early release as a reward for good behavior. The Structured Sentencing Act also eliminated parole, so the only offenders serving a non-active sentence are those on probation.

Superior Court districts (served by 95 judges).

At the Superior Court level, the exogenous variation used to create the judge instrument comes from the way judges are rotated between districts. In addition to the grouping of counties into districts, the Superior Court further groups districts into circuit. The map in Appendix Figure A.3 shows the eight circuits present in 2009. Every six months, judges within a circuit move from one district to the next based on the rotation pattern of their circuit.

In the District Court, judges are elected directly to the districts and do not rotate between them. Instead, the exogenous variation for the judge instrument comes from the random assignment of judges into the schedule of trials. The chief judge in the district assigns a judge for each slot from among those available. Each chief judge prepares the schedule for three-month periods (January-March, April-June, July-September, October-December) at a time and aims to assign a judge to the same courtroom for the entire week.

Following the judge assignment mechanism, I create the instrument conditional on court-time fixed effects. Within the Superior Court these are circuit by district by year indicators. Within the District Court they are district by year by day of week by shift indicators to address assignment of judges based on schedule availability. Thus in the Superior Court the strategy compares individuals assigned to different judges due to which judge happened to be stationed in their district at the time they were sentenced, while in the District Court the strategy compares individuals assigned to different judges due to the scheduling of cases by the chief judge. I describe the creation of this measure in detail in Section 1.4.

Judges interact with the offender for the duration of the trial, which varies between District and Superior Courts. District Court sentencing trials take only a few hours, while Superior Court sentencing trials can last between one and three days because of the added complexity of the jury trial (Kurtz-Blum, 2021). Even in cases with a jury, only judges determine the imposed sentence, subject to the criteria defined by the structured sentencing guidelines.[7] Judges also interact with offenders when their probation is revoked. In this case, offenders are randomly assigned a judge for their revocation court appearance. In the analysis sample, I identify about 13 percent of offenders who fail their probation.[8] To verify the random assignment of judges in probation violation cases, I calculate the number of offenders who are assigned the same judge for the probation revocation as they are for their original court appearance. I then test whether the frequency of same-revocation-judge assignment (8.9 percent) is different from the probability of being randomly assigned any one judge at the court-district level (7.7 percent). The one-sided test that offenders are more likely to be assigned to their original judge upon revocation has a p-value of 0.28, suggesting offenders are not more likely to be assigned their original judge. This supports the argument that revocation judges are indeed randomly assigned.

1.3 Data

The empirical analysis uses data from the North Carolina court system. Criminal records come from the Administrative Office of the Courts and allow for identification of judge assignment of mental health treatment as well as any future interactions with the criminal court system. Those records are matched with judge schedules compiled from archived documents, which provide necessary information to form the instrument. The resulting dataset is crucial for this project as it provides time variation, enough information to characterize a mental health intervention, and

[7] The exception is cases in which the death penalty is an option, at which point the jury is included in the sentencing decision. Because this study focuses on only cases for which the charged offense is mild enough to limit offenders to a non-prison sentence, it is very likely that judges are always the ones determining the sentence.

[8] This number is likely a lower bound due to data quality concerns with that variable.

a source of exogenous variation. The rest of this section summarizes the most relevant information about the data; more detailed information about collection and cleaning can be found in Appendix A.2.

The North Carolina Administrative Office of the Courts (AOC)'s Automated Criminal Infractions System (ACIS) provides information about all District and Superior Court criminal cases. For both courts, the data have information at the case, charge, and defendant level. The charge-level data provide the initial charged offense, the outcome of each charge, and the punishment for guilty charges. Using North Carolina's offense grading system, I assign each charged offense a severity and characterize each case based on its most severe charged offense. The case-level data provide initial arrest date, trial dates, sentencing date, type of attorney, points incurred from prior convictions, and judge presiding over the sentencing case. The defendant-level data provide demographic information, including sex, race, ethnicity, date of birth, and address. Using the defendant's name and date of birth, I create links between cases. I can track their prior and future offenses that occur within North Carolina between 1994 and 2009. This includes charges in different counties and jurisdictions from the current case.

The ACIS identifies the judge initials associated with each sentencing trial. I supplement that with additional information about judge tenures, schedules, and districts using archived documents from the NC Judicial Department. Using the schedules and district maps, I assign judges to their circuit (Superior Court) and district (Superior and District Courts). This information is necessary to form the instrument, which relies on variation in judges within districts.

The main outcome of interest, recidivism, measures whether the offender reappears in the records within three years of the conviction. This measure of recidivism is not specific to incarceration or probation – anyone who enters the criminal justice system again within three years will be identified. In additional analysis I restrict the definition of recidivism to only consider new charges unrelated to probation violations. Doing so addresses the potential mechanical effect of mental health treatment on returning to the criminal justice system due to violations related to not attending treatment or not being able to pay. Defining recidivism within three years covers the period of time during which recidivism is most likely while also balancing sample size for power. As Figure 1.1 shows, recidivism is most likely within the first year, and recidivism in the first three years makes up about 75 percent of all recidivism that will occur within ten years of the current conviction. In further analysis I investigate how the impact of mental health treatment changes as the definition of recidivism incorporates more years post-conviction.

Crucial to the analysis in this paper, the ACIS data include a field for special conditions of the case which contains additional details about the punishment assigned to guilty charges. I use this field to characterize aspects of the probation sentence, including whether the defendant was required to seek mental health treatment or to seek substance use disorder (SUD) treatment. Although substance abuse is often co-occurring with poor mental health, the two diagnoses are viewed separately by much of the psychology and criminal justice literature. For example, drug courts and mental health courts are two separate potential interventions that target different severely-at-risk individuals in the criminal justice. Those categorized as receiving mental health treatment in my data might also receive SUD treatment, but must receive some kind of treatment that is specific to mental health. In additional analysis I consider alternative definitions of treatment which use a more restrictive definition of mental health treatment or broaden the definition to include more types of treatment.

I observe whether defendants are mandated mental health treatment, but not whether they comply with that mandate. I also do not observe whether defendants seek mental health treatment outside of the mandate. These factors both potentially bias the estimated effect of treatment toward zero by either adding non-treated individuals to the treatment group or by adding treated individuals to the control group. I can use the special conditions field to identify cases that later violate probation and the judges associated with that violation, which identifies some of the individuals who do not comply with the mandate. In Section 1.5 I show suggestive evidence that indeed, offenders who are assigned mental health treatment are slightly more likely to violate probation, and that my estimate of the effect of mental health treatment is larger when that is adjusted for.

1.3.1 Sample Characteristics

I make two main restrictions to the data when forming the analysis sample. The first requires all defendants to have key information, crucially a judge and information about that judge's scheduled district and circuit. The second restriction limits the sample to those individuals whose only option for punishment was probation. Following the Structured Sentencing laws, offenders are assigned points based on the number of prior sentences and the severity of their charged offense. Those laws then classify defendants into punishment classes based on their points. Certain defendants are only eligible for an active sentence, meaning time in prison; others are only eligible for a community or intermediate sentence, meaning time on probation. I keep only those individuals who were not eligible for an active sentence, representing about 90 percent of misdemeanors and 60 percent of felonies. I also limit cases to those seen during the years the Structured Sentencing laws were in place, from October 1994 to December 2009. Restricting to probation cases in this way, the analysis sample represents about 70 percent of all cases seen during this time period.

The restriction to non-active sentences potentially changes the makeup of individuals in the sample. Appendix Table A.4 details these differences, providing summary statistics for the sample before and after implementing the restriction. The sample of probationers looks similar in most ways to the overall sample of criminal court offenders. The most notable difference is that probationers have committed less severe crimes, particularly fewer violent crimes, and have shorter sentences.

Table 1.1 provides descriptive statistics for the main estimation sample, restricted to observations with demographic and criminal history controls. The sample consists of about 720,000 case-level observations. I divide the sample into two groups based on whether they were mandated to seek mental health treatment. The table shows that those who were mandated to seek mental health treatment (Column 2) were about one third less likely to recidivate in the three years following their sentence. Though this provides suggestive evidence that mandated mental health treatment could reduce future crime, there are clear imbalances in the comparison groups that could lead to bias in a naive comparison of means. The group assigned mental health treatment differs in their demographic characteristics: they are less likely to be Black and slightly more likely to be female. They differ in their criminal history: they are more likely to be first-time offenders and have fewer prior arrests. They also are more likely to have private attorneys and to be assigned SUD treatment.

The most common type of crime among probationers is traffic and public order offenses, as Figure 1.2 shows. Figure 1.3 shows that mental health treatment is also most frequently assigned among traffic and public order offenses, followed closely by drug and alcohol offenses.[9] It is

[9]Figures 1.2 and 1.3 only report the main offense. For example, if drug and alcohol abuse is more commonly a

least likely among financial and fraud offenses - only about two percent are assigned treatment there, compared to almost eight percent among traffic and public order offenses. Treatment is also more common among sex offenders and offenses related to domestic violence. These differences illustrate the potential need for a source of exogenous variation to form a comparison that is not selected on characteristics, like severity of offense, that might be correlated with the likelihood of recidivism.

1.4 Method

This paper estimates the effect of judge-mandated mental health treatment on future criminal activity. A naive method to estimate that effect would use the baseline regression given by Equation 1.1, where Y_{ict} is the outcome of interest for individual i in case c in year t.

$$Y_{ict} = \tilde{\beta}_0 + \tilde{\beta}_1 \texttt{MHT}_{ict} + \epsilon_{ict} \tag{1.1}$$

Judges decide whether to mandate mental health treatment based on information they have about defendants at the time of the trial. If the information they use is correlated with future criminal activity, then the simple regression based on Equation 1.1 would produce a biased estimate of the effect of treatment. Controlling for a rich set of demographic characteristics and the criminal history of defendants would recover the causal effect of treatment on recidivism if those characteristics made up the full set of information the judge used to make their decision. In practice, I will show that indeed, the coefficient $\tilde{\beta}_1$ changes in magnitude and the adjusted R^2 increases considerably when a vector of controls $\mathbf{X}_{ict}$ is included in the regression. However, I address concerns that even after controlling for those characteristics, judges might still have more

secondary offense, could be contributing to higher rates of mental health treatment among other offenses like public order violations.

information than the data provides.[10] To identify the causal effect of mandated mental health treatment on future recidivism, I instrument for $\texttt{MHT}_{ict}$ with $\texttt{JudgeLikelihood}_{ict}$, the judge's propensity to assign treatment among other offenders.

I estimate a specification following Equation 1.2, replacing the actual mental health treatment sentence with the predicted sentence $\widehat{\texttt{MHT}}_{ict}$ from Equation 1.3. Assuming judges are monotonically more lenient across all characteristics of convicted individuals, this strategy identifies the local average treatment effect (LATE), which is the causal effect of mental health treatment for those on the margin of being sentenced to treatment as a term of probation.

$$Y_{ict} = \beta_0 + \beta_1 \widehat{\texttt{MHT}}_{ict} + \beta_2 \mathbf{X}_{ict} + \gamma_{ct} + \varepsilon_{ict} \tag{1.2}$$

$$\texttt{MHT}_{ict} = \alpha_0 + \alpha_1 \texttt{JudgeLikelihood}_{ict} + \alpha_2 \mathbf{X}_{ict} + \tilde{\gamma}_{ct} + \epsilon_{ict} \tag{1.3}$$

In Equations 1.2 and 1.3, exogenous variation comes from judge assignment within districts, as discussed in detail in Section 1.2. Accordingly, all analysis controls for court-time fixed effects (γ_{ct}). Within the Superior Court those are circuit-district-year fixed effects; within the District Court they are district-year-shift-day-of-week fixed effects. Various specifications control for the same demographic and criminal history characteristics of the defendant ($\mathbf{X}_{ict}$) as in the OLS specification described earlier. Those characteristics include race, ethnicity, sex, a polynomial in age, region of the United States, type of attorney, and indicators for: first-time offender, a prior offense in the last year, offense type (violent and property, financial and fraud, traffic and public order, drugs and alcohol, and miscellaneous), and a sex offender. Certain specifications also include county-level controls for the per capita number of psychologists, emergency visits due to mental health, Medicaid and CHIP recipients, residents without insurance, residents with

[10] In particular, the data do not have information about the individual's mental health, which could both influence the judge to assign mental health treatment and make recidivism more likely.

bachelor's degree or more, residents living in poverty, and violent crimes committed. Because counties are subsets of court districts, those controls allow for exploration of differential access to care among individuals with the same set of possible judges as a potential mediating factor in the effect of treatment.

1.4.1 Multiple Aspects of Sentence

Judges have discretion over more than the assignment of mental health treatment. This introduces concerns about the exogeneity of the instrument in the empirical design. If the instrument, which is measuring the judge's underlying tendency toward mental health treatment, is correlated with the judge's tendency toward assigning other aspects of the sentence, then movements in the instrument will affect recidivism through more channels than solely mandated mental health treatment.

As described in Section 1.2, situating this analysis during the period of time when North Carolina used Structured Sentencing rules out many areas over which judges might usually have discretion. The decision to incarcerate, for example, is governed by the severity of the convicted offense and the number of previous convictions. To assuage concerns that judges might still indirectly affect whether offenders are incarcerated by manipulating the convicted offense class, I test whether judges who differ on mental health treatment assignment differentially manipulate the convicted offense. This test regresses either an indicator for whether the charged and convicted offense are the same or an indicator for whether the charged and convicted offense result in the same sentence class on the judge's propensity to assign mental health treatment. The estimates suggest that having a judge who is more or less inclined toward mental health treatment is not correlated with manipulation of the sentence class thresholds. Nevertheless, analysis is instead based on the charged offense, which is determined earlier in the judicial process by the arresting officer and at times the prosecutor. The conviction dimension is explored further in Section 1.5.6.

The Structured Sentencing laws also govern the length and intensity of probation. For example, the least severe offense class (Class 3 Misdemeanor with no prior points) restricts sentences to "community" (non-supervised) probation for one to ten days. The most severe offense class among those that do not have the possibility of incarceration (Class I Felony with 5-8 prior points) restricts sentences to "intermediate" (supervision by a designated probation officer) probation for four to eight months. Conditional on sentence class, judges do have minimal discretion over the exact length of probation within the requirements. The magnitude of that discretion changes depending on the sentence class: for the least severe sentence, the most judges could do is cause the offender to be under unsupervised probation for nine more days. Other sentence classes are characterized by wider ranges of allowed length. Following a similar strategy as in the case of testing potential manipulation of the convicted offense classes, I find that judge tendencies toward mental health treatment are not correlated with the actual sentence length assigned. In addition, some sentence classes allow for either community or intermediate (unsupervised or supervised) probation. Testing the relationship between judge tendencies toward mental health treatment and sentence class results in an estimated correlation of -.03 (standard error .052). These results support that judge tendencies toward mental health treatment are not correlated with whether probation is supervised or unsupervised.

On the other hand, judges have open discretion over assigning other types of treatment programs. In particular, judges can require a convicted individual to seek treatment for drug abuse and addiction. A damaging relationship with drugs is very common among individuals in the criminal justice system. Using 2007-2009 National Inmate Surveys, Bronson and Berzofsky (2017) estimate that 58 percent of state prisoners in the United States and 63 percent of sentenced jail inmates met the criteria for drug abuse. In comparison, the rate of drug abuse in the general population at the time was five percent. The previous literature evaluating programs that address drug dependency in the criminal justice system is mixed. It provides suggestive evidence of improvements in recidivism rates as a result of the introduction of drug courts, which pull those with drug dependency issues into a separate court process that focuses on rehabilitation. Very little exploration has been done on the cost efficiency of such courts, which can be quite expensive to implement and run (Doleac, 2023).

In the context of this paper, drug programs present an empirical challenge in that judges have discretion over both assigning mental health treatment and assigning substance use disorder (SUD) treatment. To address the challenge I build on the literature on the identification and estimation of treatment effects in the presence of multiple treatment alternatives (e.g. Carneiro et al. (2011), Mogstad et al. (2021), Heckman and Pinto (2018), Lee and Salanié (2018), Mountjoy (2022)). My approach most closely follows work by Humphries et al. (2023) on identification of multiple treatments in the judge instrument setting. Carneiro et al. (2011) first pointed out that in the case of multiple instruments, not controlling for the other instruments incorporates their effect on the outcome into the estimator if they covary with the focal instrument. Mogstad et al. (2021) show that using one instrument while conditioning on the others preserves the Imbens and Angrist (1994) interpretation of LATE as a weighted average over compliers. They consider a scenario with multiple instruments for one treatment and introduce partial monotonicity, which requires that monotonicity be satisfied for each instrument holding the others constant. Humphries et al. (2023) expand that result to multiple endogenous treatments, showing that depending on the model of judge decision-making, certain margin-specific treatment effects can be identified. In their setting they impose what Mountjoy (2022) calls unordered partial monotonicity, which combines Mogstad et al.'s partial monotonicity with a version of Heckman and Pinto's (2018) unordered monotonicity. Unordered monotonicity requires that shifting an instrument moves all agents uniformly toward or against each possible choice.

Adapting the notation of Humphries et al. to the setting of this paper, define T_t as a binary indicator of assignment to treatment $t \in \{M, D\}$ where M refers to mental health treatment and D refers to substance use disorder treatment. Consider the continuous judge propensity instruments z_t for each treatment. Then unordered partial monotonicity with respect to mental health treatment can be defined as:

$\forall\, z_M, z'_M, z_D \in supp(Z)$ and fixing z_D, either (1) or (2) is true:

1. $T_M(z'_M, z_D) \geq T_M(z_M, z_D)$ and $T_D(z'_M, z_D) \leq T_D(z_M, z_D)$

2. $T_M(z'_M, z_D) \leq T_M(z_M, z_D)$ and $T_D(z'_M, z_D) \geq T_D(z_M, z_D)$

Because unordered partial monotonicity is margin specific, $UPM(z_M|z_D)$ refers to fixing the judge's propensity toward SUD treatment and varying the propensity toward mental health treatment. The $UPM(z_M|z_D)$ condition requires that moving from a judge who is less likely to assign mental health treatment to a judge who is more likely, holding constant the judge's SUD treatment propensity, weakly induces some individuals into mental health treatment. In addition, the second part of the assumption requires that increasing the judge's mental health treatment propensity does not move individuals out of mental health treatment. Taken together, these result in a restriction that induces complier flows into only one treatment. Notice that $UPM(z_M|z_D)$ is not the same as $UPM(z_D|z_M)$; it is possible that only some margins will be identifiable. The implication of this assumption, as Mogstad et al. (2021) and Carneiro et al. (2011) also stress, is that z_D must be conditioned on in Equations 1.2 and 1.3 to identify the effect of mental health treatment on recidivism from variation in z_M.

Several models of judge decision making could be consistent with unordered partial monotonicity, and many models would not be. For example, if the judge were to make each decision completely independently without relying on any of the same information for either choice and the necessary information for each choice were unrelated to that for the other, then unordered partial monotonicity would trivially hold because T_M would only depend on z_M, T_D only on z_D, and z_M and z_D would be uncorrelated. In this setting, that is unlikely to be an accurate model of how the judge decides on sentencing. However, here I describe a model of judge decision making, based on Vytlacil (2002) and informed by characteristics of the North Carolina judicial system and the data, which is consistent with the $UPM(z_M|z_D)$. For some functions π_t for treatment $t \in \{M, D\}$, where π_t are measurable and nontrivial functions of z_t,

$$T_D = \mathbb{1}\{\pi_D(z_D) \geq U_D\} \tag{1.4}$$

$$T_M = \mathbb{1}\{\pi_M(z_M, z_D) \geq U_M\} \tag{1.5}$$

In this latent index selection model, consider two sources of unobserved-to-the-researcher factors, U_D and U_M. The offender's relationship with drugs is summarized by U_D and the offender's mental health status is summarized by U_M. When a judge determines an offender's sentence, she first evaluates whether that offender has a drug dependence issue. If not, she has no reason to assign SUD treatment; if so, she has some likelihood of assigning treatment based on the intensity of the evidence of drug dependence. After making this determination, she moves on to consider other potential terms of probation. Information on drug abuse is likely relatively easy for the judge to learn based on the facts of the case and interaction with the offender in the courtroom, but information about the offender's mental health status might be harder to accurately determine. Because of that, the judge makes her decision about recommending mental health treatment based both on her perception of the offender's mental health and drug abuse. This model allows for a situation in which the judge views an offender with a history of drug abuse, decides that offender is more likely to have a mental illness, and orders them to seek mental health treatment. Holding z_D constant and varying z_M moves individuals across only one threshold, inducing compliers into or out of mental health treatment. Thus this framework is consistent with $UPM(z_M|z_D)$. However, holding z_M constant and varying z_D could induce movement along multiple dimensions, since both T_D and T_M depend on z_D. Thus this framework does not support $UPM(z_D|z_M)$, and the causal effect of SUD treatment, holding mental health treatment constant, cannot be identified from the model.

Unordered partial monotonicity results in testable implications. In particular, the characteristics of individuals who are assigned SUD treatment, once controlling for the judge's SUD treatment propensity, should not vary by the judge's mental health treatment propensity. This can be seen in Equation 1.4, where being assigned SUD treatment only depends on the judge's drug propensity, not on the judge's mental health propensity. Following Humphries et al. (2023), I test this implication by first predicting recidivism using a vector of individual characteristics and then regressing that predicted recidivism on the judge's propensity toward mental health care, controlling for the judge's propensity toward SUD treatment, in the subset of the sample who received SUD treatment. The p-value for the test of whether mental health treatment is predictive of the characteristics of those assigned SUD treatment is 0.84. The results of this test suggest that the mental health treatment propensity is not related to the individual characteristics that predict recidivism among SUD treatment recipients. The test supports the $UPM(z_M, z_D)$ assumption holding in the context of this study.

To summarize, the approach here relies on the independence of T_D from z_M. Then conditioning on z_D when estimating the effect of T_M instrumented with z_M identifies the local average treatment effect of mental health treatment, following Mogstad et al. (2021), Carneiro et al. (2011), Humphries et al. (2023), Mueller-Smith (2023), and others. Notice that T_D is in the error term and is correlated with T_M because both depend on z_D. Controlling for z_D breaks that correlation. This exercise can be thought of in the framework of Haavelmo causality: the causal parameter is defined using hypotheticals that fix all other inputs and vary one to determine outcomes (Haavelmo, 1943, 1944; Heckman and Pinto, 2015; Pearl, 2015). Because judge assignment of mental health treatment and SUD treatment are endogenous, there is no model of a natural experiment that varies mental health treatment assignment without also varying SUD treatment assignment. However, holding the judge's SUD propensity constant allows for conceptualizing the estimated effect of mandated mental health treatment as the partial derivative of a structural model.

1.4.2 Instrument Creation

This paper follows a growing literature that constructs an instrument based on exogenously assigned adjudicators' decisions on other cases during the same period. In this section I first summarize the general method the literature has taken to creating this instrument, then describe the changes this study makes to incorporate recent theoretical advances in the literature.

The instrument is a measure of the tendency for a randomly assigned judge to assign individuals to some outcome or group. In other papers that has been pretrial detention, incarceration, foster care homes, and more.[11] In the context of this paper, the judge assigns individuals to seek mental health treatment as a term of their probation sentence. The previous judge instrument literature commonly uses the phrase "judge leniency," but in this context it is less clear what is more lenient and what is more strict. Assigning mental health treatment could be seen as more strict, giving offenders more requirements they have to satisfy in order to fulfill probation. It could also be seen as more lenient, giving offenders a chance to seek help in changing their behavior rather than trying to work through it alone. I instead refer to the instrument as the "judge propensity" or "judge tendency" toward mental health treatment.

The measure is constructed by first residualizing the relevant judge decision on the vector of court-time fixed effects that govern the randomization of the judges. As described in Section 1.2, the court-time fixed effects ensure comparison of offenders who were at risk of being assigned to the same possible judges. In the Superior Court, they include the circuit (to which the judges are elected), the district (where they serve, and are rotated between), and the year. In the District Court, they include the district (to which the judges are elected and where they serve), year, day of week, and shift. The resulting residual is averaged over all cases seen by that judge, excluding the individual's case. Finally, the average among all judges is subtracted from the average for that judge, so that the resulting instrument variable measures how much a given judge deviates from the average judge's sentencing behavior. In the context of this paper, positive values of the instrument characterize judges who are more likely to assign mental health treatment. The

[11] See, for example: Kling (2006), Dobbie et al. (2018), Leslie and Pope (2017), Kalish (2023), Belloni et al. (2012), Autor and Houseman (2010)

leave-out measure of judge propensity is constructed by regressing the judge's sentence ($\texttt{MHT}_{icj,t}$) on the vector of court-time fixed effects ($\gamma_{c,t}$) in Equation 1.6 and then taking the mean of the residual, subtracting any cases related to that individual, in Equation 1.7. Here $n_{j,t}$ refers to the number of cases the judge oversees in a year and $n_{ij,t}$ refers to the number of cases the judge oversees for individual i.

$$\widehat{\texttt{MHT}}_{icj,t} = \texttt{MHT}_{icj,t} - \gamma_{c,t} \tag{1.6}$$

$$\texttt{JudgeLikelihood}_{ij,t} = \frac{1}{n_{j,t} - n_{ij,t}} \left(\sum_{k=0}^{n_{j,t}} \widehat{\texttt{MHT}}_{kj,t} - \sum_{l=0}^{n_{ij,t}} \widehat{\texttt{MHT}}_{lij,t} \right) \tag{1.7}$$

This describes the basic form of the judge instrument. Section 1.5.6 discusses alternative forms of the instrument that address various potential concerns and incorporate recent literature exploring the exogeneity and homogeneity properties of the judge instrument. For the main specification, following Leslie and Pope (2017) and others, I calculate the residualized mean separately by felony and offense group.[12] Calculating the instrument within types of crime relaxes the monotonicity assumption, which requires that if one judge is ranked as more likely to assign mental health treatment than another judge, that ranking will hold true for all offenders regardless of their individual characteristics. The relaxed assumption instead requires that the judge ranking be preserved within, for example, all violent felonies, and within all public order misdemeanors, but not necessarily between those two groups. Section 1.5.6 discusses further relaxations of this assumption.

Figure 1.4 depicts a histogram of the main residualized measure. Positive values represent those judges who are more likely to assign therapy. The distribution is slightly skewed right, with a minimum value of -0.33, a maximum value of 0.98, and a standard deviation of 0.03.

[12]The offense groups are the same as those referenced in Figure 1.2: violent and property, financial and fraud, traffic and public order, drugs and alcohol, and miscellaneous.

This implies that moving from a judge three standard deviations below the mean propensity to a judge three standard deviations above the mean propensity increases the probability of mental health treatment by about 18 percentage points. Most judges assign mental health treatment fairly rarely, but very few never assign it. Figure 1.5 plots the percent of cases judges assign to treatment, organizing the judges by how often they assign mental health treatment. The 50th percentile judge assigns mental health treatment less than ten percent of the time. On the right tail of the judge distribution, there are judges who assign mental health treatment more than sixty percent of the time.

1.4.3 Balance

For this strategy to identify a causal effect, the classic instrument variable assumptions must hold. The judge tendency to sentence offenders to seek mental health treatment must be correlated with actual receipt of a mental health treatment mandate and not related to other factors that predict recidivism. In addition to the histogram of the instrument, Figure 1.4 also plots the residualized rate of therapy assignment against the judge propensity measure. It represents a graphical depiction of the first stage: as the propensity measure increases, judges are more likely to assign therapy. This provides evidence that the judge instrument is predictive of mental health treatment, and indeed the F-statistic from the first stage is quite large, over nine hundred.

The randomness of judge assignment is crucial to the validity of the estimation strategy. If judges are exogenously assigned, the instrument should not be related to other factors that predict recidivism, satisfying the exclusion restriction. In order to verify random assignment, I carry out three tests of whether judges with higher propensity are assigned offenders with different characteristics.

The first test provides evidence that the judge is a stronger predictor of the offender's mental health treatment sentence than of the observable characteristics of offenders with mental health treatment. I first form the predicted probability that an offender would be assigned mental health treatment on the basis of their demographic and criminal history characteristics and the court-time fixed effects. The predicted probability is then regressed on a full set of judge indicators, controlling for the court-time fixed effects. I compare the F-statistic from that regression to the F-statistic from a regression of the actual mental health treatment assignment on the set of judge indicators. The F-statistic for the predicted probability is 0.49 (p-value 0.48), implying that there are no differences in the ex-ante characteristics of offenders who appear before different judges. The F-statistic for the actual treatment is larger and significantly different from zero (p-value 0.01), which provides evidence that judges do matter for determining the offender's assignment to mental health treatment.

For the second test, I regress the judge propensity measure and mental health treatment assignment on demographic characteristics and criminal history, conditional on court-time effects, and test the joint significance of these controls. Table 1.2 presents the results of this test. The F-statistic for the mental health treatment mandate is 405, with p-value approximately zero. The large F-statistic shows that the individual characteristics are highly predictive of mental health treatment assignment. In contrast, the F-statistic for the judge propensity instrument is 2. The small size indicates that the offender's characteristics are not significant factors in predicting the judge's propensity. Though the associated p-value is also small (0.001), the size of the estimated coefficients on the characteristics that predict judge propensity are not economically significant. The two characteristics of concern are having a private attorney and being a sex offender.

Those characteristics increase the judge propensity toward mental health treatment by one one-hundredth and three one-hundredths of a standard deviation, respectively. In addition, many of the characteristics that are predictive of being mandated mental health treatment have no impact on the judge propensity measure. Black probationers are less likely to be sentenced to mental health treatment, while first-time offenders and female probationers are more likely. None of those characteristics are predictive of the judge's propensity to assign treatment. This test suggests that the measure of judge propensity is generally balanced across demographic and criminal background controls, but identifies the potential for some sorting across judges by sex offenders and individuals with private attorneys.

For the final test, I compare the first stage with and without including a standardized version of the judge instrument. Table 1.3 presents these results. The judge instrument is predictive of mandated treatment: having a judge with a one standard deviation higher propensity toward mental health treatment increases the likelihood of being assigned mental health treatment by about 2.4 percentage points. In contrast, the judge's tendency toward SUD treatment has a much smaller relationship with mandated treatment, increasing the likelihood by about 0.1 percentage points. Comparing the final two columns of the table shows that including the instrument has little effect on the predictive power of the controls, but does increase the R^2. Similarly, comparing the second and third columns shows that adding controls has little effect on the magnitude of the estimated relationship between the instrument and treatment, but does increase the R^2. Those comparisons suggest that the connection between the instrument and mandated mental health treatment is not related to the individual demographic and criminal history characteristics of the probationers.

Taken together, these three tests suggest that the process of case assignment to judges is conditionally random. However, because of the potential for sorting related to sex offenders and offenders with private attorneys, I control for the vector of observed demographic and criminal history characteristics described above in the main analysis specification. Section 1.5.2 provides evidence that the effect of treatment is similar among offenders of different demographic characteristics, including those with private versus public attorneys. That suggests any sorting on defendant characteristics is not driving the main estimate of the impact of mental health treatment on criminal activity.

1.5 Results

1.5.1 The Effect of Mental Health Treatment on the Incidence of Future Crime

Table 1.4 presents the main results for the sample that combines District and Superior courts.[13] I estimate Equation 1.2 instrumenting for mental health treatment sentence with the leave-out residualized measure of judge propensity to assign treatment. The outcome of interest is recidivism, measured as an indicator of ever returning to the criminal court system within three years of the current offense. I limit my main analysis to cases tried between 1994-2006 so that each offender has the chance to reappear in the data by three years since the trial. All analysis, unless otherwise stated, controls for court-time fixed effects and the judge's tendency toward SUD treatment. All reported standard errors are clustered at the court-time level, and F-statistics are robust (Kleibergen and Paap, 2006).

The first column of Table 1.4 presents a simple form of the baseline OLS specification.[14]

[13] In combined-court specifications, all explanatory variables are interacted with an indicator for court.

[14] The estimate from the first column (0.10) differs from the difference in means in Table 1.1 (0.12) because it controls for court-time fixed effects.

Omitted variables could bias the estimated effect of mandated mental health treatment in either direction. On the one hand, judges might only send individuals to therapy if they seem to have a good chance of using it to their advantage, biasing the estimate away from zero (positive bias). On the other hand, judges might only send individuals to therapy if they look to be struggling more, biasing the estimate toward zero (negative bias). In practice, controlling for demographic and criminal history characteristics in column two has a large impact on the magnitude of the estimated effect, dropping it in half from ten to five percentage points. This is consistent with judges choosing to assign mental health treatment based on observable characteristics that are associated with lower likelihood of recidivism, such as being a first time offender.

If the vector of included controls were to span the full support of the information judges use to assign mental health treatment, then the estimate in column two should identify the causal effect of mental health treatment on recidivism. However, instrumenting for mandated mental health treatment with the associated judge propensity results in estimates that are larger in magnitude than estimates from the OLS with controls specification, suggesting that the included controls did not encompass the judge's full information set. The larger magnitude of the IV estimates means that the unobserved factors were biasing results toward zero in OLS. This is consistent with judges choosing to assign mental health treatment to individuals who are more likely to have a mental illness, which is related to a higher risk of recidivism. Columns three through six present the estimate from the IV specification with various included control variables. Column six presents the main specification, which controls for the judge's propensity toward SUD treatment as well as the demographic, criminal history, and county characteristics described in Section 1.4. I find that being assigned to seek mental health treatment decreases the likelihood of being rearrested within three years by about 12.1 percentage points. For reference, 34 percent of the convicted probationers over the sample period returned to the court system within three years of being sentenced. Thus the effect corresponds to a 36 percent decrease in three-year recidivism relative to the mean.

Overall, these results suggest that there is a group of offenders for whom mental health treatment is an effective means of deterrence from future crime. The results are consistent with recent evidence on the interaction between behavioral programs, financial access to healthcare, and recidivism. Jácome (2020) finds that low-income young men who lose Medicaid eligibility are 22 percent more likely to be involved in the criminal justice system, and among men with mental illness that number rises to 40 percent. Arbour (2021) finds even larger effects in his study of a Canadian prison's cognitive behavioral therapy (CBT) program: a 58 percent decrease in recidivism compared to prisoners who do not participate in the program. Heller et al. (2017) identify a 20 to 30 percent decrease in crime due to a CBT intervention among low-income Chicago youth. The effect I estimate falls in line with those estimates, though the referenced studies are estimating shorter-term effects. They focus on recidivism during and in the months following the program, whereas my definition of recidivism includes years after the average probation sentence has ended. Heller et al. (2017), for example, find that the effect they estimate disappears after the program ends, whereas my estimate persists beyond the mandatory period of treatment.

The results from this study are also consistent with studies of alternative courts for individuals with mental health or substance abuse concerns. A review of 27 drug court evaluations by the GAO (2005) found rearrest reductions of about 10 to 30 percentage points. An evaluation of the Calgary Diversion Program, a mental health court, by Mitton et al. (2007) found that it was correlated with decreased court reappearances of between 25 to 48 percent. Both programs focus on different contexts than that of this paper. Individuals routed through drug courts do not necessarily have a mental illness, whereas the Calgary Diversion Program required individuals to have a serious mental illness. In contrast, the mental health of the population studied by this paper likely falls somewhere in between those of the drug court and mental health court. In addition, this paper specifically studies the treatment of mental health, rather than the different court processing of those with mental illness. Nevertheless, the cited studies provide an anchor to which to compare this study's estimates, and show that they are in line with previous work on mental and behavioral interventions in the criminal justice system.

The main estimate should be thought of as an intent-to-treat estimate in the sense that there is no information on how many offenders assigned to mental health treatment follow through and take up treatment. For a rough proxy of whether the offender follows through with treatment, I estimate the effect of mental health treatment on whether the offender fails probation in Table 1.5. Probation can be violated for many other reasons, such as failing to pay court fines, traveling somewhere forbidden, or committing another crime. And reported violations are in part determined by probation officers, who differ in their strictness. With those caveats, I find suggestive evidence that offenders sentenced to mental health treatment are more likely to fail their probation. Indeed, when I estimate the effect of mental health treatment on three-year recidivism where I instead treat probation violation cases as *not* committing a future crime, my estimate is slightly larger. I discuss probation violations further in Section 1.5.6.

1.5.2 Heterogeneity in the Effect of Mental Health Treatment on Recidivism

I explore heterogeneity in the treatment effect by estimating two-stage least squares regressions among various subsamples defined by demographic characteristics, criminal history, and characteristics of the court district. In general, my results are suggestive of little treatment effect heterogeneity. Offenders with many different characteristics benefit similarly from mandated mental health treatment.

Figure 1.6 plots the estimated effect of mandated mental health treatment on three-year recidivism within each group of individual characteristics. I find negative effects of treatment on recidivism for most subgroups, though some estimates are imprecise. The point estimates range from a decrease of about 10 percentage points to a decrease of about 20 percentage points. The estimates provide suggestive evidence of slightly larger effects among repeat offenders, women, younger offenders, and offenders with a public attorney, but the estimates are not statistically different from each other. The estimate of the effect of mental health treatment among Hispanic offenders is particularly imprecise, with a confidence interval that spans from a decrease of 45 percentage points to an increase of 15 percentage points. The imprecision is most likely due to the small number of observations - only about two percent of the sample is Hispanic, consistent with the demographics of North Carolina in the early 1990s.[15] Overall, the results in this figure suggest that mental health treatment is similarly effective among many types of individuals.

The results regarding the heterogeneous effects of mental health treatment among types of offenders differ from previous literature in one particularly notable way. Some prior work (see,

[15]While the population of Hispanics in North Carolina experienced enormous growth starting in the 1990s, only 1.2 percent of the population was Hispanic in the 1990 Census. That number had increased to 4.7 percent by 2000. (Census, 1990, 2001)

for example, Arbour (2021)) finds that first time offenders are more responsive to interventions that decrease recidivism. In contrast, in the context of mandated mental health treatment, the effect is not concentrated among first time offenders but instead proves similarly, or if anything more, beneficial for repeat offenders.

The homogeneity of the treatment effect provides some assurance that the judges are not being assigned based on characteristics of defendants. For example, one potential concern could be that first-time offenders, who are less likely to commit future crimes, are assigned certain judges who are more prone toward mandating mental health treatment. In that case, the effect of mental health treatment among first-time offenders should be larger; instead it is indistinguishable from the effect among repeat offenders. Similarly, defendants with a private attorney might be less likely to commit future crimes because they have additional support and more financial security. If their attorneys were able to bargain for a judge who is less prone to mandating mental health treatment, the effect of treatment would be biased toward zero. Instead, the estimated effect of mental health treatment is similar among defendants with private and public attorneys. Heterogeneity of the effect would not necessarily indicate a violation of the exclusion restriction, but the lack of heterogeneity at least rules out some violations.

I also explore whether different kinds of offenders are more or less impacted by the potential for mandated mental health treatment to result in a probation failure due to lack of attendance or inability to pay. Figure 1.7 shows that there is more evidence for heterogeneous treatment effects when the outcome is failing probation. The strongest evidence is in the comparison of private and public attorney. The positive relationship between mental health treatment and probation violations is driven entirely by individuals with a public attorney, who are almost 20 percentage points (confidence interval [10, 30]) more likely to fail probation as a result of being assigned mental health treatment. By contrast, being assigned mental health treatment has no effect on probation violations among individuals with a private attorney. This evidence is consistent with financial constraints being an important mediator in the implementation of mental health treatment. I explore further evidence of the importance of financial constraints in Section 1.5.3.

I find the most evidence of heterogeneous treatment effects among different offense types. Table 1.6 presents the results from splitting the sample by whether the individual was charged with a misdemeanor or a felony. The estimates suggest mandated mental health treatment is more effective among individuals who were arrested for more serious crimes. The estimated effect of mental health treatment for those individuals is about 18 percentage points (50% of the mean among felonies), whereas the estimated effect among individuals who committed misdemeanors is only about 10 percentage points (30% of the mean among misdemeanors). Table 1.7 shows the results from estimating the main equation separately by the class of the convicted offense. These estimates suggest that mental health treatment is most effective among offenders who have been arrested for violent, property, or financial crimes. Treatment is also effective among offenders who have been arrested for drug and alcohol crimes. The particularly strong effect of mental health treatment among probationers charged with financial crimes could be explained by financial deterrence effects. Being required to pay for treatment may be seen as a particularly salient punishment to those who are convicted of financial crimes. In contrast, mental health treatment is much less effective among those who commit traffic and public order violations. I estimate an increase of 2.5 percentage points in recidivism among that offense group, but the estimate is not distinguishable from zero. Overall, these results show that individuals charged with more severe offenses experience more of a benefit from mental health treatment, especially when the crimes are felonies.

1.5.3 The Effect of Mental Health Treatment on Recidivism over Time

The main results focus on recidivism within three years. To understand how the effect of mental health treatment evolves over time, I explore several specifications. The results suggest that the effect of mental health treatment persists in the first three years after trial, during which offenders are at the highest risk of recidivism, and then dissipates. As Figure 1.1 shows, recidivism is most likely within the first year, and recidivism in the first three years makes up about 75 percent of all recidivism that will occur within ten years of the current conviction.

To understand how the effect of mental health treatment differs over time, I consider two measures of recidivism. One is cumulative: it characterizes recidivism up to x years after the trial. When $x = 3$, this corresponds to my main measure of recidivism. Panel A of Figure 1.8 shows the results from the cumulative measure looking up to five years after trial. Mental health treatment is similarly effective as time elapses, with point estimates between a six and twelve percentage point decrease in recidivism. The confidence intervals are large enough that the different estimates are not statistically distinguishable from one another. Appendix Table A.5 presents the point estimates corresponding to the figure. Appendix Figure A.5 presents a version of this exercise based on estimates from an OLS regression of recidivism on mandated mental health treatment. Although the magnitudes of the point estimates differ from the IV estimates (as discussed in Section 1.5.1), the OLS results show similar trends in the effect over time.

The relatively flat trend over time could be consistent with several different explanations, as it incorporates changes in the effect, in the frequency of the dependent variable, and in the sample.[16] In Appendix Figure A.4, I repeat the same exercise using a balanced sample of cases

[16] The sample changes because the measure of recidivism requires the trial to have been at least x years before the last year in the data.

from 2004 and earlier. The results are less precise, but the pattern is similar to that in the main sample, with estimates hovering around 10 percentage points. The comparison with the balanced panel suggests that the results are not driven by changes in the cases included in the sample. Turning to the frequency of the dependent variable, Panel B of Figure 1.8 shows the results as a percent of the mean in the relevant time period. In part because the cumulative rate of recidivism increases over time, the estimated effect in percent terms decreases over time from about 30 percent to about 25 percent of the mean.

To understand the mechanisms underlying the decrease in the effect in percent terms over time, I turn to a second measure of recidivism over time. This measure divides time after the trial into single-year periods, so that $x = 3$ would correspond to only recidivism between two and three years after the trial. The results from this exercise are shown in Figure 1.9. Panel A suggests that mental health treatment has the largest estimated effect on recidivism in the first year since trial, when the offender is significantly more likely to commit a new crime. The estimate in the first year since trial suggests mental health treatment reduces recidivism by about 9 percentage points. The results in the following two years are smaller and less precise, suggestive of a decrease in recidivism by about 3 percentage points for each successive year. However, as Panel B shows, the effect of mental health treatment measured as a percent of mean recidivism is quite similar across the first three years. The evidence from this exercise suggests that the effects of mandated mental health treatment persist for three years after the trial and dissipate after. Returning to the percent effects in Panel B of Figure 1.8, comparison to the second measure of recidivism suggests the persistence of the effect over five years is driven by reductions in recidivism in the first three years.

Because probation spans between five to nine months on average, the persistence of the effect during the first three years after trial suggests that treatment is effective even beyond the period during which it is mandated. It also suggests that mandated mental health treatment is not just shifting back when recidivism occurs, as would be the case if the estimated effect changed signs after the initial years. Lacking information on take up of treatment, I cannot say whether the persistence is driven by skills learned and a reduction in mental illness which continues after treatment ends, or instead by the continued use of treatment after the mandated period is over. Nevertheless, the persistence stands in contrast with evaluations of other behavioral programs which find that effects mostly disappear after the end of the program. Blattman et al. (2017) show that CBT effectiveness depreciates over time, and the cognitive-behavioral intervention among low income youth studied by Heller et al. (2017) lost effectiveness quickly after the youth exited the program. The interventions studied in previous research differ from the intervention studied in this paper in several ways, as discussed in Section 1.1. The focus on treating underlying mental illness, the freedom of individuals to choose a specific type of therapy, and the setting among individuals on probation could all contribute to the longer persistence of the estimated effect of treatment.

To understand potential mechanisms driving the effectiveness of mental health treatment over time, I compare the trajectory of the effect for individuals with private versus public attorneys. Viewing private attorney as a rough proxy for the offender having a higher socioeconomic status, these results are consistent with the effectiveness of treatment being related to access to resources. Figure 1.10 presents the percentage point and percent estimates over time for individuals with a private attorney and individuals with a public or waived attorney. The effect of mental health treatment in percentage points increases slightly over time for individuals with a private attorney, whereas for individuals with public attorneys the effect starts to diminish after three years since trial. When incorporating differences in the likelihood of recidivism over time for these two groups of offenders, the difference in the trends is amplified: the effect in percent terms diminishes more sharply over time for those with a public attorney, whereas it increases after the first year for those with a private attorney and then remains flat for the following four years. Offenders with private attorneys could be more able to follow through with treatment during probation and to continue treatment after the mandated period because of the monetary cost associated with treatment. Indeed, these results suggest that treatment remains effective for years after probation among individuals with a private attorney, whereas it diminishes in effectiveness over time for individuals with a public attorney. Although the point-in-time estimates are, for the most part, not statistically different across the subsamples, the different shape of the trends over time is consistent with the importance of financial access to mental health treatment as a moderator for the effectiveness of the treatment. This evidence is also consistent with a potential financial deterrence effect of the monetary cost of treatment. Individuals who can pay may nonetheless have a distaste for doing so, and may be incentivized to avoid future contact with the criminal justice system in order to escape future financial penalties.

1.5.4 The Effect of Mental Health Treatment on the Type of Future Crime

The main results focus on how mental health treatment affects the extensive margin of criminal behavior, measured by the appearance of a future criminal conviction. In this section, I consider various ways of characterizing the intensive margin of crimes committed in the future. I first exchange the dependent variable for a measure of how many crimes were committed within the three years following the trial. Table 1.8 presents these results. Including cases that do not recidivate, the average number of future crimes is about 0.5.[17] I estimate that being mandated mental health treatment results in about a 43 percent decrease in the number of crimes committed.

In addition to the number of future crimes, I evaluate the length and intensity of future punishment and the severity of future crimes. Table 1.9 presents the unconditional and conditional results for the effect of mandated mental health treatment on the average future sentence length and the likelihood of a future active sentence (incarceration). Consistent with the negative effect of mandated treatment on the extensive margin of recidivism, I show that it also reduces the overall length of sentences and likelihood of an active sentence, by 55 days (57%) and 6.2 percentage points (43%) respectively. Probationers assigned mental health treatment also appear to serve shorter sentences and fewer prison sentences conditional on having committed a crime, though conditioning on recidivism reduces the sample such that the estimates are imprecise. Turning to the characteristics of the offense, I estimate the effect of mental health treatment on different classes of crime in Table 1.10. These results are unconditional, and so incorporate both the overall effect of mandated mental health treatment on recidivism as well as any differential effects on separate types of crimes. I find that being assigned mental health treatment is most effective at reducing future violent or property and financial or fraud crimes (an 11 and 18 percentage point reduction, respectively). The estimates suggest that mental health treatment has a smaller effect on future traffic offenses, which are likely less impacted by underlying poor mental health. Finally, I estimate the effect of mental health treatment on future misdemeanors and future felonies in Table 1.11. I find that mental health treatment is more effective at reducing future felonies, with no statistically distinguishable effect on reducing misdemeanor crimes.

These results are consistent with evidence from previous studies of behavioral interventions

[17] Conditional on committing at least one crime in the future, the average number is 1.8.

that have identified effectiveness at reducing violent crimes (see, for example, Blattman et al. (2022)). Even more than violent crimes, mental health treatment is most effective at reducing financial crimes. Jácome's (2020) study of Medicaid provision among low-income young men came to a similar conclusion, showing that losing Medicaid eligibility lead to increases especially in financially-motivated crimes. Overall, my analysis of the effect of mental health treatment on the intensive margin of crime suggests that treatment is effective by reducing the number and severity of crimes in addition to the incidence of any crime.

1.5.5 Returning to the Contribution of Substance Use Disorder (SUD) Treatment

In the above analysis, I control for the judge's propensity toward SUD treatment in order to identify the margin-specific local average treatment effect of mandated mental health treatment. As discussed in Section 1.4.1, this approach follows Humphries et al. (2023), who build on Mogstad et al. (2021) and Carneiro et al. (2011), among others. Here, I investigate how important controlling for the drug propensity is in practice. Table 1.12 presents results analogous to Table 1.4, but without controlling for the judge's propensity toward SUD treatment. The specification including demographic, criminal history, and county controls estimates a 14.1 percentage point decline in recidivism as a result of being assigned mental health treatment. That estimate is about two percentage points larger than the main estimate controlling for drug propensity, though the difference is not statistically significant. The direction of the bias suggests that ignoring the dependence of mental health treatment assignment on the judge's propensity toward SUD treatment overestimates the effectiveness of mental health treatment. Estimates from the literature evaluating drug courts suggest that SUD treatment reduces certain types of future crime (Frank and McGuire, 2010; Mitton et al., 2007). Since mental health treatment and SUD treatment positively covary, the 14.1 percentage point estimate is likely picking up some bias from the benefits of SUD treatment in reducing recidivism. However, the bias is not large, consistent with the relatively small correlation between mental health treatment and SUD treatment.

1.5.6 Robustness in the Estimated Effects

I carry out several robustness exercises to ensure that the results are not sensitive to the choice of specification or sample. In this section I discuss those exercises, grouped into four categories: the design of the exogenous variation, the monotonicity assumption of the instrument, judge discretion over conviction, and the choices made when creating the analysis sample and forming the outcome and treatment variables. Overall I find that the results are not driven by these choices. Across all robustness exercises, the results show that mandated mental health treatment contributes to a lower likelihood of criminal activity.

1.5.6.1 Setting of Exogenous Variation

The design of the exogenous variation in this study relies on judges being assigned as-good-as-randomly at the court-time level. The randomization level varies between the District and Superior court systems. In the main specification I define the court-time fixed effects governing the randomization level as district-year-shift-day-of-week groups in the District court, and circuit-district-year groups in the Superior court. I consider two alternate definitions of the court-time level.

The baseline specification addresses the potential that certain judges are, for example, more available in the mornings, when one type of case happens to be frequently scheduled. The first alternative specification removes time variables, using only the district or circuit-district to form the groups. That would be consistent with no scheduling constraints that are correlated with certain types of cases or offenders. In the Superior court, judges are randomly assigned at the district level but based on a pool of judges determined by the circuit to which the district belongs. Accordingly, the baseline specification controls for the circuit as well as the district. The second alternative specification removes circuit, grouping both District and Superior courts by the district alone. Appendix Table A.6 presents the results of this exercise. Changing the court-time level, which affects both the variables on which the instrument is residualized and the level of clustering for the standard errors, has a minimal impact on the estimated effect of mental health treatment. Compared to a baseline estimate of 12.1 percentage points, the first alternative level results in an estimate of 13.9 percentage points, and the second alternative results in an estimate of 15.2 percentage points.

The construction of the instrument also depends on the amount of judge information incorporated. Following the previous literature on the judge instrument strategy, the baseline construction of the instrument uses information from all cases the judge sees over the sample period, removing information from cases related to each probationer. Here I discuss two groups of alternative formulations to address concerns about exogeneity and to incorporate recent methodological work on judge instrument design.

First, I address concerns that the judge's underlying propensity toward mental health treatment might be changing over time and that change might be correlated with the offenders the judge sees. I consider alternate formulations of the instrument that use only cases from the judge's first year, that omit cases from future years, that only use cases in the current year, and that only use cases in three-year groups. Appendix Table A.7 presents results from these specifications. Using only the judge's first year, the instrument is much weaker and the estimate is not distinguishable from zero. However, other approaches to creating the instrument result in strong first-stage relationships and similar estimates of the effect of mental health treatment on recidivism.

Second, following Frandsen et al. (2023), I create a cluster-jackknife version of the instrument, in which I leave out not only the individual's case but all other cases in the individual's randomization cluster (for example, all cases in District 3 in 1997 on Monday mornings in the District Court). This addresses concerns that the types of individuals grouped into that cluster might differ in meaningful ways from the types of individuals in another cluster, and that those differences would be incorporated into the instrument, injecting endogeneity. Using this instrument, I estimate a 12.7 percentage point decrease in recidivism. The slightly smaller F-statistic compared to my baseline specification, and the point estimate that is slightly more different from the OLS estimate do not rule out that the baseline instrument might have been biased by the inclusion of variation in the types of individuals in a cluster, but the similarity of the estimates suggests that it was not a large source of bias.

1.5.6.2 Monotonicity of the Instrument

The validity of the empirical design for identifying a local average treatment effect (LATE) relies on the assumption of monotonicity (or homogeneity) of the instrument. Section 1.4.1 characterized a type of monotonicity that allows for the identification of margin-specific treatment effects, holding constant the judge's propensity toward SUD treatment. In this section, I discuss how the monotonicity assumption changes when incorporating control variables.

The instrument variable strategy identifies the margin-specific local average treatment effect, which is the causal effect of mental health treatment for those on the margin of being sentenced to treatment as a term of probation, holding constant the judge's propensity toward treatment. However, when the specification includes covariates, the LATE interpretation does not necessarily apply. As Blandhol et al. (2022) note, the linear IV instead estimates treatment effects for both compliers and always-takers, where some always-taker effects are negatively weighted. Two-stage least squares with covariates produces an average of covariate-specific LATEs only in the case where the specification is saturated in the covariates. Furthermore, in the presence of heterogeneous nonlinear treatment effects, even nonparametrically controlling for covariates is not sufficient to identify a causal effect; the first stage must also be monotonicity-correct. The previous literature using the judge instrument design has partially addressed this second point by creating the instrument within groups of offenders. In the main specification, those groups are formed by the felony status and class of the offense. Here, I fully address the monotonicity-correctness by expanding the definition of the groups within which the instrument is formed. In addition, I address Blandhol et al.'s first point by fully saturating the model in the covariates.

Appendix Table A.8 presents results estimated with a "rich covariate specification" in which all covariates are interacted. The specification includes single, double, and triple combinations of race, ethnicity, sex, first time offender, whether there was a previous charge in the last year, sex offender, six offense groups, felony, age, court, year, season, day of week, and shift. The table also presents estimates from a specification in which the instrument is interacted with the covariates in the first stage in order to reproduce the direction of monotonicity conditional on the covariates. Interacting the instrument with covariates produces the same result as creating it within many groups of individual and case characteristics, Mueller-Smith's (2023) approach to relaxing monotonicity. The new monotonicity assumption then requires that if one judge is ranked as more likely to assign mental health treatment than another judge, that ranking will hold true for all offenders of a certain type defined by those groups of characteristics. For example, among all Black first-time offenders, one judge must uniformly assign treatment more frequently than another judge.

In practice, interactions between the covariates and instruments result in a large number of potential instruments and many potential controls. In order to avoid overfitting or many-instruments bias, I use Lasso with k-fold cross-validation to partial out those instruments with little explanatory power for predicting mental health treatment and those controls with little explanatory power for predicting recidivism. Following Chernozhukov et al. (2018), I implement a doubly-robust procedure that guards against model selection mistakes and requires only one model to be correctly specified. It relies on Neyman orthogonality combined with sample splitting so that the estimators of the nuisance parameters do not contribute to the limiting distribution of the parameter of interest. Of the 73 possible instruments, Lasso chooses 13. It exclusively chooses interaction terms, particularly interactions of race with other demographic characteristics and of offense groups with criminal history characteristics. The results in Table A.8 are estimated using the Lasso-selected instruments and control variables. Correctly saturating the model in covariates and instruments has little effect on the magnitude of the estimated effect of treatment on recidivism. While the specifications with saturated instruments have smaller F-statistics, the results are qualitatively similar to estimates using only the within-offense-and-felony-groups instrument, suggesting that incorporating differences by crime types is sufficient to address any monotonicity violation.

1.5.6.3 Judge Discretion in Conviction

Judges have discretion over whether to convict a defendant, which could introduce bias to the estimates by threatening the exogeneity of the instrument. The design of the empirical strategy requires that the judge's tendency toward mental health treatment must only impact recidivism through its relationship with the assignment of an offender to mental health treatment. If the judge's tendency toward mental health treatment is correlated with their tendency toward convicting, then ignoring that conviction tendency would introduce endogeneity into the instrument. In practice, the correlation between those two tendencies of the judge is very small, at around 0.07. Nevertheless, I explore two alternatives to the main analysis specification. Appendix Table A.9 presents those results.

The first group of two columns presents results from estimating Equation 1.2 on a sample restricted to only those probationers who were convicted. This approach eliminates concerns that the comparison group includes both individuals who were sentenced to probation without mental health treatment and those who were not convicted. However, the sample is selected on an outcome of the judge's decision set, which introduces concerns about selection bias. Results presented in the second group of two columns instead rely on the main analysis sample, but treat the judge conviction tendency like the SUD treatment tendency, conditioning on the judge's likelihood of convicting other offenders in the sample. In both alternative specifications, the resulting estimate of the effect of mental health treatment on recidivism is slightly larger than (but not distinguishable from) the baseline estimate, at about 14 percentage points. The conclusion from this analysis is that any bias introduced by the judge's tendency toward conviction is small and does not change the implications of the results.

1.5.6.4 Analysis Sample and Treatment Variable

Many choices are made when building the analysis sample from the raw court files. Here I discuss several potentially important decisions, and show that the results are robust to alternative choices.

I first address the variable identifying whether an offender is required to seek mental health treatment. The variable is formed by searching for key words in a long string of special conditions and notes that are attached to each case. Appendix Table A.10 addresses concerns that the results could be sensitive to which key words are included in the definition of mental health treatment. In order to isolate the differences from changing the mental health treatment variable definition only, this table presents estimates from the specification without controlling for the judge's SUD treatment propensity. The baseline estimate from that specification is a 14.1 percentage point decrease in recidivism due to mandated mental health treatment. Table A.10 considers various ways to restrict the key words that define mental health treatment, and shows that the estimated effect remains similar across those definitions. When the definition is more restrictive, the estimated effect is about the same, at 13.4 percentage points. When the definition is less restrictive, the estimated effect is slightly higher, at 15.2 percentage points. When some SUD treatment key words are excluded from the definition, the estimated effect is also larger, at 16.5 percentage points. All estimates have strong first stages, and are well within one standard deviation of the baseline estimate. The preferred specification takes a moderate restrictiveness toward what is considered mental health treatment, which corresponds to a fairly conservative estimate of the effect of treatment.

I now turn to exploring the definition of the SUD treatment variable. Appendix Table A.11 presents estimates from the main specification which controls for the judge's propensity toward SUD treatment, and which uses the baseline definition of mental health treatment. The table considers definitions of SUD treatment that vary the restrictiveness of the key words that are included. The baseline estimated effect here is a reduction in recidivism by 12.1 percentage points. Using two less restrictive definitions, the estimate becomes 9.6 and 9.3; using a more restrictive definition, the estimate is slightly higher, at 12.7. These estimates all have strong first stages and none of the effect sizes are statistically distinguishable from one another. The preferred specification again takes a moderate restrictiveness toward what is considered SUD treatment.

The main results explore the effect of mental health treatment conditional on the judge's propensity toward SUD treatment. Another way to conceptualize treatment is as the combination of both special conditions under the judge's control. Appendix Table A.12 presents estimates of the effect of being assigned mental health treatment or SUD treatment. This specification implies a different model of the judge's decision-making. Here the judge jointly chooses to assign some sort of treatment (could be counseling, drug rehab, or both) or to assign no treatment, and that decision is dependent on one underlying factor that is unobserved by the researcher. Instrumenting with the judge's propensity to assign any treatment isolates variation that moves some offenders into receiving a treatment. I estimate that being assigned any treatment decreases recidivism by 12.9 percentage points, very similar to the baseline estimate of the effect of mental health treatment. This exercise provides suggestive evidence that the estimate of the effect of mental health treatment is not sensitive to the specific model I impose on judge behavior.

An alternative view more in the style of Heckman and Pinto (2018) would be to characterize treatment as four mutually-exclusive and exhaustive categories. Abbreviating mental health treatment as MHT and substance use disorder treatment as SUDT, the categories are: (1) MHT alone, (2) SUDT alone, (3) MHT and SUDT, and (4) neither MHT nor SUDT. Instruments for these four treatments can be created using the judge's tendency toward each specific category. For example, the instrument for being assigned MHT alone would be the judge's tendency to assign MHT alone among all other offenders in the sample. Appendix Table A.13 presents results from this approach, where "neither" is the left-out category. Estimating each of the treatment categories separately suggests that MHT alone is similarly effective as MHT paired with SUDT, with estimated 16.2 and 15.8 percentage point declines in recidivism, respectively. SUDT alone is slightly less effective, associated with a decline of 11.1 percentage points. However, the results including all three treatments together show that the effect is concentrated in MHT alone. The specification with the three treatments suggests that MHT alone, compared to no treatment (the left-out category), reduces recidivism by 15.1 percentage points. The estimates for MHT and SUDT, and for SUDT alone, are not statistically distinguishable from zero. The comparisons from this exercise suggest that the effect of mental health treatment on future crime is operating primarily through the channel of mental health treatment rather than through an interaction with SUD treatment.

The last major choice related to sample definition is the characterization of the outcome variable. For the majority of the analysis, recidivism is defined by whether the offender appears in the data with a later trial date. That definition includes probation violations that lead to a future court appearance. However, it is not clear whether violating probation should count as recidivism. Here, recidivism is instead defined to exclude cases in which the future trial is a probation revocation case. Table 1.5 presents those results. It first shows suggestive evidence that being assigned mental health treatment slightly increases the likelihood of having a probation violation. Because seeking mental health treatment adds an additional requirement to the probation terms, it could mechanically cause more violations. If probationers are unable to pay for counseling, or do not have the time or access to transportation, they could have their probation revoked. It is unclear how often this happens in practice, as probation officers have discretion to give warnings rather than an official charge of a violation. I find that mental health treatment increases the likelihood of violating probation by 2.6 percentage points, but that estimate is not statistically significant. The next two columns of the table show that the estimated effect of mental health treatment on recidivism is slightly larger when the measure of recidivism does not include probation violations. I estimate a 12.9 percentage point decrease in non-violation recidivism. This exploration of the characterization of recidivism suggests that the preferred specification results in a conservative estimate of the effect of treatment. In a scenario in which the onus of payment were not on the probationer, mental health treatment could potentially be more effective because it might not induce mechanical violations related to financial barriers. However, it is possible that being responsible for payment makes individuals more inclined to participate in the treatment. In evaluating the effect of mental health treatment, the preferred specification counts violations in recidivism, since it characterizes the total effect of treatment under the current mandated environment.

1.6 Policy Implications

The preceding analysis suggests that mandating mental health treatment for a heterogeneous group of individuals on probation would decrease their criminal involvement over a fairly long horizon. In this section I contextualize those results by considering whether the benefits of treatment outweigh the costs of providing it. In North Carolina, probationers are generally required to pay for their own mental health treatment even if it is court-mandated. For the following cost-effectiveness exercises, I first carry out an accounting of the costs and benefits associated with mandated treatment. I then discuss assumptions under which the estimates from this paper can be applied to consider a policy that would instead pay for mental health treatment for the judge-chosen individuals.

I compare the cost of providing treatment with the benefits, which include lower judicial costs, lower social costs to victims, and lower costs to offenders from fewer future criminal justice system interaction. Importantly, I consider the benefits associated with reductions in crime separately by type of crime, since the studied population of probationers is more likely to commit less severe crimes. Ignoring the characteristics of the crimes would result in an overestimate of the benefits of paid provision of mental health treatment. Appendix A.1 describes the approach taken here in more detail and provides ranges for each estimate discussed below. All monetary amounts referenced are in 2022 dollars.

1.6.1 Costs of Providing Mental Health Treatment

To calculate the per-person costs of mental health treatment, I rely on Medicaid reimbursement tables. The main cost of the policy comes from the provision of counseling and psychiatric medication. I define treatment as one initial intake evaluation combined with weekly counseling sessions for the duration of probation. I include the average cost to Medicaid of daily antidepressants scaled by the likelihood of using those medications.[18] On average, providing treatment for one individual would cost the government $3,042. I make two adjustments to that initial estimate. First, I include the marginal excess burden due to deadweight loss if all those funds were raised through an increase in taxes. Second, I remove the costs associated with the approximately 36 percent of probationers who would have already been eligible for and would have taken up Medicaid. That fraction reflects the amount of probationers who both meet the low income threshold and are parents or disabled, since North Carolina did not expand Medicaid eligibility beyond those groups. Combining the two adjustments, my final estimate of the costs of treatment for one individual is $3,233 in 2022 dollars.

1.6.2 Benefits of Providing Mental Health Treatment

To calculate the per-person benefits from the reduction in future criminal propensity, I use separate estimates of the effect of mental health treatment on different classes of crimes. I estimate Equation 1.8 for future offenses of violent and property, financial and fraud, traffic and public order, drugs and alcohol, and miscellaneous types. The estimation procedure follows the main specification, but here the outcome is an indicator for the individual's future offense being of a certain type and the comparison is to not committing a future crime.[19] These five crime categories are the same groups within which the instrument is created, as described in Section 1.4.2.

[18]Psychiatric medication is used by about half of the individuals seen by counselors who work in the criminal justice system, based on conversations with several of those counselors.

[19]In other words, I estimate recidivism in the sample of offenders who either commit a future crime of that type or do not commit a future crime.

$$\texttt{Offense}_{ict+1} = \beta_0 + \beta_1\widehat{\texttt{MHT}}_{ict} + \beta_2\mathbf{X}_{ict} + \gamma_{ct} + \varepsilon_{ict} \tag{1.8}$$

Assigning a monetary value to the costs of the criminal justice system is a difficult task. On the one hand, information can be gathered on the financial outlays toward operating the courts, police, jails and prisons, probation, and other aspects of the judicial system. On the other hand, the cost of crime to victims and society as well as the detrimental effects of incarceration on offenders are much less straightforward to value. I pull from an important literature characterizing the costs of crime in order to form my cost estimates. I consider the judicial costs, sourced from Hunt et al. (2017) and McCollister et al. (2010); the loss of tax dollars due to decreased productivity of offenders while incarcerated, sourced from McCollister et al. (2010); the social costs, sourced from Heaton (2010), Hunt et al. (2017), and McCollister et al. (2010); and the costs to the offenders, sourced from McLaughlin et al. (2016). Social costs include both direct losses to victims as well as indirect losses due to feeling unsafe, etc. Whenever possible, I use cost information disaggregated by crime type and the associated estimate of the effect of mental health treatment on future crimes of that type. I then aggregate the cost estimates weighting by the relative prevalence of each offense group among repeat offenders. On average, the costs avoided by providing mental health treatment total $16,131.

1.6.3 Combining Costs and Benefits

I first consider a simple accounting exercise comparing the costs to the offender versus the total benefits through reductions in crime. I make an adjustment to the costs to the offender calculated above, since that estimate gives the cost to the government of providing treatment through Medicaid. Mental health treatment is much more costly out-of-pocket – about twice as costly as listed in Medicaid reimbursement tables. When thinking about offenders' willingness to pay for treatment themselves, the relevant figure would be \$5,411.29. Still, comparing the cost to the offender with the benefits (\$16,131) shows that mandated mental health treatment is much more financially beneficial than it is costly.

However, many of those benefits accrue to society overall rather than to the offenders themselves. Therefore, I consider a policy that would shift the burden of payment to the government. In order to rely on the same estimated benefits from the reductions in crime, I make two simplifying assumptions about behavioral responses. I assume that both the effect of mandated mental health treatment and the assignment of offenders to treatment by judges do not depend on who pays. Because the causal effect estimated in this paper is that of *being mandated to seek* mental health treatment, the financial burden is part of the treatment. I have shown suggestive evidence in Section 1.5 that for more financially disadvantaged individuals, being required to pay induces more probation violations, which suggests the effect of mandated mental health treatment would be larger if the financial burden were shifted off of those probationers. On the other hand, I have also shown evidence consistent with a financial deterrence effect on more financially advantaged individuals, which suggests the effect could be smaller if the financial burden were taken away. This policy exercise could be viewed as imposing that any effects of the financial aspect of treatment average out to zero in the population overall. Alternatively, a policymaker might prefer to view this exercise as applying only to those more financially disadvantaged probationers, in which case any behavioral response to the change in payment incidence would render these an underestimate of the cost-effectiveness of the policy.

Figure 1.11 presents a graphical depiction of the magnitudes of the costs of treatment to the government and benefits of treatment through the reduction in crime. Comparing the costs relative to the benefits, I find that paying for mental health treatment would about equalize with the benefits earned from the reduction in direct costs of prison alone. Incorporating societal costs of crime results in a benefit-to-cost ratio of about 5 to 1. The high ratio suggests that among probationers, mental health treatment for the marginal individual is highly beneficial for society.

Another way to evaluate the cost-effectiveness of the policy is through the marginal value of public funds (MVPF), which measures the amount of revenue that can be delivered to the relevant beneficiaries per dollar of government spending on the policy (Hendren and Sprung-Keyser, 2020). The MVPF is the ratio of the individual's willingness to pay to the net costs of the policy to the government. In the case of mental health treatment among probationers, my estimate of the net costs to the government ranges from -$790 to $2,721. The large range comes from two sources, uncertainty due to the estimated effect of treatment and uncertainty due to the disagreement in the literature on the magnitude of the costs of crime. If the savings due to reductions in crime are greater than the cost of the program, then the net cost to the government is negative and the marginal value of public funds is infinite (as long as the willingness to pay is positive). In the case of the lower bound cost estimate, the MVPF of mental health treatment among probationers is infinite. To form the MVPF associated with the upper bound estimated net costs to the government, I turn to estimating the individual's willingness to pay.

To form an estimate of the individual's willingness to pay, I consider both offenders and a representative member of society. The offender benefits by avoiding the negative effects of incarceration, including lost earnings, increased risk of sexual abuse and mortality, and broken family connections. Whether their case is dismissed, they are sent to prison, or they are put on probation, all offenders benefit from the reduced likelihood of shouldering court fees and fines – just the conviction-related court fines alone average $869 for misdemeanors and $1,628 for felonies (Rafael, 2023). The representative member of society benefits in many ways related to the decreased likelihood of being a victim of a crime or a family member of an incarcerated person. Tangible benefits include reduced likelihood of the loss of property and medical treatment. Intangible benefits, much harder to measure, include reductions in fear of crime and the psychological effects of being a victim of crime. I quantify these benefits using the estimates from Section 1.6.2 for the total social costs and costs to offenders weighted by the effect of treatment on future crime. The willingness to pay for reductions in crime associated from mental health treatment is estimated at $14,526 (with a range between $4,069 and $31,650). Finally, combining the willingness to pay with the net cost to the government, I estimate a MVPF of 19.3, with a range between 1.5 and infinity. That the lower bound estimate of the MVPF is still greater than one provides strong evidence of the cost-effectiveness of the policy.

The findings from the traditional cost-benefit analysis and the marginal value of public funds exercise confirm that spending one dollar on mental health treatment for probationers returns more than one dollar to the beneficiaries of the policy. For comparison, Jácome (2020) estimates an MVPF of 1.8 for a policy related to reducing crime through the provision of health insurance. Her policy involved expanding Medicaid eligibility for two years to all low-income young men in South Carolina. The policy I consider here has much lower costs than Jácome's, which likely contributes to the larger estimated MVPF. Drug courts are another related policy with much higher costs, at about $21,000 per person (Prins et al., 2015). Studies have found potential benefits of these courts, but those benefits are smaller than the estimated effect of mental health treatment on recidivism (see Deschenes et al. (1995); Gottfredson et al. (2003); Prins et al. (2015)). However, holding the distribution of future crimes in the sample constant, the drug courts would need to be about twice as effective as mental health treatment in order to return an

MVPF greater than 1.

1.6.4 Areas for Further Research

I view the cost-benefit analysis carried out here as an underestimate of the cost-effectiveness of mental health treatment among probationers for several reasons. These cost-benefit exercises focus only on the effect of mandated mental health treatment on criminal behavior. If receiving mental health treatment makes it easier for probationers to find and keep employment or achieve more education, then the benefits of mandated therapy would be even larger. In addition, the estimates in this section are based on benefits due to reductions in recidivism in the three years since trial, but the results in Section 1.5.3 suggest that recidivism risk in later years remains lower for probationers who are sentenced to mental health treatment. The true lifetime benefits of mental health treatment are therefore likely larger than calculated here.

This exercise is also limited in that it takes the assignment of treatment as given. Other potential policies could involve expanding assignment of treatment to more probationers, to other types of offenders, or even to individuals outside the criminal justice system. Because individuals on probation are by definition committing less serious offenses (in order to be assigned to probation by the Structured Sentencing laws), the estimated monetary value of the benefits of reduced crime are much smaller than they could be. For example, the social cost of violent crimes like homicide, rape, and assault is nearly \$1 million, but only about five percent of repeat offenses in the analysis sample are violent crimes. In a population in which more serious offenses were committed, the benefits of treatment could be even larger. However, the benefits also depend on how effective mental health treatment is in those settings, which this paper cannot answer. The results in this paper, finding that mental health treatment is similarly effective across many demographic and criminal history characteristics, suggest that treatment could be effective among individuals in different scenarios as well. They also suggest limited scope for targeting the provision of treatment to specific types of offenders, since I do not identify any individual characteristic that is relatively more predictive of the effectiveness of the treatment. However, this paper has shown that probationers are a group for which judge-targeted treatment proves a cost-effective policy to reduce crime.

1.7 Conclusion

This paper evaluates the causal impact of mental health treatment on individuals' likelihood of committing a future crime. Using randomized judge assignment to North Carolina criminal cases, I find that mandated mental health treatment as a term of probation decreases future crime in both the short and longer term. I estimate that being assigned to seek mental health treatment decreases the likelihood of three-year recidivism by about 12 percentage points, or 36 percent. That effect persists over time: by five years after conviction, offenders are about 11 percentage points less likely to recidivate. My estimation strategy addresses the interaction of mental health treatment with another aspect of probation, substance abuse disorder treatment. I show that, with some assumptions about the model of judge decision-making, I can recover a margin-specific treatment effect. I also explore different dimensions of heterogeneity, finding that treatment is similarly effective in many subgroups. I show that the effect of mental health treatment is not concentrated among a certain class of offenses, though the evidence in this paper suggests that mental health treatment is particularly effective at reducing more serious offenses. I validate these results in a series of robustness checks, all which suggest that mandated mental health treatment contributes to lower criminal propensity.

Taken together, the results suggest that mandating mental health treatment as a term of probation can improve the transition of convicted criminals into civilian life. I quantify the costs and benefits of a policy paying for mental health treatment for all judge-chosen probationers, using separate estimates of the costs and the effects of treatment among offense groups. I estimate that the benefits from the reduction in crime would more than recover the costs spent to provide mental health treatment, at a rate of about 5 to 1. The associated marginal value of public funds is 19.3. Among probationers, mental health treatment for the marginal individual is highly cost-effective for society.

Almost half of jail inmates in 2012 had a history of mental health problems (Bronson and Berzofsky, 2017). Individuals with mental illness can struggle to make rational, welfare-maximizing decisions, making traditional incentive-based strategies for crime deterrence less effective. Mental health treatment helps align their preferences with their actions; it can teach strategies to avoid high-risk situations, improve decision-making, support long-term behavioral change, and correct the perceived costs and benefits of behaviors. The findings of this study illustrate the potential of mandated mental health treatment as a beneficial and cost-effective alternative to those traditional crime deterrence strategies.

1.8 Tables

Table 1.1: Offender Characteristics by Mental Health Treatment Sentence

	All	MHT	No MHT
Age	30.53	32.58	30.43
	(10.15)	(10.66)	(10.12)
Female	0.20	0.21	0.20
	(0.40)	(0.41)	(0.40)
Black	0.45	0.35	0.46
	(0.50)	(0.48)	(0.50)
Hispanic	0.02	0.03	0.02
	(0.15)	(0.16)	(0.15)
Sex Offender	0.01	0.03	0.01
	(0.10)	(0.17)	(0.10)
First Time Offender	0.61	0.70	0.60
	(0.49)	(0.46)	(0.49)
Private Attorney	0.31	0.48	0.31
	(0.46)	(0.50)	(0.46)
Assigned MH Treatment	0.05	1.00	0.00
	(0.21)	(0.00)	(0.00)
Assigned SUD Treatment	0.04	0.37	0.02
	(0.20)	(0.48)	(0.15)
Recidivism within 3 Years	0.34	0.23	0.35
	(0.47)	(0.42)	(0.48)
Failed Probation	0.06	0.08	0.05
	(0.23)	(0.27)	(0.23)
Observations	727,184	33,886	693,298

Note: This table presents summary statistics for the main estimation sample. The sample is divided into two categories: those who were assigned mental health treatment and those who were not. Observations are required to have non-missing control and treatment variables and to have been seen at least 3 years before the end of the data so that recidivism can be measured. First-Time Offender is based on first appearance within the sample period beginning in 1994, which could misclassify offenders who were arrested in years prior. "SUD" stands for Substance Use Disorder. Source: 1994-2009 North Carolina ACIS data.

Table 1.2: Balance of Judge Assignment across Controls

	Actual MHT	Judge Propensity
Age	-0.001***	0.000
	(0.000)	(0.000)
Female	0.008***	0.000
	(0.001)	(0.000)
Midwest	0.031***	0.001
	(0.009)	(0.002)
South	0.007	-0.001
	(0.005)	(0.004)
West	-0.004	0.001
	(0.012)	(0.002)
Black	-0.020***	-0.000
	(0.001)	(0.000)
Hispanic	-0.008***	-0.000
	(0.002)	(0.000)
Domestic Violence or Sex Offense	0.051***	0.003***
	(0.003)	(0.001)
First Time in System	0.009***	0.000
	(0.001)	(0.000)
Prior Arrest in the Last Year	-0.002***	-0.000
	(0.001)	(0.000)
Private Attorney	0.017***	0.001***
	(0.001)	(0.000)
Joint F-Stat	404.9	2.0
P-value	0.000	0.001
Observations	727,184	727,184

Note: This table presents results from regressions using either the judge propensity to assign mental health treatment or the mandate to seek mental health treatment as the dependent variable. I regress these on the same vector of controls used in the main specification (age cubic, sex, region of United States, race/ethnicity, first time offender, sex offender, offense groups, prior appearances, attorney) and the court-time fixed effects. I report the F-statistic and associated p-value from a test of joint significance of the controls. The F-statistic is much larger for predicting mental health treatment than for predicting judge propensity toward mental health treatment. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.3: First Stage

	Mandated Mental Health Treatment			
MHT Propensity	0.024***	0.024***	0.022***	
	(0.001)	(0.001)	(0.001)	
SUDT Propensity		0.001**	0.001**	
		(0.001)	(0.001)	
Age at Date of Judgement Issuance			-0.000	-0.000
			(0.001)	(0.001)
Female			0.006***	0.006***
			(0.002)	(0.002)
Midwest			0.050	0.048
			(0.031)	(0.031)
South			-0.023	-0.023
			(0.018)	(0.018)
West			-0.057	-0.061*
			(0.036)	(0.036)
Hispanic			0.008	0.010
			(0.007)	(0.008)
Black			-0.041***	-0.041***
			(0.002)	(0.002)
Domestic Violence or Sex Offense			0.054***	0.062***
			(0.007)	(0.007)
First Time Offender			0.023***	0.024***
			(0.002)	(0.002)
Prior Arrest in Last Year			-0.001	-0.000
			(0.002)	(0.002)
Public Attorney			-0.001	-0.002
			(0.002)	(0.003)
Private Attorney			0.033***	0.033***
			(0.003)	(0.003)
Observations	727,184	727,184	727,184	727,184
Adj. R^2	0.063	0.063	0.082	0.071

Note: This table presents results for the first-stage of the main specification. These results come from regressing the mandate to seek mental health treatment on the standardized judge propensity instruments for mental health treatment (MHT) and substance use disorder treatment (SUDT) and individual characteristics. The controls included are race, ethnicity, sex, age, region, first-time offender, attorney, offense types, whether the offender had a prior offense in the past year, and county-level descriptors: per-capita psychologists, emergency room visits related to mental illness, Medicaid coverage, uninsurance rate, poverty rate, violent crime rate, and proportion with at least a bachelor's degree. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.4: Effect of Mental Health Treatment on Recidivism

	OLS		IV	
Assigned MHT	–0.102***	–0.047***	–0.162***	–0.121***
	(0.003)	(0.003)	(0.045)	(0.044)
1st Stage F-Stat			1024	970
Mean of Outcome	0.342	0.342	0.342	0.342
Court-time FEs	X	X	X	X
Demographic		X		X
Criminal History		X		X
County		X		X
Observations	727,184	727,184	727,184	727,184
Adj. R^2	0.032	0.095	0.022	0.066

Note: This table presents results from OLS and IV regressions of three-year recidivism on being mandated to seek mental health treatment. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The third column has no controls other than the court-time fixed effects and offense types; the fourth includes controls for race, ethnicity, sex, age, and region; the fifth adds controls for first-time offender, attorney, offense types, and whether the offender had a prior offense in the past year; and the sixth column adds county-level descriptors: per-capita psychologists, emergency room visits related to mental illness, Medicaid coverage, uninsurance rate, poverty rate, violent crime rate, and proportion with at least a bachelor's degree. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.5: Effect of Mental Health Treatment on Recidivism, Excluding Probation Violations

	Fail Probation		Recidivate, Not Counting Fail	
	OLS	IV	OLS	IV
Assigned MHT	0.008***	0.026	–0.050***	–0.129***
	(0.003)	(0.036)	(0.003)	(0.045)
1st Stage F-Stat		970		970
Mean of Outcome	0.056	0.056	0.325	0.325
Observations	727,184	727,184	727,184	727,184
Adj. R^2	0.091	0.062	0.097	0.068

Note: This table presents results from OLS and IV regressions of either failing probation or recidivism (defined so that probation violations do not count as committing a future crime) on being mandated to seek mental health treatment. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.6: Effect of Mental Health Treatment within Charged Felony Type

	Misdemeanor		Felony	
	OLS	IV	OLS	IV
Assigned MHT	–0.071***	–0.101**	–0.050***	–0.177**
	(0.003)	(0.048)	(0.005)	(0.071)
1st Stage F-Stat		769		589
Mean of Outcome	0.31	0.31	0.37	0.37
Prop. Assigned MHT	0.07	0.07	0.03	0.03
Observations	305,420	305,386	421,764	421,425
Adj. R^2	0.080	0.032	0.067	0.041

Note: This table presents results from OLS and IV regressions of three-year recidivism on mental health treatment, estimated separately by misdemeanor and felony offenses. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.7: Effect of Mental Health Treatment Among Offense Groups

	Violent/Property	Financial/Fraud	Drugs/Alcohol	Traffic/Public Order	Miscellaneous
Assigned MHT	−0.846**	−0.972*	−0.175**	0.025	0.069
	(0.353)	(0.691)	(0.084)	(0.098)	(0.174)
1st Stage F-Stat	116	63	285	252	70
Mean of Outcome	0.42	0.42	0.29	0.29	0.27
Prop. Assigned MHT	0.03	0.02	0.05	0.07	0.03
Observations	195,659	115,313	129,436	264,023	21,154
Adj. R^2	−0.035	−0.132	0.022	0.009	0.025

Note: This table presents results from IV regressions of three-year recidivism on mental health treatment, estimated by five offense groups which are mutually exclusive and exhaustive. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.8: Effect of Mental Health Treatment on Number of Crimes

	OLS		IV	
Assigned MHT	–0.191***	–0.084***	–0.351***	–0.214**
	(0.006)	(0.005)	(0.093)	(0.083)
1st Stage F-Stat			1149	1126
Mean of Outcome	0.539	0.539	0.539	0.539
Controls		X		X
Observations	727,184	727,184	727,184	727,184
Adj. R^2	0.032	0.098	0.001	0.069

Note: This table presents results from OLS and IV regressions of the number of future crimes on mental health treatment. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.9: Effect of Mental Health Treatment on Future Sentence

	Unconditional Sentence Length		Conditional Sentence Length		Unconditional Active Sentence		Conditional Active Sentence	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Assigned MHT	−17.971***	−55.366*	−30.155***	−110.984	−0.026***	−0.062**	−0.053***	−0.091
	(1.829)	(28.528)	(7.004)	(105.882)	(0.002)	(0.031)	(0.006)	(0.094)
1st Stage F-Stat		970		352		970		352
Mean of Outcome	96.693	96.693	282.352	282.352	0.142	0.142	0.416	0.416
Observations	727,184	727,184	249,029	248,797	727,184	727,184	249,029	248,797
Adj. R^2	0.032	0.017	0.060	0.023	0.071	0.045	0.109	0.046

Note: This table presents results from OLS and IV regressions of the qualities of the future sentence (average sentence length measured in days and whether the sentence is an active prison sentence) on mental health treatment. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.10: Effect of Mental Health Treatment on Type of Offense Committed

	Violent/Property		Financial/Fraud		Drugs/Alcohol		Traffic/Public Order		Miscellaneous	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Assigned MHT	–0.012***	–0.111	–0.017***	–0.175**	–0.012***	–0.040	–0.006***	–0.056	–0.000	0.000
	(0.002)	(0.074)	(0.002)	(0.083)	(0.001)	(0.062)	(0.002)	(0.082)	(0.000)	(0.014)
1st Stage F-Stat		348		348		348		348		348
Mean of Outcome	0.10	0.10	0.09	0.09	0.07	0.07	0.08	0.08	0.00	0.00
Observations	727,184	727,184	727,184	727,184	727,184	727,184	727,184	727,184	727,184	727,184
Adj. R^2	0.080	0.052	0.076	0.037	0.034	0.017	0.042	0.012	0.015	0.003

Note: This table presents results from OLS and IV regressions of the type of future crime on mental health treatment. These results are *not* conditional on having committed a future crime. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.11: Effect of Mental Health Treatment on Severity of Offense Committed

	Misdemeanor		Felony	
	OLS	IV	OLS	IV
Assigned MHT	–0.008***	0.003	–0.039***	–0.120***
	(0.002)	(0.024)	(0.003)	(0.043)
1st Stage F-Stat		970		970
Mean of Outcome	0.07	0.07	0.27	0.27
Observations	727,184	727,184	727,184	727,184
Adj. R^2	0.077	0.019	0.112	0.067

Note: This table presents results from OLS and IV regressions of whether the future crime was a misdemeanor or a felony on mental health treatment. These results are *not* conditional on having committed a future crime. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

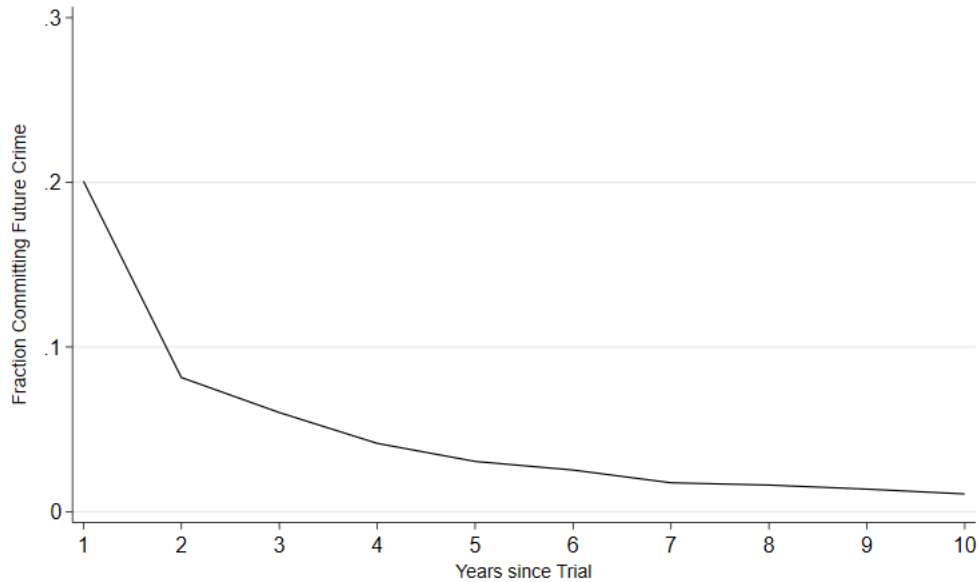
Table 1.12: Effect of Mental Health Treatment on Recidivism, Without Controlling for Drug Propensity

	OLS		IV	
Assigned MHT	–0.102***	–0.047***	–0.175***	–0.141***
	(0.003)	(0.003)	(0.042)	(0.041)
1st Stage F-Stat			1169	1126
Mean of Outcome	0.342	0.342	0.342	0.342
Controls		X		X
Observations	727,184	727,184	727,184	727,184
Adj. R^2	0.032	0.095	0.021	0.065

Note: This table presents results from OLS and IV regressions of three-year recidivism on mental health treatment. The specification is identical to the main specification except that the judge propensity to assign substance use disorder treatment is not included as a control. All specifications condition on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

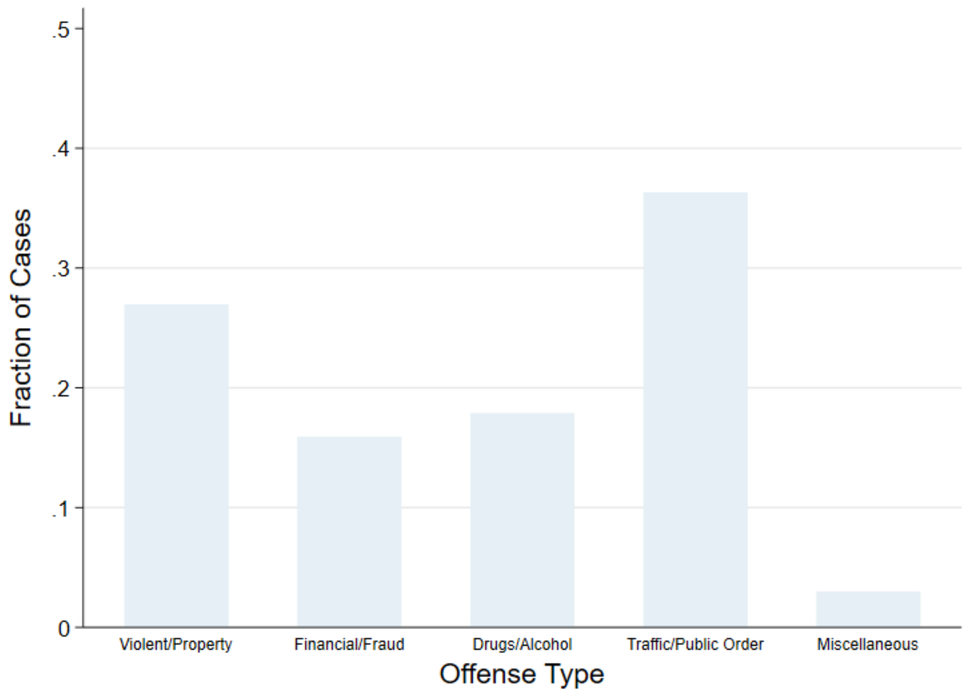
1.9 Figures

Figure 1.1: Rate of Recidivism over Time Since Trial



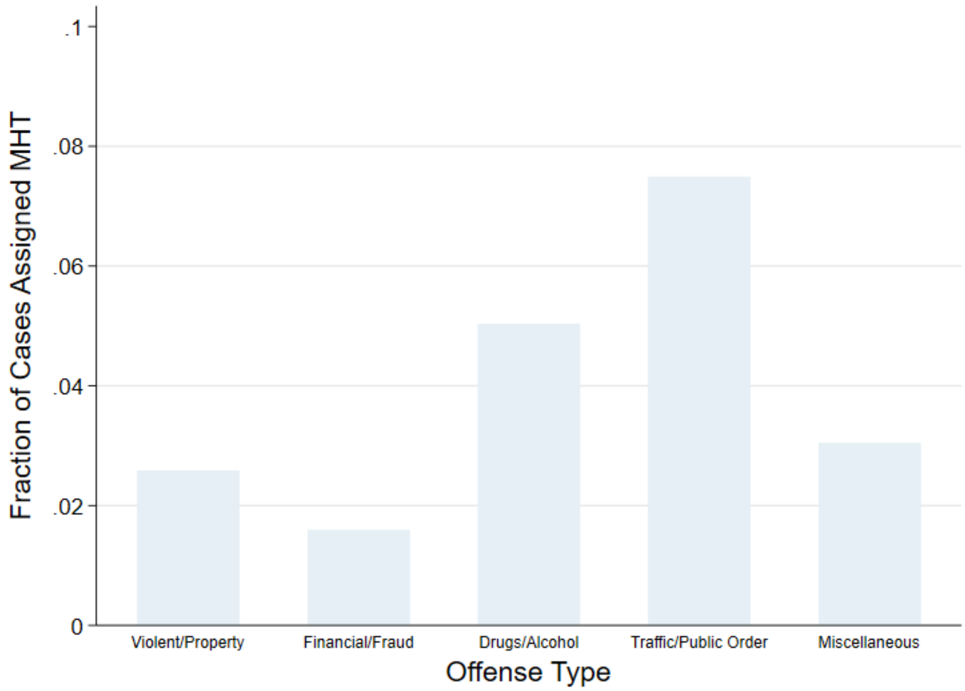
Note: This figure shows how many probationers are convicted of a new crime in each year since the trial. The highest risk of recidivism by far is in the first year following a conviction. Source: 1994-2009 North Carolina ACIS data.

Figure 1.2: Distribution of Offense Classes Among Probationers



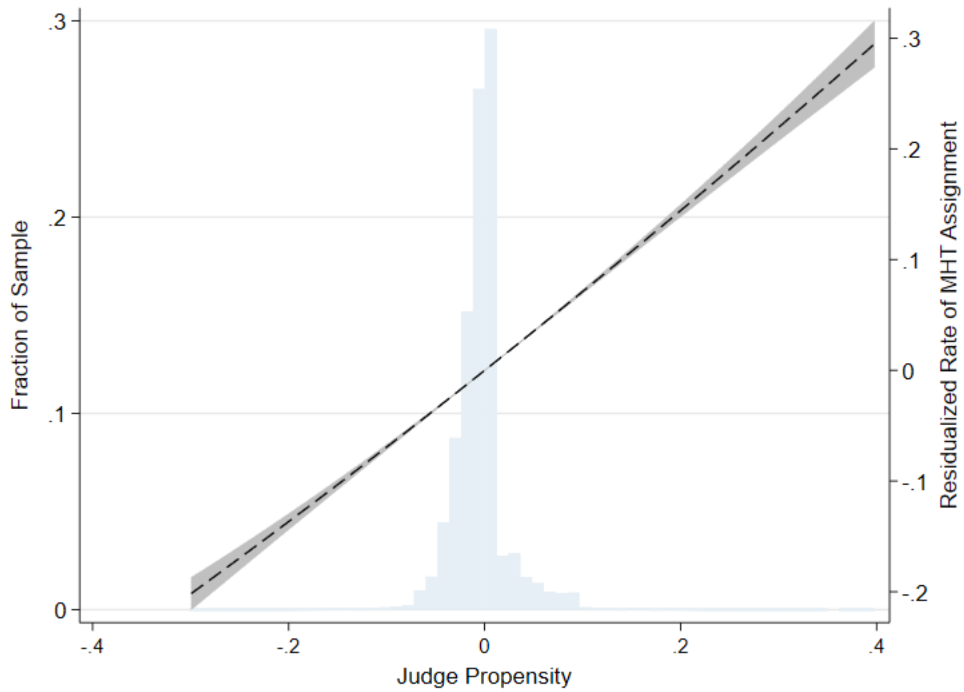
Note: This figure shows how often each offense type was committed by probationers. For each case, the defining offense is chosen to be the most severe offense from among those charged. Source: 1994-2009 North Carolina ACIS data.

Figure 1.3: Frequency of Mental Health Treatment Among Offense Classes



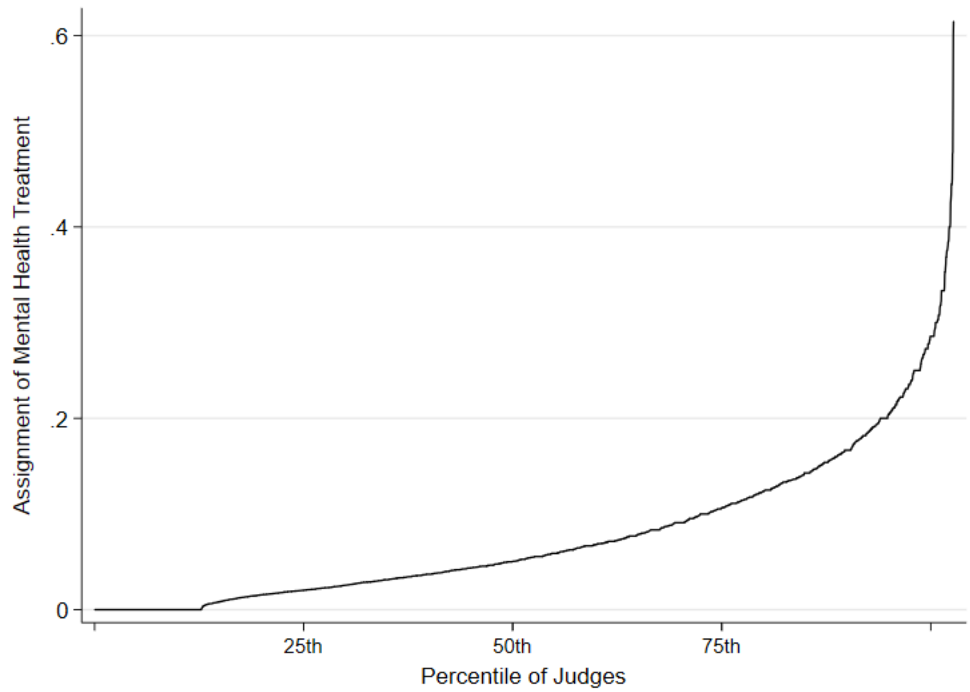
Note: This figure shows how often mental health treatment was assigned as a term of probation among different offense types. For each case, the defining offense is chosen to be the most severe offense from among those charged. Source: 1994-2009 North Carolina ACIS data.

Figure 1.4: Judges with a Higher Propensity Measure Assign More Mental Health Treatment



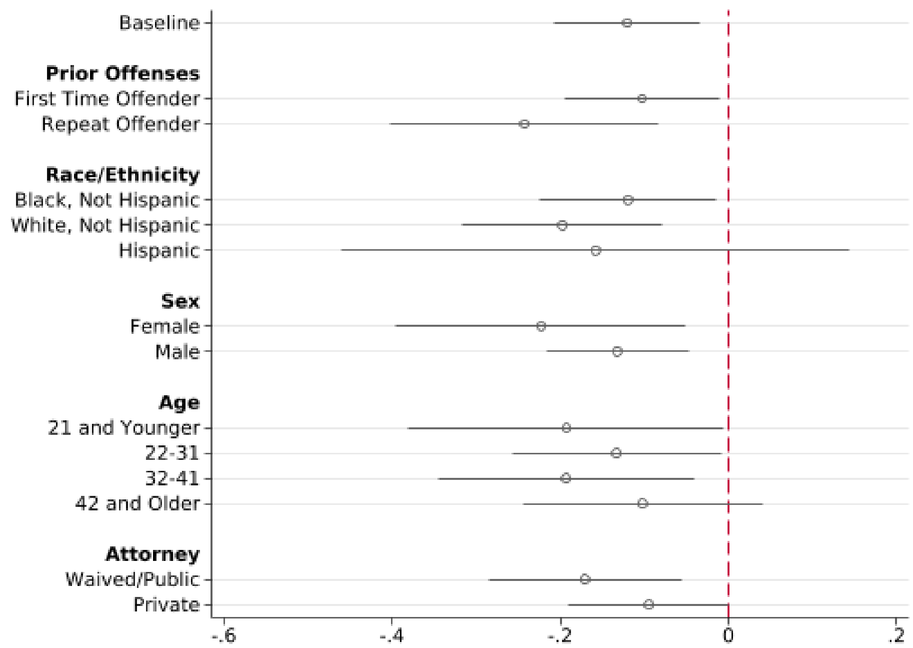
Note: On the left axis, this figure plots a histogram of the instrument, which is a measure of the judge tendency toward mandating mental health treatment. For the purposes of this graph I remove right-tail outlier judges (.7% of the sample). Positive values represent those judges who have a higher tendency toward mental health treatment. On the right axis, this figure plots mental health treatment, residualized on controls, against the measure of judge propensity. This provides a graphical representation of the first stage: judges with a higher propensity are more likely to mandate mental health treatment. Source: 1994-2009 North Carolina ACIS data.

Figure 1.5: Proportion of Time Judges Assign Mental Health Treatment



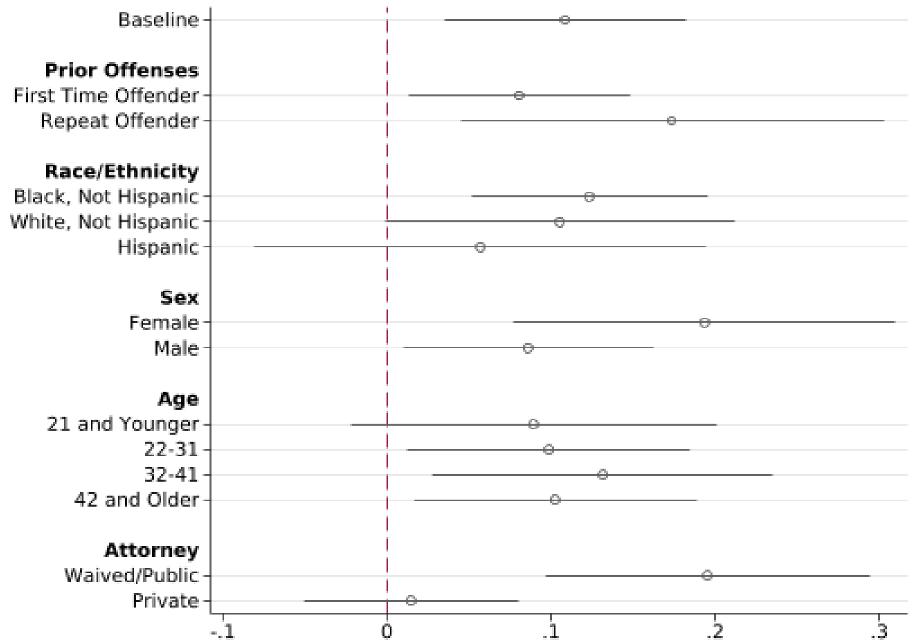
Note: This figure plots each judge's rate of mental health treatment assignment, ordered from those judges who are least likely to those who are most likely to assign mental health treatment. Source: 1994-2009 North Carolina ACIS data.

Figure 1.6: Effect of Mental Health Treatment on Recidivism, by Individual Characteristics



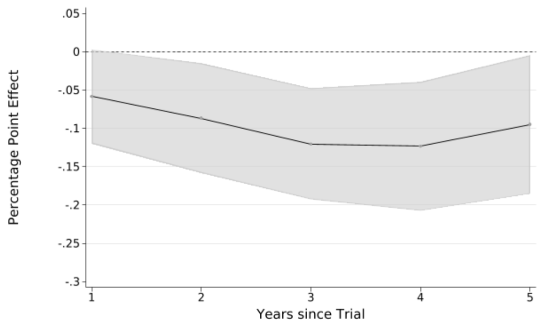
Note: This figure plots the percentage point decline in three-year recidivism due to mandated mental health treatment, estimated from the main IV regression with full controls. I estimate the effect separately for groups defined by five characteristics: first-time offender, race/ethnicity, sex, age, and attorney type. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data.

Figure 1.7: Effect of Mental Health Treatment on Failing Probation, by Individual Characteristics

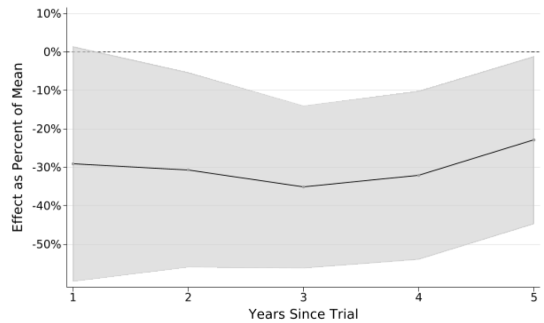


Note: This graph plots the percentage point effect on violating probation due to mental health treatment, estimated from the main IV regression with full controls. I estimate the effect separately for groups defined by five characteristics: first-time offender, race/ethnicity, sex, age, and attorney type. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data.

Figure 1.8: The Effect of Mental Health Treatment on Recidivism over Time



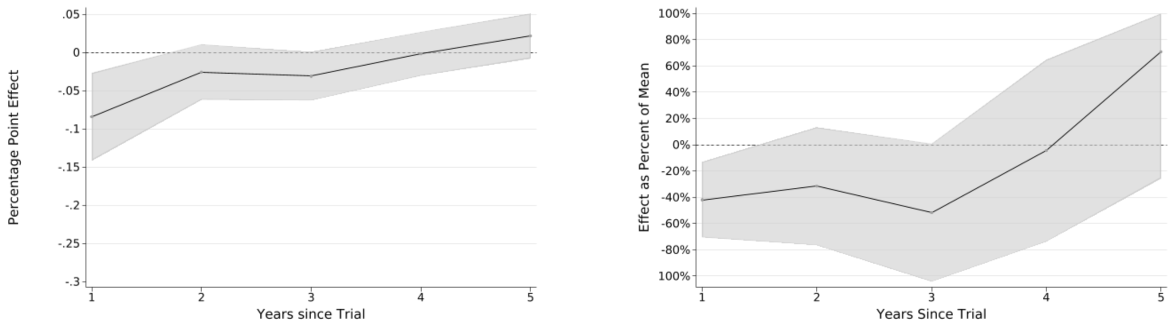
(a) Percentage Point



(b) Percent of Mean

Note: This figure plots the percentage point (left) and the percent (right) decline in recidivism due to mental health treatment, estimated from the main IV regression with full controls. Recidivism is defined first looking only one year out from the trial, then two years, and so on up to five years. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data.

Figure 1.9: The Effect of Mental Health Treatment on Recidivism in Each Year Since Trial

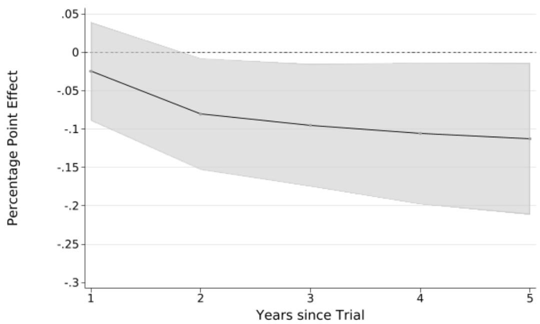


(a) Percentage Point

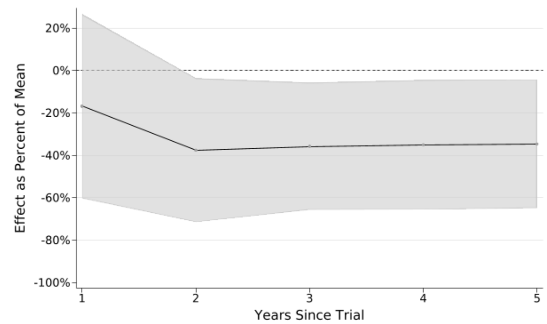
(b) Percent of Mean

Note: This figure plots the percentage point (left) and the percent (right) decline in recidivism due to mental health treatment, estimated from the main IV regression with full controls. Recidivism is defined first looking only within one year out from the trial, then between one and two years, then between two and three years, and so on up to five years. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data.

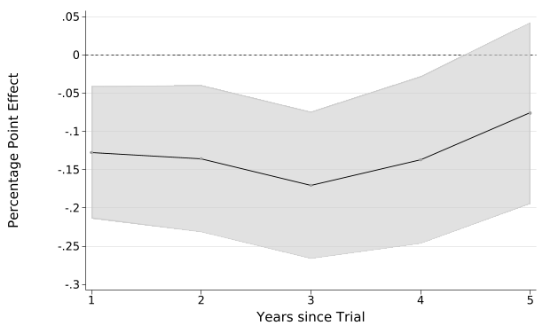
Figure 1.10: The Effect of Mental Health Treatment on Recidivism over Time, by Type of Attorney



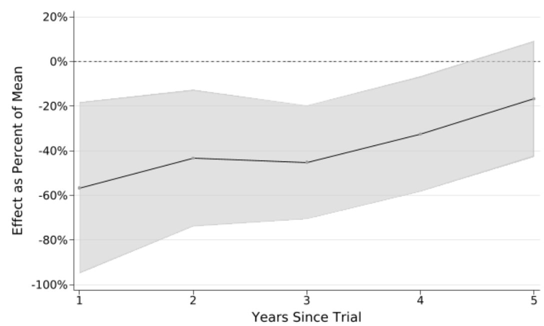
(a) Private Attorney, Percentage Point



(b) Private Attorney, Percent of Mean



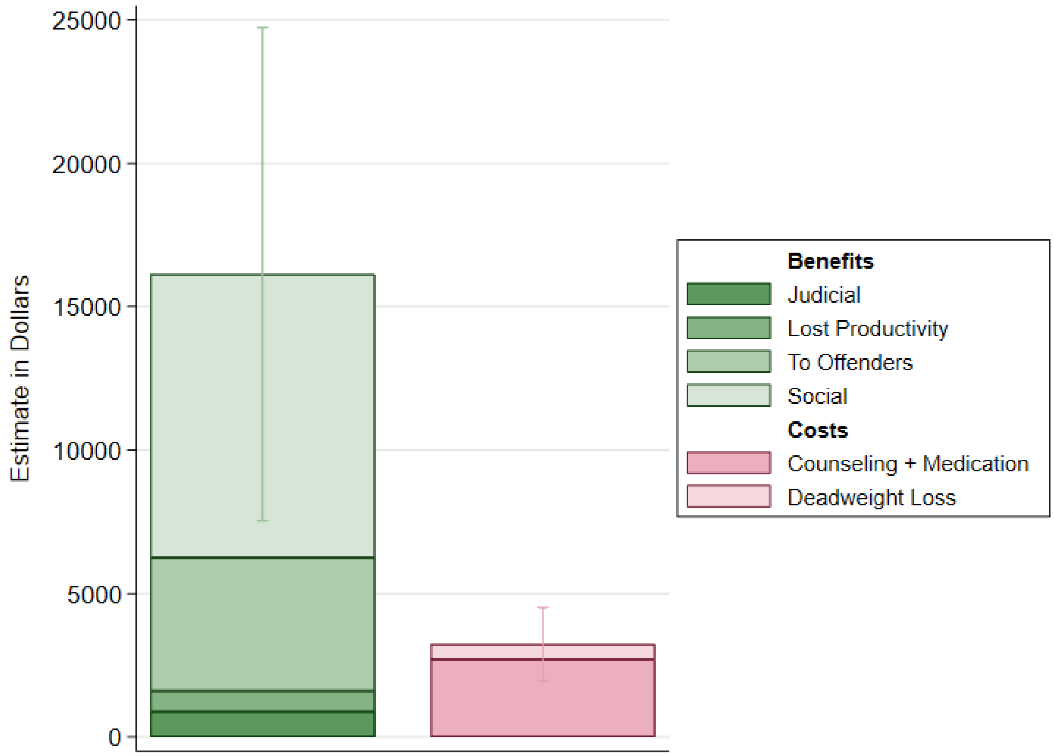
(c) Public Attorney, Percentage Point



(d) Public Attorney, Percent of Mean

Note: This figure plots the percentage point (left) and percent (right) decline in recidivism due to mental health treatment, estimated from the main IV regression with full controls. The effect is estimated separately among individuals with private attorneys (top row) and those with public attorneys (bottom row). Recidivism is defined first looking only one year out from the trial, then two years, and so on up to five years. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data.

Figure 1.11: Costs and Benefits of Mental Health Treatment During Probation



Note: This figure plots the costs to the government of providing mental health treatment (in red) alongside the monetary value of the benefits of mental health treatment through the decreased incidence of crime (in green). The large confidence intervals incorporate both sampling variation from the empirical estimates as well variation from the cost-of-crime literature as to the monetary valuation of different types of crimes. Comparing the size of the bars suggests that mandated mental health treatment is about 5 times as beneficial as it is costly.

Chapter 2: Same-Sex Couples and Parental Earnings Dynamics

2.1 Introduction

When a woman living in the United States has a child, her earnings fall by on average 21-61% depending on her state of residence (Kleven, 2022). This phenomenon, often referred to as the "child penalty," has been observed in various countries and is significant and persistent even in places where safety nets and job protections are robust (Kleven et al., 2021). Although the child penalty results in a large and sustained increase in the earnings gap within opposite-sex couples (Angelov et al., 2016; Chung et al., 2017), less is known about the child penalty for same-sex couples.[1] By comparing the earnings dynamics of opposite-sex and same-sex couples, we aim to understand the extent to which the child penalty is driven by earnings and/or work-life preferences versus by societal gender norms or equality norms that impact the labor and home production decisions when childcare responsibilities are introduced.

Based on our naïve a priori assumptions, we expect that the earnings and work-life preferences of same-sex couples will be comparable to those of opposite-sex couples. However, we expect that the dynamics of gender norms within same-sex couples will differ from those within opposite-sex couples. This difference allows us to focus on the household bargaining framework

[1] Throughout the paper, we adopt the naming convention from the American Community Survey (ACS) of "same-sex" and "opposite-sex" couples. Thus our convention defines the experience, for example, of a gay male with a female partner as being part of an opposite-sex couple. We acknowledge that this convention cannot capture the full range of complex identities and behaviors.

and explore how pre-parenthood earnings, as a measure of expected post-child earnings, are related to the labor/home production tradeoffs of the partners. Specifically, we explore whether one member of a same-sex couple assumes a primary caregiver role, resulting in an increase in the earnings gap after the child enters the household, or whether the additional demand on their time is more equally shared than in opposite-sex couples.

In addition to examining how individuals' earnings change, we also explore the effect of a child on the combined earnings of the couple. We aim to determine if household earnings trends change with the arrival of a child. Although most studies on the child penalty do not explicitly measure this, if male partners experience no change in their earnings after having a child and female partners experience a significant decrease, we would expect to see a decrease in household earnings in opposite-sex households. Our study explores whether this potential decrease in household earnings also exists for same-sex couples. Furthermore, the path of household earnings when a child enters has not been documented for same-sex couples in the U.S. and may depend on how the couple distributes home and market production within the household bargaining framework.

We develop a simple theoretical model to explore the factors that influence the division of time between home production and paid work when a child enters a household. This simple model abstracts away from the differences in preferences for children and constraints across opposite-sex, male same-sex, and female same-sex couples. We consider four possible decisions regarding the allocation of additional child-related home production work: allocation based on equality, allocation based on comparative advantage of earnings, allocation based on work-life preferences, and allocation based on societal gender norms.

We use an event-study design to estimate the earnings trends around the entrance of the first child. Following the methodology of Kleven et al. (2019), Andresen and Nix (2022), and others, we compare earnings in quarters after a child enters the household to earnings six quarters before the first child. In our main specification, we use inverse propensity weights to ensure that same-sex couples and opposite-sex couples have similar background characteristics, enabling us to compare couples whose income trajectories should be the same if not for the introduction of a child.

Our empirical analysis uses data from the American Community Survey (ACS) to obtain measures of couples (both spouses and unmarried partners), their identifying characteristics, and their children. We supplement the ACS data with date of birth data from the Census Numident and Internal Revenue Service (IRS) 1040 tax returns. This supplementary data allows us to be more precise about the date of birth of the first child or the date when an adopted child first enters the household. Finally, we define our outcomes of interest using individual quarterly earnings from the Longitudinal Employer-Household Dynamics (LEHD) dataset covering years 2005-2021. These data include quarterly earnings records collected by state unemployment insurance (UI) programs and provide coverage for over 95% of private-sector workers, as well as almost all state and local government employment (Abowd et al., 2009).

To preview our results, we find that the child penalty in opposite-sex couples applies almost exclusively to the female partner even if she was the primary earner before parenthood. In contrast, neither partner in same-sex male or same-sex female couples experiences as strong or persistent a drop in earnings as females in opposite-sex partnerships. If anything, our results are suggestive of a trend of equalization of earnings post-child, with the primary earners experiencing no change in earnings and the secondary earners increasing earnings. However, we also see that among same-sex equal earners, there is a shift toward specialization after the child enters. At the household level, we find that both opposite-sex and same-sex households experience a reduction in earnings in the year following the child's arrival, regardless of how the earnings losses are distributed across partners.

Disaggregating the results into intensive and extensive margins of earnings changes, we see that the sustained losses of women in opposite-sex couples primarily come from movement out of the labor force. For those that stay in the labor force, earnings recover within a year of the child's arrival. For men in opposite-sex couples, the upwards earnings trend post-child is even greater for primary and equal earner men when only considering the intensive margin. Secondary earner men are more likely to be employed post-child than pre-child, so the growth at the extensive margin is larger for them than the growth at the intensive margin. For both male and female same-sex couples, we see growth in earnings at the intensive margin for primary earners, while the trend in total earnings remains close to zero. This is also true of equal earners in same-sex couples. For secondary earners in same-sex couples, the change in earnings at the intensive margin is very close to the overall change in earnings, suggesting that there is limited extensive margin adjustment amongst secondary earners in same-sex couples after a child enters the household.

The paper is organized as follows: Section 2.2 provides a literature review and describes our contributions to that literature; Section 2.3 outlines our theoretical framework; Section 2.4 describes our data and methodology; Section 2.5 presents our results; and Section 2.6 provides discussion of the implications of our results and possible future research directions.

2.2 Literature Review

We contribute to two strands of literature by examining the earnings impact of the first child for same-sex (both female and male) versus opposite-sex couples in the United States. First, we provide a novel view of parental earnings penalties in the U.S., estimating effects for both members of the couple. Second, we contribute to the growing literature describing the characteristics of same-sex couples. Our paper builds on previous work combining these two threads that uses Scandinavian data; we analyze the earnings dynamics in the U.S. context, where the social safety net for new parents is more porous both with respect to parental leave and childcare. The larger U.S. population also allows us to separately explore the patterns in earnings of same-sex male couples with children which was not possible in the Scandinavian countries studied.

Much of this work builds on insights from Becker's 1965 classic household comparative advantage model. According to this model, we would expect the partner with higher earnings to specialize in paid market work and the partner with lower earnings to assume greater responsibility for home production work. It becomes challenging to disentangle the effects of gender norms from those of comparative advantage when higher earnings tend to accrue to one sex.

Previous studies have shown that the earnings gap widens among individuals in opposite-sex couples when they have children (Angelov et al., 2016; Chung et al., 2017). Kleven et al. (2021) find that the child penalty exists even when couples adopt. Other studies focus on mothers and show that their earnings decline, presumably leading to an increase in the earnings gap within couples (Byker, 2016; Hotchkiss et al., 2017; Kleven et al., 2019; Kuziemko et al., 2018; Neumeier et al., 2018; Sandler et al., 2019). Kleven et al. (2019) show that opposite-sex couples in Scandinavian countries experience smaller child penalties than in the United States.

Our analysis comparing patterns between opposite-sex couples and same-sex couples highlights the role of gender norms, which Kleven (2022) shows has an important role in the geographical distribution of the child penalty across the U.S. Using both spatial and temporal variation, he finds that gender norms are an important determinant of the presence and magnitude of a child penalty. For example, in Utah the earnings penalty is 61% while in Vermont it is only 21%.

Most of the evidence regarding the parental earnings gap for same-sex couples comes from Scandinavian countries. However, these studies have been unable to investigate male same-sex couples with children due to their small numbers. Therefore, this paper makes another important contribution by investigating parental earnings gaps among male same-sex couples. Andresen and Nix (2022) explore the differences between three groups of couples (opposite-sex adopting; opposite-sex not adopting; and female same-sex) to better understand the mechanisms underlying the child penalty. Using Norwegian data, they find that females in opposite-sex couples experience a substantial child penalty upon the birth of the child, but males in opposite-sex couples do not. Both parents in female same-sex couples experience a child penalty, but it is larger for the birth mother. They conclude that the long-term child penalty is not driven by giving birth or male labor market advantage, since there is recovery in even the birth-mother's earnings in same-sex couples. Instead, they suggest that preferences, gender norms, and discrimination play a significant role.

A number of studies have identified characteristics of same-sex couples that lead us to anticipate differences in the child penalty between opposite-sex and same-sex couples, as well as to anticipate differences within same-sex couple by gender. Same-sex couples are more likely to assortatively match on some dimensions, leaving less opportunity for specialization within couple. For example, Ciscato et al. (2020) examine homogamy in couples over race, age, and education using 2008-2012 ACS data for California. They find that same-sex couples, especially female same-sex couples, are more likely to have similar educational attainment. Black et al. (2007) likewise note that, based on Becker's 1991 theory of specialization, "we expect to observe less household specialization in gay and lesbian couples" (p. 61). On the other hand, same-sex couples are less similar on other dimensions; Ciscato et al. (2020) document that they are more likely to be different races and ages than opposite-sex couples.

Relatedly, same-sex couples are more likely to have smaller earnings gaps compared to opposite-sex couples. Antecol and Steinberger (2013) and Jepsen and Jepsen (2015) use data from the 2000 Decennial Census to show that same-sex couples in the U.S. have smaller earnings gaps than opposite-sex couples, mostly due to differences in the presence and number of children. Black et al. (2007) find that same-sex couples are more likely to be dual-earner households than opposite-sex couples. Though when they are not dual-earner households, the "stay-at-home" partner is more likely to be the one with lower education for both opposite-sex and same-sex partners. Both same-sex and opposite-sex couples are more likely to be single-earner households once they have children.[2] In terms of empirical evidence of specialization, Hofmarcher and Plug (2022) use the ATUS and find that opposite-sex couples tend to specialize more than same-sex couples. This difference is driven by strongly specializing couples where the female partner is out of the labor force. Giddings et al. (2014) use ACS and Decennial data and find that there is less specialization in same-sex couples as compared to opposite-sex couples, but this gap is shrinking over time.

[2] Additional evidence from Hansen et al. (2020) shows that marriage equality laws increased specialization within same-sex female households. Since married same-sex couples are more likely to have children than unmarried, that increase in specialization could be in line with Black et al.'s 2007 application of Becker's theory.

Kurdek (1993) surveys about 300 *childless* couples in the U.S. and finds three patterns for the distribution of household tasks: same-sex female couples follow an "equality pattern" where either partner is equally likely to undertake a task; same-sex male couples follow a "balance pattern" where partners specialize in tasks that are equally distributed; and opposite sex couples follow a "segregation pattern" where one partner does most of the tasks (and this partner is the woman). Thus, same-sex couples distribute tasks more equally than opposite sex couples but the equality is attained in male same-sex couples through specialization and in female same-sex couples by sharing or alternating tasks.

Traditional gender norms that place the time cost of childcare on the female partner will function differently, if at all, in same-sex couples. However, other norms may take their place. For example, in female same-sex couples, there could be a norm that the birth mother takes on more of the childcare duties, as Andresen and Nix (2022) find in the short run for same-sex female couples in Norway. Similarly, in male same-sex couples, there could be a norm where one partner takes on more of the childcare if they are genetically related to the child.

Prickett et al. (2015) use the American Time Use Survey (ATUS) to examine how parents spend time with children. They find that women in both same-sex and opposite-sex couples and men in same-sex couples spend significantly more time with children than men in opposite-sex couples. Additionally, there may be a stronger culture of coparenting in female same-sex couples. Genadek et al. (2020) use the ATUS to examine the amount of time couples spend together and find that female same-sex couples spend more time together overall and coparent more than opposite-sex couples, who in turn coparent more than male same-sex couples.[3] Using

[3] Both studies using the ATUS caution about interpretation given the very small sample sizes for same-sex couples.

international data and focusing on four household tasks (not including childcare), Bauer (2016) finds patterns similar to those in Kurdek (1993): that female same-sex couples share housework tasks more than male same-sex couples and more than opposite-sex couples (the percent of shared housework tasks are 41%, 33%, and 28% respectively). However, Bauer finds the presence of children "traditionalize the domestic work pattern," that is, increase task segregation for both female same-sex couples and opposite sex couples (but not for male same-sex couples).

Another difference between same-sex and opposite-sex couples is their family structure. Same-sex couples tend to have fewer children and the children they have are typically older. The children of same-sex couples are more likely to have come from earlier relationships (Genadek et al., 2020). Same-sex couples are also more likely to have adopted children or stepchildren than opposite-sex couples (Badgett et al., 2021). Walker and Taylor (2021) estimate that 17.2% of married same-sex couples (5.9% of unmarried) live in households with only adopted children, compared to 1.7% of married opposite-sex couples (0.6% of unmarried). Same-sex couples are also more likely to foster children, with 1.7% of married same-sex couple households (1.2% for unmarried partners) fostering children compared to 0.2% of married opposite-sex couple households (0.1% for unmarried partners).

2.3 Theoretical Framework for Household Choice after Entry of First Child

We develop a theoretical framework for understanding household choices made around the arrival of the first child in a household. Our focus is on decisions made after the entry of the child, so we do not model differences in preferences for having children. Since we do not have information on the identity of the birth parent, we do not model differences based on this important, but omitted, characteristic.[4] We also do not explicitly model any additional constraints that same-sex couples may face in becoming parents, which may differ from those faced by opposite-sex couples.[5]

We define two types of people, male (M) and female (F), who each receive a random initial draw from earnings distributions with mean $\bar{w_M}$ and $\bar{w_F}$ where $\bar{w_M} > \bar{w_F}$. Earnings are distributed log-normally except for a wage floor at zero; variance is equal across the two types. After the initial wage draws from their respective earnings distribution, individuals' wages increase annually based on the number of hours they work.[6] Full-time workers see a larger growth in their wages than part-time workers each year. We consider three types of couples consisting of different combinations of types M and F individuals: opposite-sex couples (MF), male same-sex couples (MM), and female same-sex couples (FF).

The utility function for the household can be represented as:

$$U = U(C, l, h) \tag{2.1}$$

Where:
C is the household's consumption level
l is the household's leisure time
h is the household's pre-child home production hours

We assume that the utility function is separable and additive, and exhibits constant returns to scale. The household's consumption level is determined by a Cobb-Douglas production func-

[4] Like opposite-sex couples who face fertility issues, same-sex couples also face biological issues that can affect which, if either, of the members of the couple will be biologically related to the child and, for female same-sex couples, if either member of the couple carries the child through pregnancy. For couples who adopt, the timing and age of the child at adoption are also potential choice variables.

[5] Some of these additional constraints reflect existing laws that treat parental rights differently for same-sex couples. For example, while all 50 states allow same-sex couples to adopt children, 17 states have no explicit protections against discrimination based on sexual orientation or gender identity, and 12 states allow state-licensed child-welfare agencies to refuse to place children in families if doing so conflicts with their religious beliefs (Project) In addition, since same-sex couples are less likely to live near their places of birth (Black et al., 2007), they may be less able to rely on extended family resources for childcare, limiting their choice set.

[6] A more complicated model would allow for wages to vary with flexibility (i.e., a lower paying job that provides greater time flexibility). For now, we think of our model as proxying for the choice of lower pay/more flexibility by simply modeling the adjustment in total hours.

tion that combines the earnings of both individuals in the couple. The household's leisure time is determined by a CES production function that allows for complementarity or substitutability between the leisure time of the two individuals. The household's pre-child home production hours are determined by a linear production function that sums up the home production hours of both individuals.

The household faces a time constraint given by the change in household production, which can be represented by the following equations:

$$k_1 + k_2 + (h_T + Z) + s_1 + s_2 + l_1 + l_2 = 48 \tag{2.2}$$

$$h_T + Z = h_1 + Z_1 + h_2 + Z_2 \tag{2.3}$$

Where:
Subscripts refer to adult individuals $i = 1$ and 2 in the couple
k_1 and k_2 are the hours worked per day
s_1 and s_2 are the hours used for sleeping
l_1 and l_2 are the hours per day used for leisure
h_T is the household's pre-child home production hours
$h_T + Z$ is the household's post-child home production hours
Z_1 and Z_2 are the additional hours per day used for home production after the child's arrival
48 represents the maximum available hours per day at the household level

The household maximizes its utility subject to the time constraint above and to the budget constraint:

$$W_1 k_1 + W_2 k_2 = C \tag{2.4}$$

Prior to the arrival of their first child, both type M and type F individuals work k_i hours per day, resulting in annual earnings of $Earnings_i = w_i * k_i * 5 * 52$. The remaining time is allocated to sleeping (s hours, assumed to be fixed for both M and F at 8 hours), home production (h hours), and leisure (l hours). In the pre-parenthood period, the joint contribution of each couple to household production is $h_T = h_1 + h_2$ hours.

When the first child enters the household, the amount of time required for household production increases. The joint household production that occurred pre-parenthood can be reduced or performed concurrently with childcare some of the time, so the total household production required after having a child is $h_T + Z$ hours where $Z < 24 - s_i$. Parents may use their leisure time for some of this extra production, but even if the leisure time were non-overlapping between the members of the couple and could be used efficiently for childcare, there will still be a deficit of home production that will need to be compensated for by reducing the number of hours spent on paid work.

The couple will choose how to allocate the additional household production hours among themselves using one of four possible scenarios. (1) Split Equally: both partners reduce their work hours equally by switching to a part-time work schedule. (2) Specialization in Earnings: the higher-earning partner maintains their full-time job while the lower-earning partner performs all of the extra household production. (3) Specialization in Preferences: the partner with the stronger preference for home production or weaker preference for work, regardless of their earnings, performs all the extra household production. (4) Gender Norms: the couple follows traditional gender norms to determine which partner takes on the extra household production.[7] These four scenarios are further formalized below.

2.3.1 Scenario 1: Split Equally

In this scenario, the couple will reduce their work hours by the same amount and allocate the additional time to home production. The constraint for this scenario can be represented as:

[7] If we had information on birth parent or donor parent, we could broaden this to include other social norms such as the birth parent being more likely to leave the labor market (relevant for FF couples) or the donor parent being more likely to leave the labor market (relevant for MM couples).

$$k_1' = k_1 - Z/2 \tag{2.5}$$

$$k_2' = k_2 - Z/2 \tag{2.6}$$

Which reduces the budget constraint so that

$$W_1 k_1' + W_2 k_2' = W_1(k_1 - Z/2) + W_2(k_2 - Z/2) = C - ((W_1 Z)/2 + (W_2 Z)/2) \tag{2.7}$$

This scenario predicts that earnings of both members of the couple decrease as does the total household budget after the child enters the household.

2.3.2 Scenario 2: Comparative Advantage in Earnings

In this scenario, the partner with the lower earnings will take on all the additional home production, while the other partner's hours will remain unchanged. The constraint for this scenario can be represented as:

$$k_1' = k_1 \tag{2.8}$$

$$k_2' = k_2 - Z \tag{2.9}$$

The household budget constraint is reduced to

$$W_1 k_1' + W_2 k_2' = W_1 k_1 + W_2(k_2 - Z) = C - W_2 Z \tag{2.10}$$

Where $W_1 > W_2$. This scenario predicts that the earnings of only the partner with lower pre-child earnings will decrease.

2.3.3 Scenario 3: Specialization in Work-Life Preferences

In this scenario, the partner with a larger preference for home production or smaller preference for work, regardless of earnings level, will take on all of the additional home production, while the other partner's hours will remain unchanged. Let $i = 1, j = 2$ if $U_1(h) > U_2(h)$, otherwise $i = 2, j = 1$. The constraint for this scenario can be represented as:

$$h_T + Z = h_i + Z_i + h_j \tag{2.11}$$

And the household budget constraint is reduced to

$$W_i(k_i - Z) + W_j k_j = C - W_i Z \tag{2.12}$$

This scenario predicts that the earnings of only the partner with lower preference for work will decrease. The total household budget will be reduced by the wages of the individual with higher household production preferences for every hour of time that is used for household production rather than labor.

2.3.4 Scenario 4: Gender Norms

In this scenario, the couple will follow gender norms in allocating additional home production hours. We assume that this scenario applies only to opposite-sex couples. Let $i = W$ if the individual is female. The constraint can be represented as:

$$h_T + Z = h_t + Z_i \tag{2.13}$$

And the household budget constraint is reduced to

$$W_i(k_i - Z) + W_j k_j = C - W_i Z \tag{2.14}$$

This scenario predicts that the earnings of only the female individual in the MF couple will decrease. The total household budget will be reduced by the wages of the female individual.

2.3.5 Expansion to Dynamic Model

In a two-period model (before and after child), all four scenarios result in earnings losses at the household level if the two spouses earn the same amount. When individual 1 earns more than individual 2 and earnings are not perfectly correlated with preference for work or gender, the second scenario results in the smallest household earnings loss.

To expand beyond the two-period model into a dynamic setting, we simulate this model using a synthetic set of 2,000 individuals with $\bar{w_M} = 30, \bar{w_F} = 25, \sigma_M = \sigma_F = 10$. The individuals are categorized such that there are 1,000 men, 1,000 women, 500 opposite-sex couples, and 250 each of male same-sex and female same-sex couples. We categorize individuals within a couple as primary or secondary earners based on their initial earnings draw. We simulate the four scenarios for all three couple types with the exception of Scenario 4 which we only simulate for opposite-sex couples.

After the first year, those couples that earn enough (in our simulations, determined by earnings being larger than average earnings in that period) can opt to pay for childcare. Childcare can cover half of the hours deficit in the second through fourth years and the full deficit starting in the fifth year. Thus, the initial increase in required number of hours for household production that was brought on by the child declines over time when childcare can be purchased (and as long as the household does not add other children). Some families choose not to use childcare, even if they can afford it, if the partner working less than full time has a high preference for home production and/or a low preference for work. In this model, a high preference for home production can also be conceptualized as a distaste for paid childcare.

Figure 2.1 shows results from our simulation exercises for both opposite-sex and same-sex couples (panel A and panel B, respectively). The overall trends are similar across same-sex and opposite-sex couples. In Scenario 1 (Split Equally), all individuals experience a small decrease in earnings after having a child, with a subsequent recovery once childcare becomes available. In Scenario 2 (Comparative Advantage in Earnings), primary earners (both male and female) maintain their pre-parenthood earnings trajectory, while secondary earners suffer large earnings losses that do not fully recover. In Scenario 3 (Specialization in Preferences) there is a decrease in average earnings for all groups, with a somewhat smaller decline for secondary earners (relative to pre-parenthood earnings). The recovery in earnings is steadier than in the equal split scenario, since one member of the couple continues to earn full-time wages and experiences earnings growth each period. Finally, in Scenario 4 (Gender Norms) among opposite-sex couples both primary and secondary earner women experience a decline in earnings. The earnings growth for secondary earner men is slower than that for primary earner men, as earnings growth is based on pre-parenthood earnings levels.

In our empirical exercises, we are able to test for Scenarios 1, 2, and 4. Unfortunately, we do not have a good measure to determine which member of a couple has a stronger preference for market work versus household work, so we are unable to explicitly test for Scenario 3. Since preference over market work is likely correlated with pre-parenthood earnings, simply comparing the empirical trends to the simulation results does not provide a strong test of Scenario 3. For example, the model predicts that earnings would decrease for all groups, but if empirically those who have a lower preference for market work are more likely to be secondary earners, seeing a larger drop in earnings for secondary earners would also be consistent with Scenario 3.

The model presented here focuses on the choices made by couples, rather than the choices of employers, except through the lower average pre-parenthood wage for women. However, discrimination in the labor market, particularly against women and individuals in same-sex couples, has been well-documented in the literature. This and other factors could drive a wedge between the patterns we see in the actual data and the results from this simple model of household decision-making.

2.4 Data and Methodology

We rely on two primary data sources, the American Community Survey (ACS) and the Longitudinal Employer-Household Dynamics (LEHD) dataset. The ACS provides us with demographic and relationship information, while the LEHD provides us with administrative earnings data used to construct our outcome variables. We also use administrative data from Internal Revenue Service (IRS) 1040 forms and the Census Numident to verify key variables from the survey data, in particular the date of birth and date of child entry into the household. We link these various sources of data using unique identifiers called Personal Identification Keys (PIKs) to construct a comprehensive dataset, as we detail in Appendix B.1.

2.4.1 American Community Survey (ACS), IRS forms, and Numident

For our analysis, we focus on a sample of couples with children from the 2013-2019 ACS. The ACS is sent to approximately 290,000 addresses each month, totaling about 3.5 million addresses annually. Statistics from the ACS are currently published in annual formats of 1-year and 5-year estimates; we use the 1-year estimates for 2013-2019. The ACS collects detailed information for up to 5 people in the household and less detailed information for up to 7 additional household members.

We identify couples and children in the ACS using question 2, which asks about the relationship to "Person 1." The question includes 15 checkbox responses for 2013-2018, with the options "Husband or wife" and "Unmarried partner" for identifying couples. From 2019 onward, the relationship question responses are disaggregated into further categories: "Opposite-sex" and "Same-sex" (see Appendix B.1, and especially Figure B.1, for a detailed discussion of this question over time). Thus, prior to 2019 we determine "opposite-sex" and "same-sex" using the relationship question combined with the question about sex, while in 2019 we can simply use the relationship question.[8] In all years, we define a couple as "spouses" or "unmarried partners." This definition is based on a point-in-time indicator of couplehood, as we do not have a longitudinal measure of couple status. This means we cannot confirm that couples are together during the entire earnings period in our sample.[9]

The question about sex (question 3) provides checkboxes for male and female. We are concerned that survey errors may lead to discrepancies between the survey-reported same sex couple status and actual same-sex couple status. Kreider et al. (2017) note that "when two groups are related, and a very small proportion of the large group makes mismarks, their answers can affect the estimates of the smaller group." As a robustness exercise, we carry out a second approach to defining couples. We identify couples via the IRS 1040 forms, characterizing those who file jointly as couples.[10] We then use the Census Numident, which sources sex from Social Security

[8]Until 2013, people of the same sex who reported themselves as married in the American Community Survey were edited to be unmarried partners.

[9]In order to pin down partners in the household during the period around the entrance of the child, we could narrow our focus to married couples and use the date of marriage question in the ACS. However, this is not a viable strategy at this time because of the insufficient same-sex married couple sample size.

[10]We require the household to have at least three people (partners and at least one child).

Administration (SSA) administrative records, to define both the sex of the individuals and the date of birth of the child.[11] Basing our sample on joint tax filers gives us a much larger sample, but limits the focus to married couples only. Our analysis using these two different ways of constructing the sample leads to very similar results.

The ACS relationship question includes four checkbox responses for children (biological, adopted, step, and foster children). In our analyses, we include biological, adopted, and stepchildren, but exclude foster children due to difficulties identifying when they enter and exit a household, as well as the different relationship dynamics involved. We use birth dates in the ACS to identify the timing of the child's birth for up to three children per household and we validate those birth dates using the Census Numident.[12] To accurately identify when the child enters the household in the case of adoption, we use IRS 1040 data, using the year in which the child is first claimed by the head and/or partner as the year of entry for cases in which the child was adopted.[13]

According to data from the 2019 ACS, there are around 66 million couples in the U.S., with the majority being married couples (58 million as compared to 8 million unmarried partners) (Walker and Taylor, 2021). Of these 66 million couples, 65 million are opposite-sex couples. The remaining one million are same-sex couples, comprised of 462,215 male same-sex couples and 518,061 female same-sex couples. This means opposite-sex couples account for roughly 98.5% of couples, with male same-sex couples and female same-sex couples comprising 0.7% and 0.8%

[11] Note in both survey and administrative data, the concept collected is sex, not gender. By including only those people who have both survey-reported and administrative sex that match, we would likely exclude a subsample of transgender couples. Without more detailed questions on sexual orientation and gender identity (SOGI), these individuals, as well as LGBTQIA+ individuals that are not part of a couple at the time they are surveyed, are difficult to identify in the data. An excellent overview of current research into this measurement question is the National Academies of Sciences, Engineering, and Medicine report: https://www.nationalacademies.org/our-work/measuring-sex-gender-identity-and-sexual-orientation-for-the-national-institutes-of-health.

[12] Question 4 of the ACS asks about the age of the person and their date of birth. Since we are constraining our sample to couples, the couple accounts for two of the five persons about which detailed questions are asked.

[13] See Appendix B.1 for more details about the measurement of child-timing for adopted children.

respectively. It is important to note that our sample will differ from the above estimates because we are only including couples with children, we pool our sample over multiple ACS years, and we only include couples who both match to the LEHD employment data pre-parenthood. Table 2.1 provides the number of same-sex male, same-sex female, and opposite-sex couples in our ACS sample. It shows that opposite-sex couples account for about 99.4% of couples in our sample, with 0.1% male same-sex couples and 0.4% female same-sex couples.

Fisher et al. (2018) show that only 0.48% of joint tax filers are same-sex couples, but ACS survey data shows a higher rate. There are several reasons why we would expect survey data to yield a higher number of same-sex couples than joint tax filings. First, one member of the couple may not earn taxable income. Second, they may choose not to file jointly; they may have to file separately for their state taxes and choose to file federal taxes separately as well for logistical ease. They also may be following a practice of filing separate taxes that they established before 2013 when they could not file federal taxes jointly. Fisher et al. study tax filings from 2013 to 2015; those years do not fully capture the additional time it takes couples to adjust their filing practices after the legalization of same-sex marriage, which did not happen at the federal level until 2015.

Fisher et al. (2018) also find that 7% of male same-sex joint tax filers and 28% of female same-sex joint tax filers have dependents. Interestingly, given the differences in the total percentages, the rate of same-sex couples with children in the tax data is roughly similar to the 10% of male and 22% female same-sex couples with children in the 2008-2018 ACS survey data (Badgett et al., 2021). Joint tax filers may overrepresent couples with children, since there are several tax breaks that make filing jointly more beneficial for those with dependents. On the other hand, in couples with dependents, one of the partners is more likely to not have taxable income (Black et al., 2007).

2.4.2 Longitudinal Employer-Household Dynamics (LEHD) Data

We use data from the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) to measure earnings. These data provide quarterly earnings records collected by state unemployment insurance (UI) programs, linked to establishment-level data from the Quarterly Census of Employment and Wages (QCEW). LEHD data cover over 95% of private-sector workers and almost all state and local government employment (Abowd et al., 2009). We use LEHD data for all private-sector and public-sector employers from 2005-2021. The earnings measure we use throughout the paper is aggregate earnings from all employers in a given quarter and are real (2010) earnings. We winsorize earnings at the 99th percentile.

Our sample of parental earnings is an unbalanced panel. We include births as early as 2000 in our sample and as late as 2019 the last year of ACS available to us. We estimate 20 quarters (5 years) forward, so couples with children born in 2016-2019 will not be represented in all quarters. The analysis using a balanced panel, excluding these partially covered couples, gives qualitatively similar results.

2.4.3 Estimating the Earnings Path

We adopt the identification strategy used by Kleven et al. (2019), Andresen and Nix (2022), and other researchers estimating the child penalty. We rely on the exact timing of the event, namely, the year in which the first child enters the household, which creates a sharp discontinuity in household production requirements. Our identifying assumption is there are no other factors that would cause a discontinuous change in earnings when the child arrives, except for the additional household production required by the arrival of the child itself. In the following discussion of our strategy, we use the term "arrival" or "entrance." It represents date of birth for biological and stepchildren, but for adopted children the relevant event is when the child becomes part of the household.

$$y_{it} = \sum_{k=T_l}^{T_u} [\alpha_k \mathbb{1}(t - eventquarter_i = k)] + age_i + quarter_t + u_{it} \tag{2.15}$$

We estimate Equation 2.15 where the summation term represents a series of dummy variables indicating the distance in time from the quarter in which the first child arrives. We set the lower bound to be more than 2 years before the arrival of the child (quarter -9) and the upper bound to be more than 5 years after the arrival of the child (quarter 21). The "treatment" of becoming a parent is absorbing, so that those who have a child remain treated. This both matches our perception of reality and avoids the measurement error from using those who have already been treated as controls for those currently being treated, as discussed in Goodman-Bacon (2021). We omit quarter -6 so that our estimates represent total earnings relative to earnings a year and a half before the arrival of the first child. To control nonparametrically for life-cycle and time trends we include age of parents at child arrival and fixed effects for each quarter. Our identifying variation then comes from variation in the ages of the parents and the exact quarters in which their children join the household. The disruption from having a child comes at different points in their age/earnings trajectory and at different calendar times.

For easier interpretation, we transform earnings levels into the percentage of the counterfactual earnings in the absence of a child. We use this approach instead of using log earnings to retain those in the sample who have zero quarterly earnings after the entrance of the child.

Following Kleven et al. (2019), we define

$$P_t = \frac{\hat{\alpha}_t}{E[(\tilde{y_{it}}|t]} \tag{2.16}$$

where $E[\tilde{y_{it}}|t]$ represents the average predicted earnings from a regression without the event dummies, t refers to the event time (measured in quarters since arrival of first child), and $\hat{\alpha}_t$ is the estimated coefficient on the respective event time dummy taken from estimating Equation 2.15 above.

2.4.4 Defining Comparable Within-Couple Earner Types

We estimate Equations 2.15 and 2.16 separately for twelve different groups, as identified in Table 2.1. We first separate individuals into two groups based on their relationship status: opposite-sex couples and same-sex couples. This allows us to compare the earnings trends for these two types of couples, which is the focus of this paper. Our method for identifying same-sex individuals does not allow us to impute sexual orientation of uncoupled individuals, so we do not include uncoupled individuals in our analysis. Among each couple type, we further distinguish the partners by their pre-parenthood earnings, creating another three groups each.

By separating couples by pre-parenthood specialization patterns, we can test the model introduced in Section 2.3, since the predicted patterns vary by whether couples divide the child-induced additional household labor by equal allocation (Scenario 1), comparative advantage in earnings (Scenario 2), work-life preferences (Scenario 3), or following gender norms (Scenario 4). As discussed in Section 2.3, because we do not have a reliable measure of preferences over market work, we can provide only suggestive evidence regarding Scenario 3. As a proxy for the specialization that has occurred in the partnership in advance of the child, we identify the partner with the higher earning potential based on their earnings before the entrance of the child. This split by earnings potential can also proxy for the bargaining power each partner has at the point when the child enters the household. We characterize individuals as the primary or secondary earner using their annual earnings from two years prior to the entrance of the child (that is, quarters -7 to -4). Primary earners are defined as those who make at least 10% more than their partner; secondary earners are those who make at least 10% less than their partner. Partners who make within 10% of each other in the period before the entrance of their first child are defined as equal earners. The size of these relative groups is shown in Table 2.1.

The higher earnings potential designation based on pre-parenthood earnings is preferred over other approaches to differentiate between partners in same-sex couples such as using the Census-designated householder or the birth mother as the primary/secondary partner. The former is highly correlated with the earnings-based designation for opposite-sex couples but only minimally correlated for same-sex couples (Antecol and Steinberger, 2013), while the latter, though used by Moberg (2016) and Andresen and Nix (2022), is only applicable to female same-sex couples and does not address male same-sex couples or the many couples who adopt. A simpler version of our measure, without the equal earner category, has been used by Antecol and Steinberger (2013) to study earnings gaps among same-sex and opposite-sex couples; by Hansen et al. (2020) to investigate marriage premiums among gay and lesbian couples; and by Martell and Nash (2020) to examine the effect of legalized marriage on hours worked.

Table 2.2 is a transition matrix showing how many individuals move from being in a specialized couple to being in an equal-earner couple after the entrance of the child. We define pre-child specialization using the earnings in the one-year period starting one and a half years before the child enters; similarly, we define post-child specialization using the earnings in the one-year period starting one year after the child enters. For both same-sex and opposite-sex couples, the majority of individuals do not switch earner type. About 86% of same-sex couples and about 88% of opposite-sex couples remain specialized or remain equal earners.

If our measure of higher earnings potential was highly variant over time, we might be concerned about relying on such a definition by which to differentiate partners. However, the opposite is true: our measure is fairly constant over time, lending support to our use of this measure to proxy for pre-parenthood specialization. Though most do not change, there are some transitions between specialized and equal earner types brought on by the entrance of a child. Among opposite-sex couples, slightly more individuals would be defined as equal earners after the entrance of the child (5.30% as compared to 6.45%). Among same-sex couples, the opposite is true: individuals are slightly more likely to be specialized after the entrance of the child (8.51% as compared to 5.92%).

2.4.5 Propensity Score Weighting

Individuals in same-sex couples differ from those in opposite-sex couples in various ways as discussed in Section 2.2. Table 2.3 provides summary statistics for four groups: men in opposite-sex couples, women in opposite-sex couples, men in same-sex couples, and women in same-sex couples. Men in same-sex couples are more likely to have a college degree than men in opposite-sex couples (62% versus 53%). Females in same-sex and opposite-sex couples are about equally likely to have a college degree (61% versus 63%). Women and men in same-sex couples are more likely to be Hispanic or Black than their counterparts in opposite-sex couples. Women and men in same-sex couples also earn more than their counterparts in opposite-sex couples. However, women and men in opposite-sex couples are much more likely to be married, with about 90% of individuals in opposite-sex couples being married, compared to 78% of men and 65% of women in same-sex couples.[14] On average, individuals in same-sex couples have fewer children (about 1.7 compared to 1.9 for opposite-sex couples) and are older when their first child enters the household (about 2 years older than their opposite-sex counterparts).

Table 2.4 shows the pre-parenthood values of our main outcome variables, earnings and employment, disaggregated further into our twelve analysis subsamples defined by sex, couple type, and earner type. All earner types in same-sex couples have higher average earnings than their counterparts in different-sex couples. Primary and equal earners of both sexes and couple type have employment rates at six quarters before the birth of over 90%, highlighting that this is a selected sample of parents that are generally highly attached to the labor force pre-birth. However, secondary earner men in opposite-sex couples are the least likely to be employed pre-birth, at 68%.

We have described several ways in which opposite-sex and same-sex couples differ based on characteristics that are observable in our data. For our analysis, we want to verify that any differences in earnings trends that we find are not due to same-sex couples simply being different from opposite-sex couples in pre-parenthood characteristics that are related to their income paths after their child's arrival. In other words, we want to ensure that same-sex couples and opposite-sex couples have similar background characteristics so that we are comparing couples whose income paths ought, other than for the introduction of a child, to be the same as one another. Therefore, we weight our main results by inverse propensity scores, which creates a sample of

[14] Same-sex marriage was only available nationwide during part of our sample, with the Obergefell decision in 2015, which may explain much of this difference in marriage rates. The most recent cohorts, who had marital rights for a longer period before our sample period, may look more similar on this margin.

opposite-sex couples that is more similar to our sample of same-sex couples.

We first predict the likelihood of being in a same-sex couple using the following observable characteristics of couples: the oldest age among the couple, the gap in ages within the couple, the highest years of education, the gap in years of education, race, whether the couple is multi-racial, ethnicity, whether the couple is multi-ethnic, marital status, the gap in pre-child earnings, and the highest pre-parenthood income among the two. Using this predicted probability, we assign weights to give more predictive power to individuals who are more like individuals in the other type of couple. Opposite-sex couples who are more like same-sex couples are more highly weighted; same-sex couples who are more like opposite-sex couples are likewise more highly weighted. Specifically, the weight is calculated as the inverse of the predicted probability for same-sex couples and the inverse of one minus the predicted probability for opposite-sex couples. We combine these inverse propensity weights with ACS sample weights to form the final weights used in our regressions. Results that do not control for differences between same-sex and opposite-sex couples (that is, with no propensity weights) are presented in the appendix (see Figure B.2) and yield qualitatively similar results.

2.5 Results

We start by presenting the percent change in earnings results for opposite-sex couples and same-sex couples. For each of the types of couples, we show the percent change in earnings after the entrance of the first child at the individual level, then extend the analysis to consider the impact at the household level. We then explore intensive and extensive margin responses and discuss various ways in which our results are robust to observable differences between opposite-sex and same-sex couples. Finally, we highlight implications for our theoretical model.

2.5.1 Opposite-Sex Couples Earnings

Figure 2.2 Panel A shows the percent change in earnings after the entrance of the first child for opposite-sex couples. These earnings changes are relative to six quarters before the entrance of the child. The solid lines correspond to men and dotted lines correspond to women, while the color-symbol pairs (red circle, green square, and blue triangle) differentiate the earner types (primary, secondary, and equal). First focusing only on the difference between the male and female lines and not the earner types, we can see support for the findings of the previous literature. Men in opposite-sex couples continue their upward pre-parenthood earnings trajectories, while women experience a significant decrease after the entrance of their first child followed by a partial recovery of earnings. One year after the child's entrance, women's earnings are about 20% lower than they were before the arrival of their first child and even after five years their earnings remain about 18% lower than before their first child. This "motherhood penalty" has been well-documented in previous studies on the parental earnings gap.

We extend that literature by showing that the penalty exists across the split into earner types. All three types of opposite-sex female earners - primary, secondary, and equal -experience a decrease in earnings. In contrast, secondary earner opposite-sex men experience an increase in earnings beginning just before their child arrives, while primary and equal earner men see a consistent upward trend in their earnings that appears unchanged by the child entry.

The percent change in earnings is very similar in magnitude across earner types for women but varies more for men. Secondary earner men experience the largest increase in earnings in percent change terms relative to before the arrival of their child. By one year after the child joins the household, primary and equal earner men are earning 4% and 10% more than they were earning pre-child, respectively, whereas secondary earner men are earning about 30% more. That increase in earnings starts slightly before the child arrives; by the time the child enters the household, secondary earner men are already earning about 25% more than they were earning before the child. By five years after, though all men have a positive trend in earnings, secondary earner men are earning about 40% more than they were before their first child, while primary earner men are earning about 10% more, and equal earner men are earning about 20% more.

We find no evidence of a fatherhood earnings premium, or a discontinuous increase in earnings around child arrival, among primary and equal earner men in opposite-sex couples, but some evidence among secondary earner men. This finding adds to the existing literature on the circumstances under which a fatherhood premium arises. While previous research has documented the existence of a fatherhood premium (Glauber, 2018; Pal and Waldfogel, 2016; Yu and Hara, 2021), much of that work uses either cross-sectional analysis or small samples, making it difficult to control for selection into fatherhood. Mari (2019) shows that the impact of parenthood on fathers' earnings varies by country with wage premiums found in the U.S. and Canada but not in Norway or Denmark. Our findings suggest that in addition to heterogeneity in fatherhood premiums by country, there is also heterogeneity by household earner dynamics. We find that the fatherhood premium exists for fathers whose female partners earned at least 10% more than they did before the arrival of their first child. Similarly, Killewald (2013) uses panel data in the U.S. to explore heterogeneity in the fatherhood premium between different types of fathers and finds that the fatherhood premium exists only for married, residential, biological fathers whose wives do not remain full-time workers.[15]

Secondary earner fathers in our sample are significantly less attached to the labor force before the entrance of their first child, as Table 2.4 shows. We require individuals in our sample to have positive earnings at some point in the two years before the arrival of the child, but secondary earner men in opposite-sex couples are the least likely to be employed in our reference quarter (6 quarters before the entrance of the child). 68% of secondary earner men in opposite-sex couples are employed in that quarter. In comparison, 94% of primary earner men in opposite-sex couples are employed. Even among secondary earners, men in opposite-sex couples are less often employed than women in opposite-sex couples (81%), men in same-sex couples (79%), and women in same-sex couples (74%). In addition, secondary earner men in opposite-sex couples are earning the least six quarters before the entrance of the child. They earn on average \$5,206, less than half of their primary earner counterparts (\$12,250). That could explain why we see such an increase in earnings among secondary earner men in opposite-sex couples after child arrival. We explore this idea more in Section 2.5.4.

2.5.1.1 Opposite-Sex Couples Earnings Household

At the household level, Figure 2.3 Panel A shows a similar drop in total labor market earnings at the time of child arrival for all three types of opposite-sex couples. However, male-primary couples experience a slightly smaller decline. In the first quarter after child arrival, their household earnings decrease by about 8% compared to before the child. Female-primary and equal-earner couples face a larger drop of about 16%. In all cases, earnings begin to recover

[15] Our current definition of primary earner does not differentiate between primary earners whose partners are not in the labor market and those that are. In future work as we explore labor force participation, we can look at this distinction and compare to the existing literature.

two quarters after the first child arrives and fully recover by the end of our study period, five years later. By that time, equal-earner couples are earning approximately the same as before their child, male-primary couples are earning about 5% more, and female-primary couples are earning about 18% more. Based on the trends in individual earnings, it is likely that female-primary couples experience the largest increase in their total household earnings due to the large increase in earnings among secondary-earner men.

2.5.2 Same-Sex Couples Earnings

Figure 2.2 Panels B and C present the percent change in earnings after the entrance of the first child for female and male same-sex couples, respectively. Our estimates are less precise for same-sex couples due to the much smaller sample sizes, especially for men in same-sex couples. While our sample includes about 1.5 million opposite-sex couples, we only have about 6,000 female same-sex couples and about 2,000 male same-sex couples. We highlight two main differences between the earnings trends for same-sex couples and those for opposite-sex couples.

First, while men and women in opposite-sex couples have very different trends in earnings, men and women in same-sex couples have very similar trends. This result was not necessarily expected as previous literature suggests that male and female same-sex couples are different in ways that could be important to our analysis. For example, they enter parenthood at different rates Badgett et al. (2021). In our sample, men in same-sex couples earn more than women in same-sex couples (about $30,000 more on average) and are more often married (about 14 percentage points more). There are more Black women in same-sex couples than Black men (8% versus 6%) and fewer Hispanic women in same-sex couples than Hispanic men (10.8% versus 12.6%).

Nonetheless, the trends across sex of same-sex couples appear very similar.

Second, in contrast to opposite-sex couples we observe significant differences in same-sex couples when we look across earner types. There is a drop in earnings after the entrance of the child for both male and female individuals in same-sex couples of all earner types of around 2 to 15%, which is small relative to the 40% decline we see for opposite-sex women, but also does not follow the pattern of increasing earnings seen in opposite-sex men. However, after the initial drop, earnings trends diverge for primary, secondary, and equal earners. Same-sex secondary earners experience a quick increase in earnings relative to pre-parenthood, while primary and equal earners have no change in long-run earnings relative to pre-parenthood. By five years after the entrance of the first child, male and female secondary earners earn about 20% more than they did before the child and male and female equal earners earn about 5% more than they did before the child. Male and female primary earners are earning about the same as they were before the child.

As primary earners do not on average increase their earnings while secondary earners do increase their earnings, it appears that for same-sex couples, the response to parenthood among those who were specialized pre-child-entrance is toward an equalization of earnings rather than a divergence. The transition matrix in Table 2.2 shows that 6.5% of those who were specialized before the child would be categorized as equal earners after the child, which provides further evidence that same-sex couples who are specialized before the child on average move toward equalization after the child. However, Table 2.2 also shows that overall, same-sex couples are slightly less likely to be equal earners after the child than before (10.21 before versus 7.26 after). Figure 2.2 uses only pre-child earnings to define the earner categories for the analysis. This means that some of the individuals in the equal earner group would actually be categorized as primary or secondary earners based on post-child earnings. Because both of the individuals in a given equal-earner couple contribute to our estimate of the trend in earnings for that group, when one individual increases earnings and the other decreases earnings, those trends could net out to the zero change in earnings that we see in Figure 2.2. That suggests that the overall effect of child entrance on equalization of earnings among same-sex couples is unclear.

2.5.2.1 Same-Sex Couples Earnings Household

Figure 2.3 presents the results for household earnings which show similar changes in household earnings across same-sex versus opposite-sex couples, despite the differences at the individual level. Same-sex couples, like opposite-sex couples, experience a drop in household earnings of about 15% one quarter after their first child enters the household. By the second quarter, the drop lessens to about 5% less than earnings before the first child and household earnings fully recover by the end of the sample period (five years after the first child). In opposite-sex couples, the household earnings trends we see are a result of increases in male earnings which just about offset the decreases in female earnings, regardless of earner type. At the household level, the earnings trends for same-sex couples follow a similar pattern to those of opposite-sex couples, but the earnings trends within the couples that create these aggregate household trends are very different. Instead of a gap in earnings between individuals that aggregates to zero change at the household level, the relatively constant household earnings in same-sex couples are the result of the equalization of earnings between different earner types within the couple.

2.5.3 Intensive and Extensive Margins

To understand the contribution of changes in employment to the overall changes in earnings, we analyze the extensive and intensive margins of earnings. Specifically, we examine the percentage change in earnings due to changes in employment versus the percentage change in earnings from within employment. To achieve this, we replace quarterly earnings with the individual's pre-child earnings whenever the quarterly earnings are zero. This technique transforms our variable measuring changes in earnings from a negative to a zero value in any period where the individual has no earnings, relative to the baseline earnings level. This method is akin to that used by Groen et al. (2020) in their analysis. This substitution of earnings for zero values only works for levels of earnings, not percentage changes, so the results discussed in this section are in levels.

Figure 2.4 displays two lines: the change in total earnings for the entire sample and the change in earnings for the intensive margin. The gap between the total earnings and the intensive margin earnings is the extensive margin earnings, or the loss in earnings from those who experienced a change in their employment status from baseline.

For primary earner men in opposite-sex couples, we observe an increasing trend in earnings after the entrance of their child. However, this increasing trend would have been even higher if some men had not left their jobs during that period. Earnings one year after the child for primary earner men are \$1,700 higher if just including the intensive margin, but only \$700 higher for all earnings. At 5 years the gap is larger, with growth to \$3,700 at the intensive margin and only \$1,800 for all earnings. The earnings of equal earner men also exhibit an increasing trend that would have been even higher if not for those who experienced periods of zero earnings post-child.

In contrast, the total earnings line is above the intensive margin line for secondary earner men reflecting that secondary earner men are more likely to be employed after their child's arrival than at baseline. The increase in their earnings is partially due to an increase in earnings on the intensive margin and partially due to an increase in employment.

For women in opposite-sex couples, we observe a dip in both the total and intensive margin earnings after child arrival. However, the intensive margin earnings losses are smaller, -$1,400 versus -$3,000 losses measured for all earnings for primary earner women. They also recover to the previous level after just a couple of quarters such that primary earner women have had earnings growth at the intensive margin of $400 after one year, whereas the extensive margin earnings do not recover and are pulling down the total earnings level to a loss of -$1,600 at one year. The pattern holds for women in opposite-sex couples, regardless of whether they are the primary, secondary, or equal earner in the couple, suggesting that a major contributor to the depressed earnings of women in opposite-sex couples after child arrival is decreased employment.

In comparison, men and women in same-sex couples have small drops in earnings after child arrival on both the intensive and extensive margins. But after the first or second quarter, we see increasing trends in earnings at the intensive margin for all six groups. For primary and equal earners, there is a larger gap between total and intensive margin earnings with larger losses due to a change in employment, in comparison to secondary earners for whom there appears to be little change in employment status. For primary earners, the gap between intensive and extensive margin earnings is around $2,000 for both men and women five years after their child, while for secondary earners it is $900 for men and $600 for women. In this, the secondary earners of either sex in same-sex couples follow a pattern more like that of secondary earner men in opposite-sex couples than any of the other groups.

2.5.4 Implications for Our Model

In our model, we find that only when the division of childcare responsibilities is driven by gender norms would we observe the significant decrease in earnings of all female partners, including those with higher earning potential than their male partners, that we see in the data. This result aligns with Weeden et al. (2016) who argue that the wage gap is perpetuated by social norms that lead to gender-based differences in hours worked, and consequently earnings. Furthermore, Kleven (2022) shows that the average child penalty varies across space in the U.S. in line with local social norms around gender roles. Scenario 2 (comparative advantage in earnings) predicted that primary earners would not see a drop in earnings while secondary earners of both sexes would, but empirically we see that primary earner women do experience a drop in their earnings and secondary earner men do not. Scenarios 1 (equal allocation) and 3 (work preferences) predicted that all sexes and earner types would see a small drop in earnings, whereas empirically we see that opposite-sex men of all earner types do not see a drop in their earnings and opposite-sex female earners see a large drop. As mentioned in Section 2.3, the implications of our analysis for Scenario 3 are speculative at best, because we do not have a measure of preferences over market work that is uncorrelated with pre-parenthood earnings and gender norms. We can say, however, that the implications of Scenario 4 (gender norms) are consistent with the observed outcomes in our empirical exercises.

One inconsistency between the model and the observed outcomes is the pattern of earnings for secondary earner men in opposite-sex couples. In the model, when they follow a strategy of specialization based on gender norms, the earnings growth for these men is lower than that for primary earner men, because the earnings growth is entirely based on initial earnings. However, in the data we see that secondary earner men earn more to compensate for the loss of household income when their higher-earning partners experience a child penalty.

The results for same sex couples suggest an equal allocation of post-child household production and reject specialization based on potential earnings. However, the model does not fully capture some of the patterns seen in the data, particularly the secondary earner growth and the lack of growth for the primary earner. This may mean there are more complicated household bargaining dynamics at play that are not captured by our simple model.

The disaggregation into intensive and extensive margins shows some of those dynamics. In same-sex couples, the secondary earners are not adjusting their labor at the extensive margin. The intensive margin and total earnings lines run nearly parallel. However, the primary earners in same-sex couples are making extensive margin adjustments. We also see extensive margin adjustments among equal earners in same-sex couples, which may be why we see some shifting from the equal earner group to the specialized group in Table 2.2. This could be evidence that the same-sex couples are specializing according to work preferences, as in Scenario 3, if those preferences are not strongly correlated with the earnings categories.

2.5.5 Robustness Checks

We present three empirical analyses in this subsection as robustness checks on the results above. First, we look at results without propensity score weighting. Second, we consider results for adopted children. Third, we present results where we use future parents as a control group.

2.5.5.1 Without Propensity Score Weighting

We have so far shown that when a child enters the household, the earnings of women and men in opposite-sex couples respond differently than the earnings of women and men in same-sex couples. However, we have also shown in Table 2.3 (summary statistics; see Section 2.4 for a discussion) that women and men in opposite-sex couples are different from women and men in same-sex couples on many dimensions. Thus another interpretation of our results could be that the differences in earnings trends we identify are due to different rates of marriage, different rates of educational attainment, or different likelihoods of having a second child in the household. If that were the case, then controlling for those differences we should see no difference in the earnings trends of individuals in opposite-sex and same-sex couples.

In our main results, we use the propensity weights discussed in section 2.4 to create a sample of opposite-sex couples that are similar to same-sex couples on observable characteristics. This suggests that the differences we observe between the two groups are due to the gender mix of the couples and not to other factors. We perform two additional exercises to evaluate the influence of observable characteristics, both of which suggest that they have little influence on the differences we see.

First, we provide results without using propensity weights. Figure B.2 shows that our results remain almost identical; for individuals in both female and male same-sex couples, the unweighted earnings trends are slightly less positive than the weighted earnings trends. Earnings of primary earners more clearly show a slight decrease after the child enters the household, which remains constant over time at about 2% lower earnings than before the child for both men and women in same-sex couples. In comparison, the weighted estimates suggest no change or even an increase in earnings of about 2% for primary earners. The similarity between our results with and without weights suggests that the difference in earnings trends that we see are not driven by differences in characteristics of individuals in same-sex versus opposite-sex couples.

Second, we add controls one by one to our specifications to see which variable is driving the differences between opposite-sex and same-sex couples. We consider similar controls to those that create the propensity score: age, difference in age, education, difference in education, race, difference in race, ethnicity, difference in ethnicity, pre-child earnings, household earnings, and age of the child upon entrance to the household. We find that no variable makes a notable difference in the results. This is consistent with our comparison of earnings trends with and without propensity score weighting. Just as the propensity score did not significantly affect the overall patterns, the results with the controls added separately also show no differences in the patterns we identify.

2.5.5.2 Adoptive parents

Adoption is another dimension on which opposite-sex and same-sex couples differ. In Figure 2.5, we explore whether the difference in earnings trends can be explained by the differential rates of adoption among opposite-sex and same-sex couples. The analysis excludes biological children to focus solely on those whose first child was adopted, which includes same-sex and opposite-sex couples with adopted children. Kleven et al. (2019) found that adopting parents experience a similar long-term gender gap in earnings as biological parents, and our results confirm this finding. We go further to demonstrate that male and female same-sex couples both experience a drop in earnings after the entrance of their adopted child. Those results are consistent with the trends we see in our full sample, suggesting that adoption cannot explain the differences in earnings we observe between opposite-sex and same-sex couples.

While there are some small differences between the adoptive parent and full sample results, qualitatively the patterns are very similar. Individuals in male and female same-sex couples see a drop in earnings after the child enters, but their earnings quickly recover fully, with secondary earners seeing the largest increase in earnings relative to before the child. There is suggestive evidence that adoptive parent secondary earners in male same-sex couples see less of an increase in earnings after the child enters the household than their full-sample counterparts, but the confidence intervals do not reject that the trends are the same.

Like our main results, men in opposite-sex couples do not see a drop in earnings around the quarter in which the child enters, but instead see a steady positive trend in earnings. Secondary earner men in opposite-sex couples see the largest increase in earnings, though that increase is slightly smaller than in the full sample. The biggest visual difference between the adoptive and full samples is in the size of the drop in earnings for opposite-sex women. By five years after the entrance of the child, their earnings are about twenty percent lower than before the child, which is the same as in the full sample. But, as in Kleven et al. (2019), the initial drop in earnings is much smaller than the drop in the full sample; adoptive parent women may experience different parental leave patterns than biological parent women.

2.5.5.3 Future parents as control group

In all our specifications we treat the entrance of a child as an absorbing state, so already treated parents do not serve as controls for those who are not yet treated. In Figure B.3, we provide a robustness check that uses parents who have a child after our observation window as a control. We do this by shortening our observation period: we only include earnings up to 2014, although we continue to use births through 2019 from the American Community Survey (ACS). Parents whose children enter their households between 2014 and 2019 serve as controls for individuals who become parents before 2014. We argue that this is a better control group than individuals who never become parents. As in our main specification, we include couple type, age, and calendar quarter fixed effects so that the earnings effects are measured relative to individuals of similar age and couple type over time. We find that the patterns of earnings after the child's arrival remain similar when we use this control group.

2.6 Conclusion and Future Work

We show that same-sex couples experience different earnings trajectories after a child enters their household than opposite-sex couples. This suggests different couple types follow different decision rules on how to allocate the additional time burden of caring for a young child. Specialization in childcare and market work, with the choice of who specializes driven strongly by gender norms, seems to be the typical response of opposite-sex couples, while we see more complicated dynamics in earnings of same-sex couples.

As we have noted, same-sex couples face the same constraints as other working parents in the U.S., including lack of affordable childcare, lack of universal parental leave, and lack of mandated sick leave, which all make working and caring for children challenging, but same-sex couples may also face additional constraints. Moreover, we have made the simplifying assumption that work-life preferences across opposite-sex and same-sex couples are the same. Keeping in mind these potentially confounding factors, it is noteworthy that same-sex couples allocate their time differently than opposite-sex couples. These differences in behavior over the sex composition of couples provides suggestive support for the findings in Kleven (2022) that gender norms are one of the most important drivers of the child penalty for mothers, both in the U.S. and likely worldwide.

Future work will extend the current analysis to look at changes in the earnings gap within couples, in addition to the current analyses of individual and household level trends. We could also examine differences by more or less family-friendly occupations which could be used as a rough proxy for differences in work-life preferences as captured in Scenario 3. Finally, we appreciate the improvements in the concepts collected in the ACS that enable us to produce this detailed analysis and look forward to other possible changes in the future that could help expand this knowledge.

2.7 Tables

Table 2.1: Sample Sizes

Opposite-Sex Couples 1,519,000					
Male Primary	Male Secondary	Male Equal	Female Primary	Female Secondary	Female Equal
848,000	383,000	114,000	547,000	688,000	114,000
Male Same-Sex Couples 2,200			Female Same-Sex Couples 6,500		
Male Primary	Male Secondary	Male Equal	Female Primary	Female Secondary	Female Equal
2,000	1,400	350	5,900	4,300	850

Note: This table presents the number of observations in the analysis sample, broken apart by couple type, sex, and earner type. Same-sex couples are determined by the reported relationship status and reported sex of the individual and their partner in the ACS. Equal earners make within 10% of their partner's earnings; when the gap in earnings is larger than 10%, earners are designated as primary if they have higher earnings than their partner and secondary if they have lower.

Table 2.2: Specialization

	Opposite-Sex Couples			Same-Sex Couples		
	Specialized Pre	Equals Pre	Total	Specialized Pre	Equals Pre	Total
Specialized Post	86.72	5.30	92.03	83.87	8.51	92.38
Equals Post	6.45	1.52	7.97	5.92	1.70	7.62
Total	93.18	6.82	100.00	89.79	10.21	100.00
N	2,820,582			15,706		

Note: This table presents an evaluation of how much the earner type designation changes over time. Our designation categorizes individuals based on their earnings in the "pre" period defined as 6 and 9 quarters before the birth of the child. When the gap in earnings is 10% or less, both members of the couple are designated as equal earners; when the gap in earnings is larger than 10%, earners are designated as primary if they have higher earnings than their partner and secondary if they have lower. The "post" version instead uses earnings from 5 to 9 quarters after the child is born to categorize earner types. The resulting tabulation shows how many individuals change categories after they have their first child. Specialized refers to individuals who are either primary or secondary earners.

Table 2.3: Summary Statistics

	Opposite-Sex Couples		Same-Sex Couples	
	Men	Women	Men	Women
Age at Entrance	30.31	28.26	32.33	30.83
	(6.346)	(5.805)	(7.501)	(7.682)
Age of Child at Entrance	0.044	0.031	0.368	0.226
	(0.527)	(0.446)	(1.696)	(1.242)
Adopted Child	0.024	0.024	0.289	0.210
	(0.153)	(0.153)	(0.454)	(0.407)
Number of Children	1.903	1.902	1.750	1.649
	(0.895)	(0.894)	(0.843)	(0.827)
Prebirth Earnings	94,230	68,710	111,700	84,440
	(88,010)	(68,390)	(116,300)	(92,650)
Primary Earner	0.721	0.486	0.623	0.610
	(0.540)	(0.575)	(0.570)	(0.564)
Secondary Earner	0.364	0.598	0.468	0.469
	(0.558)	(0.574)	(0.583)	(0.572)
Equal Earners	0.085	0.085	0.090	0.079
	(0.278)	(0.278)	(0.287)	(0.270)
Married	0.905	0.905	0.782	0.647
	(0.293)	(0.293)	(0.413)	(0.478)
Black	0.056	0.046	0.060	0.080
	(0.229)	(0.210)	(0.238)	(0.271)
Hispanic	0.095	0.094	0.126	0.108
	(0.293)	(0.293)	(0.332)	(0.311)
Less than HS	0.053	0.031	0.056	0.043
	(0.225)	(0.174)	(0.231)	(0.202)
HS	0.414	0.338	0.327	0.353
	(0.493)	(0.473)	(0.469)	(0.478)
College	0.533	0.631	0.616	0.605
	(0.499)	(0.483)	(0.486)	(0.489)
Observations	1,345,000	1,350,000	3,800	11,000

Note: This table provides descriptive statistics about the analysis sample. We divide the sample into couple types (opposite-sex and same-sex) and further into men and women. All variables come from the ACS except for earnings (from the LEHD); primary, secondary, and equal earners (created using LEHD, ACS, and Numident information); and age at entrance and age of child at entrance (from the Numident and ACS).

Table 2.4: Prebirth Earnings and Employment

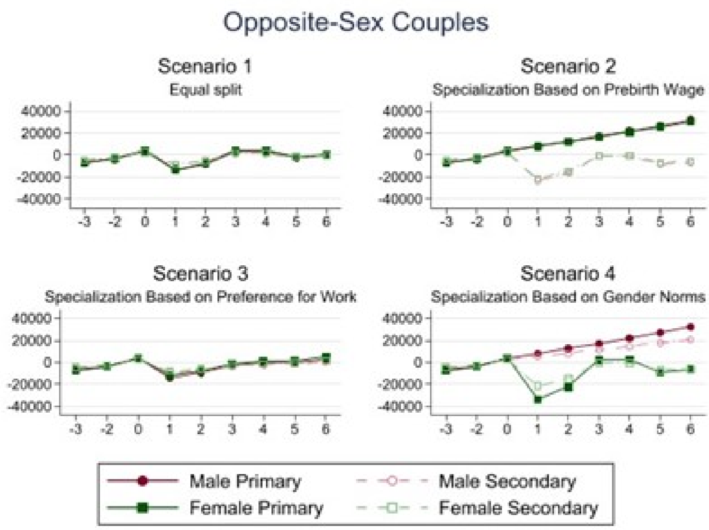
	Opposite-Sex Couples					
	Male Primary	Male Secondary	Male Equal	Female Primary	Female Secondary	Female Equal
Average Earnings	12,250	5,206	10,460	9,253	5,559	10,340
Percent Employed	94.1%	68.2%	96.2%	92.3%	80.5%	97.6%

	Male Same-Sex Couples			Female Same-Sex Couples		
	Male Primary	Male Secondary	Male Equal	Female Primary	Female Secondary	Female Equal
Average Earnings	15,270	7,321	13,410	11,390	5,610	11,420
Percent Employed	94.5%	79.0%	96.4%	92.9%	73.6%	95.8%

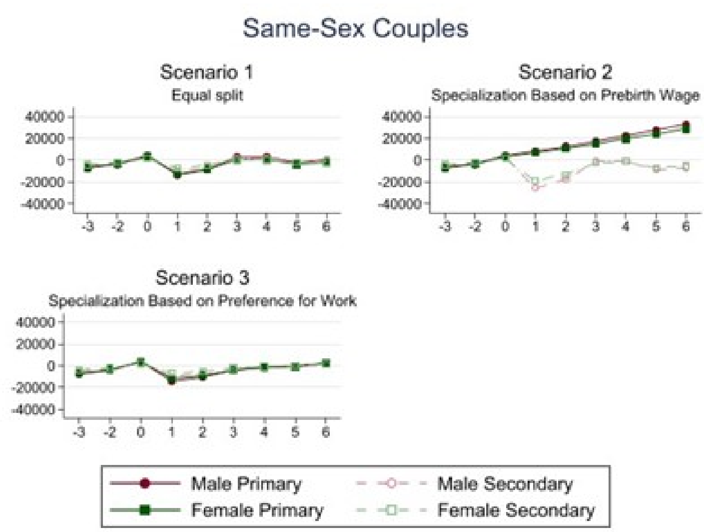
Note: This table provides descriptive statistics on the pre-parenthood values of our main outcome variables. The values are earnings and employment measured 6 quarters before the entrance of the child into the household. The sample is divided into the twelve analysis groups defined by couple types (opposite-sex and same-sex), sex (male and female), and pre-birth specialization (primary, secondary, and equal earners).

2.8 Figures

Figure 2.1: Theoretical Model Predictions



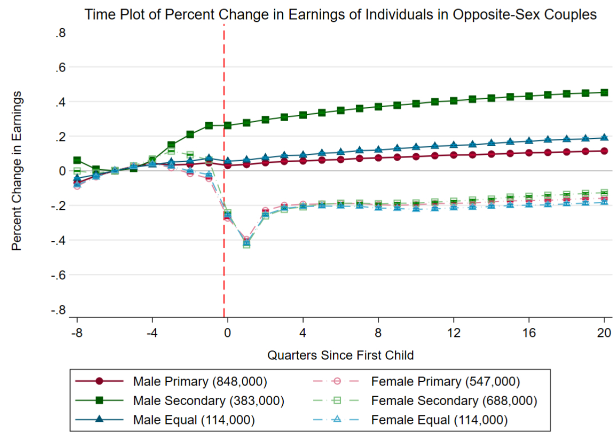
(a) Individuals in Opposite-Sex Couples



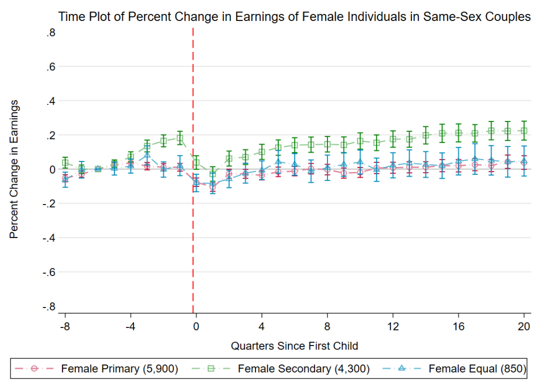
(b) Individuals in Same-Sex Couples

Note: This figure presents the predicted earnings path around child entry based on simulations of the theoretical model. The scenarios explore couples dividing their time either (1) equally, (2) based on comparative advantage in earnings, (3) based on preferences for home production, or (4) based on gender norms. See Section 2.3 for a full description of the model.

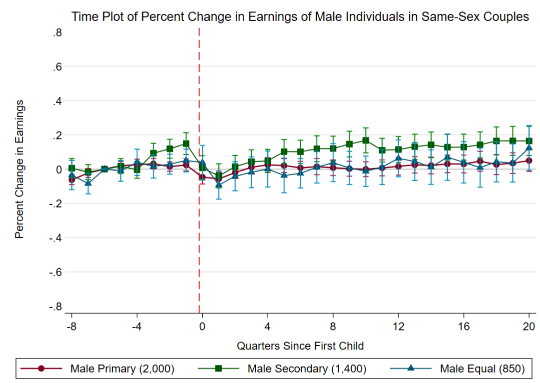
Figure 2.2: Individual Earnings after Entry of First Child



(a) Individuals in Opposite-Sex Couples



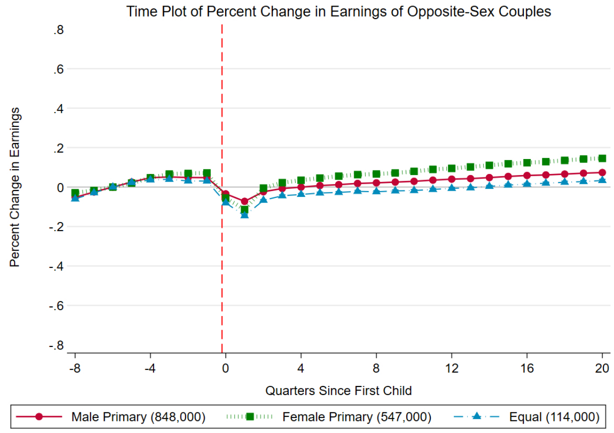
(b) Individuals in Female Same-Sex Couples



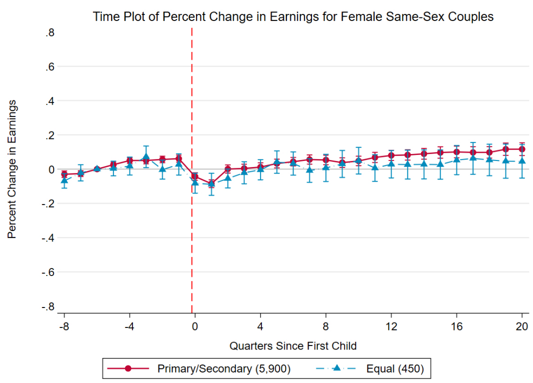
(c) Individuals in Male Same-Sex Couples

Note: This figure presents the main event studies of earnings around child entry at the individual level. For easier interpretation, earnings changes are presented as percent changes following Kleven et al. (2019). The results are weighted with propensity weights to address differences in underlying characteristics of opposite-sex and same-sex couples. When the gap in pre-birth earnings is 10% or less, both members of the couple are designated as equal earners; when the gap in pre-birth earnings is larger than 10%, earners are designated as primary if they have higher earnings than their partner and secondary if they have lower. Individuals in the analysis sample are required to have nonzero earnings in at least one quarter in the two years before the child enters the household. Source: ACS, Numident, and IRS data.

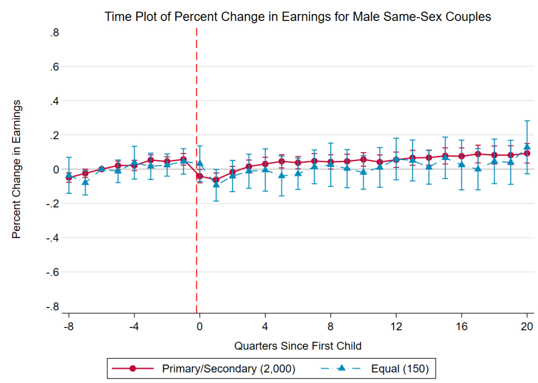
Figure 2.3: Household Earnings after Entry of First Child



(a) Individuals in Opposite-Sex Couples



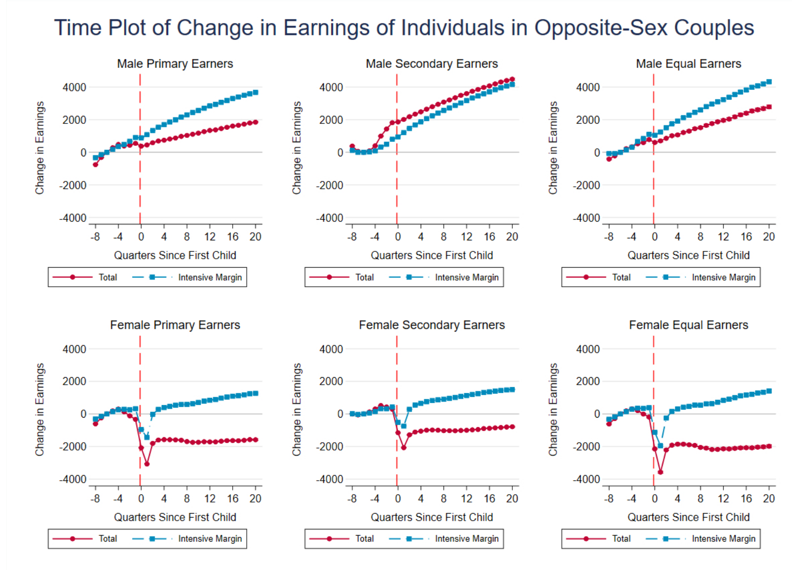
(b) Individuals in Female Same-Sex Couples



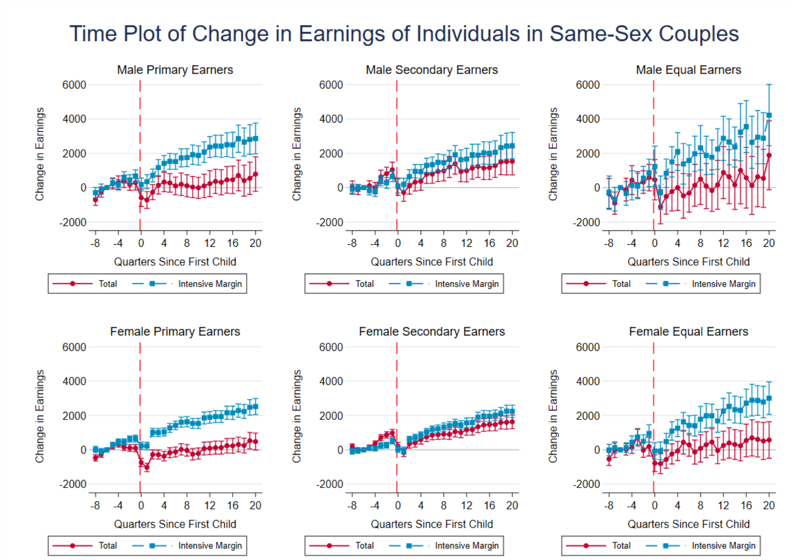
(c) Individuals in Male Same-Sex Couples

Note: This figure presents the household-level event studies of earnings around child entry. For easier interpretation, earnings changes are presented as percent changes following Kleven et al. (2019). The results are weighted with propensity weights to address differences in underlying characteristics of opposite-sex and same-sex couples. When the gap in pre-birth earnings is 10% or less, both members of the couple are designated as equal earners; when the gap in pre-birth earnings is larger than 10%, earners are designated as primary if they have higher earnings than their partner and secondary if they have lower. This analysis sample is slightly less restrictive than the individual-level analysis sample because we include individuals with no pre-birth earnings as long as their household has nonzero pre-birth earnings in at least one quarter. Source: ACS, Numident, and IRS data.

Figure 2.4: Individual Earnings after Entry of First Child, Intensive Margin



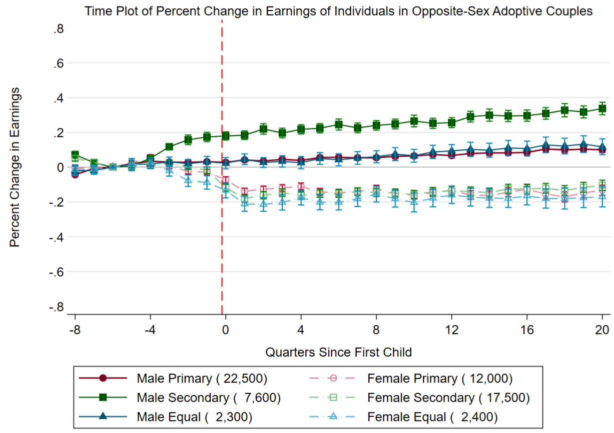
(a) Individuals in Opposite-Sex Couples



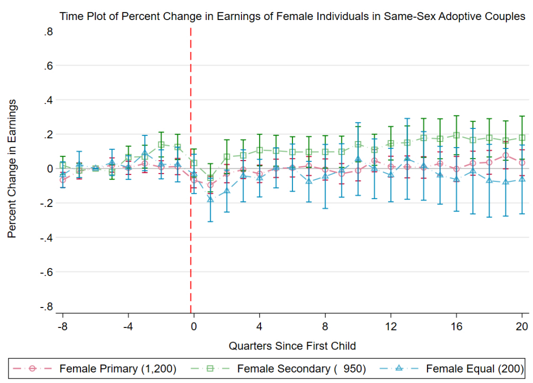
(b) Individuals in Same-Sex Couples

Note: This figure presents a decomposition of the event studies of earnings around child entry at the individual level into the intensive and extensive margins. The results are weighted with propensity weights to address differences in underlying characteristics of opposite-sex and same-sex couples. When the gap in pre-birth earnings is 10% or less, both members of the couple are designated as equal earners; when the gap in pre-birth earnings is larger than 10%, earners are designated as primary if they have higher earnings than their partner and secondary if they have lower. Individuals in the analysis sample are required to have nonzero earnings in at least one quarter in the two years before the child enters the household. Source: ACS, Numident, and IRS data.

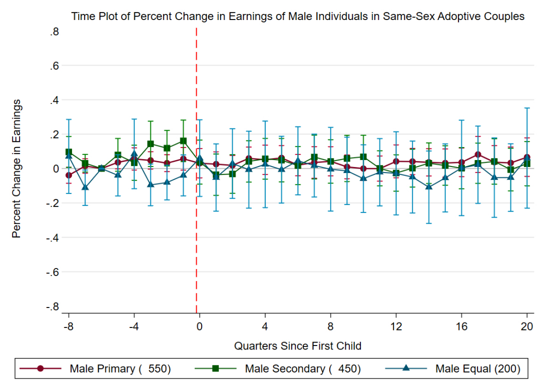
Figure 2.5: Individual Earnings after Entry of First Child, Among Adoptive Parents



(a) Individuals in Opposite-Sex Couples



(b) Individuals in Female Same-Sex Couples



(c) Individuals in Male Same-Sex Couples

Note: This figure presents the main event studies of earnings around child entry at the individual level, but limiting the sample only to parents whose first child was adopted. For easier interpretation, earnings changes are presented as percent changes following Kleven et al. (2019). The results are weighted with propensity weights to address differences in underlying characteristics of opposite-sex and same-sex couples. When the gap in pre-birth earnings is 10% or less, both members of the couple are designated as equal earners; when the gap in pre-birth earnings is larger than 10%, earners are designated as primary if they have higher earnings than their partner and secondary if they have lower. Individuals in the analysis sample are required to have nonzero earnings in at least one quarter in the two years before the child enters the household. Source: ACS, Numident, and IRS data.

Chapter 3: Getting Their ACT Together? The Impact of Mandatory College Entrance Exams on Postsecondary Achievement

3.1 Introduction

Inequality in educational achievement has been widening along several dimensions in recent decades. The achievement gap between children from high and low income families was about 40% higher for children who would be in the high school graduating cohort of 2019 than it was for children born two decades earlier (Reardon, 2011), and gaps in secondary school completion and college participation remain especially high for many minority groups (Espinosa et al., 2019). Several existing policies hope to address that education gap by encouraging disadvantaged students into college. Programs like Head Start and Project STAR increase the likelihood of attending college, but at a high financial cost – tens of thousands of dollars to induce one more students to attend college (Deming, 2009; Dynarski et al., 2013). Instead, recent research has focused on comparatively low-cost interventions to reduce informational and administrative barriers to college. Results from those interventions suggest that decreasing various barriers in the application process increases the number of low-income students who apply to and enroll in colleges (Bettinger et al., 2012; Hoxby and Turner, 2015).

In this paper, I evaluate a group of state-level reforms that made it easier for high school students to take the two main college entrance exams in the United States. Although the timing of the policy implementation was predominantly driven by state responses to federal standardized testing mandates, states have also touted this policy as a way to help disadvantaged students apply to, enroll in, and attend college. The policy represents a relatively inexpensive way to both increase students' information about their college readiness and reduce the time and financial cost associated with the college application process. However, improved access to college entrance exams would not be enough to increase college attendance if students are not prepared to take the test or remain financially constrained by other requirements of the college admissions process. It is not clear from previous work how large of a barrier the exams present compared to other portions of the application process. The evidence I provide in this paper suggests that lifting the financial and time barriers induced by the exam does not sufficiently reduce the application cost for the marginal student.

I study the period from 2001 to 2017, during which 30 states implemented versions of compulsory exam policies which required schools to offer either the ACT or SAT reforms for free during the school day. Even as more university systems move away from standardized test scores as part of their application package, states continue to institute these policies, with four more implemented since 2019.[1] To identify the causal effect of increased exam access on college attendance, I compare students in graduation cohorts and states with and without the mandatory exam policy. I characterize individuals' graduation cohort and educational achievement using American Community Surveys (ACS) from the Census Bureau, and make use of the Census Bureau's Master Address File (MAF) to precisely identify in which state each individual lived

[1] My analysis ends with policies implemented in 2017, which would have affected the cohort graduating in 2018, in order to have data on their post-high school outcomes.

when they took the exam as a junior in high school.[2]

My evaluation of the state exam policies finds no effect of increased exam access on college attendance. Using a full set of individual and state-time varying controls, I estimate an increase in college attendance of 0.1 percentage points with a standard error of 0.3 percentage points. When I focus on disadvantaged students (proxied by mother's education), who are more likely to be on the margin of applying to college if barriers were reduced, I find that this group was not differentially affected by the policy. In addition, my analysis is robust to several definitions of college attendance and control states, including recent approaches to addressing bias in estimation due to variation in the timing of treatment (especially Borusyak et al. (2024)). The lack of an effect is not explained by demographic characteristics of the students or state-level differences in access to guidance counselors and test-taking aid.

My main results pertain to the effect of the policy on college attendance. To explore other possible margins of response, I consider whether access to the exam induces students to adjust the type of college they attend. I find no effect of the policy on the choice between private and public schools or the choice between two-year and four-year schools. Similarly, along various dimensions of heterogeneity I estimate no effect of the policy: the policy does not behave differently among ACT and SAT states, among early- and late-adopting states, or among students of different sexes, race, or ethnicities.

I also evaluate the cost efficiency of the policy, pretending as if the point estimate were statistically significant, in order to provide a comparison with other programs aimed at addressing the education gap. Compared to high-cost, time-intensive programs such as Head Start, provision

[2] Due to the timing of the Census Bureau disclosure process, the results presented in this version of the paper do not use the MAF to correct for the high school state, but results with the correction look very similar.

of the ACT or SAT is more cost efficient despite the small size of the point estimate. However, other low-cost programs have had much larger (and statistically distinguishable from zero) positive effects on both college attendance and the competitiveness of the schools in which students enroll. Those programs are much more cost efficient than the ACT/SAT policy I evaluate here. Overall, my analysis suggests that the mandatory exam policy on its own is not an effective option to improve college attendance among low-income students who would otherwise not attend.

The effectiveness of the policy requires that there exist a group of students who would attend college if taking the SAT/ACT were less costly. This group of students would have to both be "college-ready" in terms of their academic performance and aptitude, and would have to be uninhibited by information barriers and other financial costs of the college application process. Several studies have provided evidence supporting the first point. Hyman (2017) and Cook and Turner (2019) both predict the distribution of test scores among non-takers to show that offering the test to all students would increase the number of college-ready test-takers by about 20%. Further, Hyman estimates that among poor college-ready students, about a third do not take the admission test. Cook and Turner, studying a state which has no standardized testing policy (Virginia), use their predictions to consider several counterfactual policies. They find that by offering the exam for free to all students in the state, about 24 % of non-takers would have scores above 1000, which they define as "college-ready." They consider more targeted policies and find that targeting students with high enough end-of-course test scores or districts with low test-taking would capture more than half of those high-achieving non-takers. This evidence is in line with other studies that show many high-achieving disadvantaged students do not attend college or attend a less selective school than they could (Bowen et al., 2009; Dillon and Smith, 2017; Hoxby and Turner, 2015; Hoxby and Avery, 2013; Pallais and Turner, 2006).

Whether these high-achieving non-takers would attend college if they were required to take a standardized test is less clear. Several studies evaluating policies in early-adopting states have provided mixed evidence. Klasik (2013) compares aggregate trends in college enrollment in Illinois, Colorado, and Maine to enrollment trends in matched control states. His estimates range from a decrease of seven percentage points to an increase of eighteen percentage points. Goodman (2016) and Hurwitz et al. (2015) also compare treated states to matched control states (Goodman evaluates Illinois and Colorado; Hurwitz et al. evaluate Maine). Using aggregate data, Goodman finds no effect on overall college enrollment. Using student-level data, Hurwitz et al. find an increase in enrollment of two to three percentage points, but no effect on college completion. However, they cannot disentangle the effect of offering the SAT for free during the school day from other simultaneous policy changes in Maine, including offering online prep courses, introducing the PSAT the year prior, and instituting a laptop initiative (Bulman, 2015).

Instead of comparing to matched control states, Hyman (2017) evaluates Michigan's exam policy by comparing students at schools that were and were not test centers before the policy. He finds that students without an in-school test center before the policy were 0.6 percentage points more likely to enroll in college, but that estimate is marginally significant and not clearly comparable with the other research designs, since his design differences out the overall effect of the policy that is general to all students. Bulman (2015) finds the most positive evidence for these policies. He evaluates three district-level policies in Florida, Texas, and California, finding that enrollment increased by about three percentage points, which represents ten percent of new test-takers. It is not clear how his results compare to the effect of a state-wide policy, however, since the state-level policies likely have larger general equilibrium effects. For example, states could respond to the policy by changing the supply of spots at their public universities, whereas districts do not have that option. In contrast, in this paper I present analysis evaluating policies across a large number of states and incorporating administrative data to identify individual-level variation in college outcomes. My results are most in line with the findings of Goodman (2016), but also with Hurwitz et al. (2015) and Hyman (2017), who identify only small changes in college enrollment and none in college completion. By combining information from multiple states' policy changes, I can explore whether different aspects of the policies contribute to their effectiveness, finding robustly that across many characteristics, these policies did not promote college attendance.

3.2 College Entrance Exams in the United States

In 2001, the No Child Left Behind Act expanded state-mandated standardized testing.[3] Both the law and its 2015 successor, the Every Student Succeeds Act, instituted the requirement that states provide standardized assessments in reading/language arts and mathematics at least once during high school (Larson 2002; ESSA 2016). These laws do not identify a specific exam but instead put forth criteria that any qualifying exam must fulfill. Nonetheless, states have turned to college entrance exams – in particular, the American College Test (ACT) and the Scholastic Achievement Test (SAT) – in order to meet federal testing requirements. Students generally take these exams for the first time in their junior year of high school.

Colorado and Illinois were the first states to implement mandatory ACT testing in 2001. By the end of 2019, 32 states had experimented with mandatory college entrance exams (two states, Minnesota and Missouri, ended their policies after a couple years; see Table C.1 in the Appendix

[3] Title I of the Elementary and Secondary Education Act is commonly referred to as the No Child Left Behind Act

for a breakdown of all state policies). Figure 3.2 depicts the policy variation over time. In total as of the 2019-2020 school year, 24 states implemented an ACT policy and 12 implemented an SAT policy. Four of those 24 ACT states later switched to the SAT. While the switch is important to take into account when examining test taking behavior and test scores, college attendance should not be affected by switching from one test to another since the exams are approximately equivalent in terms of meeting college entrance criteria. All colleges in the United States accept both the ACT and the SAT (Moody, 2020). Further, Lenning and Maxey (1973) found that the ACT and SAT do not differ in predicting first semester college grades. Although the exams do cater to different strengths, the evidence suggests that the choice between the two exams does not determine whether students are accepted by at least one of the colleges to which they apply.

Since the policies were primarily instituted due to state bureaucratic choices about which test best fulfills the federal mandate for the lowest cost, the adoption of the policy is likely not associated with state college enrollment counts. The main motive for implementing these exams was simply to satisfy federal standards. A positive externality often emphasized by the press has been to help disadvantaged students apply to, enroll in, and attend college. In the 2019-2020 school year, taking the ACT or SAT cost $52 to $68, a nontrivial amount for low income students.[4] However, the analysis in this paper suggests those claims of increases in college attendance are unsubstantiated.

The standardized testing policies are characterized by three components: eliminating registration fees, moving the testing dates from Saturdays to during the school day, and switching registration practices from opt-in to opt-out. The impact of the policies depends on the number of students for whom the total cost of applying is larger than their willingness to pay when the

[4] The test fee depends on the inclusion of the optional writing section.

exam is offered on the weekend for a price but smaller when the exam is administered in school for free. The willingness to pay and the monetary and time costs of the exam relative to the total application cost are key parameters. The theoretical prediction of the policy impact depends on whether and by how much the mandatory exam policies decrease the financial and informational barriers faced by marginal students.

Some evidence suggests those financial and informational barriers could be large. Most college applications request a test score from one of those two exams, and students from low socioeconomic backgrounds are much less likely to take either exam. In their sample, Avery and Kane (2004) find that only 32% of minority urban students, compared to 98% of wealthy suburban students, took a college entrance exam despite professing similar college aspirations. If students from lower income backgrounds are more constrained by the financial, time, and informational requirements of college applications, then removing those constraints could increase their access to college by raising their likelihood to apply to and match with schools. However, there are many steps in the decision to apply for and attend college, and taking the entrance exam is only one step along the way. The evidence from this paper suggests that improved access to college entrance exams is not sufficient to increase college attendance.

3.3 Data

3.3.1 ACT and SAT Test Scores and Take-Up

The number of ACT takers increased dramatically during the 2000-2019 period, as Figure C.1 shows. The ACT gained popularity partially by marketing to states looking to fulfill the federal requirement for high school standardized testing. The SAT recently implemented several changes in an attempt to reclaim its status as the primary US standardized test; this is apparent in the sharp increase in SAT takers around 2018. SAT and ACT scores and counts of test takers come from a multitude of sources, detailed briefly here.

Counts of test takers and scores for the SAT come from the College Bound Seniors State Profile Reports published yearly by the College Board.[5] The reports are published for each state-year and contain aggregated measures for the state as a whole. Average ACT scores by state come from the annual ACT Condition of College and Career Readiness reports. Counts of test takers from 2010 onward come from the ACT data visualization tool available on their website.[6] Earlier years' information comes from the annual High School Profile Report: HS Graduating Class State Composites, which provide average test scores and total number of test takers for the 5 previous years.[7] Enrollment counts come from the National Center for Education Statistics (NCES) Table 219.20. Graduates are projected beginning in the 2013-2014 year, which introduces some measurement error into the proportion of test takers in each state-year.[8] I have no reason to believe that the measurement error would be differential across treatment status.

3.3.2 College Outcomes

I use American Community Survey (ACS) files from 2000 to 2019 to estimate the causal impact of the mandatory exam policies on college attendance.[9] The main outcome of interest is

[5] These reports changed name to the Suite of Assessments Annual Report in 2017; https://www.collegeboard.org/

[6] https://www.act.org/content/act/en/research/services-and-resources/data-and-visualization/grad-class-database.html

[7] https://www.act.org/content/act/en/college-and-career-readiness.html

[8] Source: U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), "State Nonfiscal Survey of Public Elementary/Secondary Education," 1981-82 through 2005-06; "State Dropout and Completion Data File," 2005-06 through 2012-13; Public School Graduates and Dropouts from the Common Core of Data, 2007-08 and 2008-09; and State High School Graduates Projection Model, 1980-81 through 2026-27.

[9] Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek. IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/

college attendance one year or later after graduating high school. Measuring college attendance instead based on two years beyond high school yields similar results. As an alternative outcome of interest, I define two year college as the education level being an associate's degree. I differentiate this from four year college, which I define as having some college two years after the cohort year.

I characterize individuals based on the year they would have graduated high school if they progressed through the education system at the standard rate, which I call their cohort year.[10] These graduating class cohorts span 1990 to 2018, which provides ten years of pre-period data for the earliest policy states. I restrict the sample to individuals aged 17 or older who have completed or left high school, so that their ACS survey responses capture any college attainment. I refer to all school years based on the spring semester year (so that the 2012-2013 school year becomes the 2013 year). Because I characterize individuals based on the year they would have been seniors in high school rather than their survey year, pooling the ACS across years serves simply to increase my sample size. The large sample size contributes to the tight standard errors I estimate around the null effect. To account for potentially differential effects based on the year individuals were interviewed, I include survey year fixed effects and survey year interacted with cohort year in my preferred specification. The final estimation sample is 11,488,252 individuals who were seniors between 1990 and 2018 for whom I have information about their college attendance.

I also characterize students by their state. It is not uncommon for individuals to attend college in a different state from where they attended high school. Among 19 year olds in my sample, 31% moved between states in the last year. I adjust for migration using the Census Bureau's Master Address File (MAF).[11] The MAF combines information from several adminis-

D010.V10.0

[10] I add 18 to their birth year or 19 for those born in the 4th quarter of the year. Adding 17 and 18 or 19 and 20, respectively, results in similar estimates.

[11] As mentioned earlier, due to the timing of the Census Bureau disclosure process, the results presented in this

trative data sources to characterize where all individuals in the United States live each year. It may identify several possible addresses for one individual, but ranks those by likelihood of being accurate. I choose the most likely address for each individual-year and match the MAF to the ACS using Personal Identification Keys. Each individual is assigned the MAF state that was measured in the year closest to the individual's cohort year, and is assigned their ACS state of birth for the small number of cases who do not match to the MAF. As a robustness check, I limit my sample to those who were interviewed one year after high school and use the ACS migration variable which gives the state in which the individual lived the previous year. The limited sample should have the most accurate information about the state in which students took the exam.

Table 3.1 presents descriptive statistics and differences in means for treatment and control states, where control states are those which never enacted a policy and treatment states are those which enacted a policy at some point from 2001 to 2017. There are several states which have enacted a policy since 2017, but I do not have enough post-period data to analyze those states. The groups are well-balanced in most criteria. Treatment states are less white and have lower family incomes, but have more citizens and speak better English. The existence of these differences on levels is not necessarily a threat to the identification strategy, which relies on similarities in trends over time between treatment and control groups. For more power and to create more similar comparison groups, I control for demographic characteristics in my preferred model specification. Those demographic characteristics include income measures, which I convert to 2018 dollars when applicable using the CPI-U. I also control for state time-varying characteristics, where time is relative to the year when the student graduated high school. Total employment,

version of the paper do not use the MAF to correct for the high school state, but results with the correction look very similar.

GDP, employment-population ratio, and unemployment rate come from the Bureau of Economic Analysis (BEA).[12] State guidance counselor ratios come from the National Center of Education Statistics (NCES).[13]

3.4 Empirical Methodology

3.4.1 Difference-in-Differences Specification

The policy mandating that college entrance exams be administered free during the school day occurred at the state level in different years for different states. Figure C.2 depicts the variation in treatment timing of the policy. The treatment group changes by state and year starting in 2001. Taking differences appropriately across the two dimensions of variation should identify the true effect of interest: the average impact of access to college entrance exams on subsequent college attendance.

The key identifying assumption is that, absent the exam policy, trends in college attendance would be the same for treated and control states. That assumption encompasses two implications. First, there should be no other policy enacted at the same time which would impact college attendance differently for treatment states. If another policy aimed to improve exam performance or college attendance in the same states implementing ACT or SAT testing, my estimates would be biased upward, making it harder for me to detect no effect. Second, the trend in college attendance for the group of states who chose to implement mandatory exams should not differ from that of the states who chose not to implement mandatory exams. I provide evidence in section 3.4.3 that those trends are parallel in the pre-policy period.

[12] https://apps.bea.gov/regional/downloadzip.cfm
[13] https://nces.ed.gov/ccd/elsi/

I compare individuals with and without exam access due to policy implementation that differs by their graduating class cohort and high school state. State and cohort fixed effects isolate my source of identifying variation. I implement these difference-in-differences comparisons in a regression framework by estimating Equation 3.1 below using ordinary least squares.

The baseline model is

$$\texttt{attendance}_i = \alpha + \delta\texttt{access}_{sc} + \eta_s + \lambda_c + \{\texttt{X}_i\beta + \texttt{S}_{sc}\zeta + \gamma_t\} + \varepsilon_{isc} \tag{3.1}$$

where i references individual, c references cohort, t references interview year, and s references state. η_s is the state fixed effect and λ_c the cohort fixed effect. ε_{isc} is an idiosyncratic error term. I allow for arbitrary correlation among individuals within a state by clustering standard errors at the state level. All models are estimated using ACS individual weights. The parameter of interest is δ, the effect on college attendance of being in a graduating class cohort with full access to an exam compared to being in a cohort without such access.

In further specifications I include demographic characteristics that vary by individual (X_i), school and business-cycle characteristics that vary by state and cohort year (S_{sc}), and interview year fixed effects (γ_t) to address the different ages at which students are later interviewed in the ACS.

The individual demographic characteristics include family income, marital status, sex, race, ethnicity, poverty level, a categorical variable for degree of English speaking, an indicator for food stamp receipt, number of children, mother's education, and citizenship status. The state characteristics which vary by cohort year are the gross domestic product (GDP) and employment-to-population ratio, which control for differential effects of the macroeconomy on decisions to attend college. I also include each state's ratio of students to guidance counselors in consideration that students who receive more or higher quality college counseling might be better able to take advantage of access to the free exam.

For average treatment effects by year, I modify Equation 3.1 slightly. I replace the indicator for having access to the policy with dummy variables for policy access in each cohort relative to the policy implementation.

$$\texttt{attendance}_i = \alpha + \sum_{k=-10}^{k=10} \delta_k \texttt{access}_{sk} + \eta_s + \{\texttt{X}_i\beta + \texttt{S}_{sc}\zeta + \gamma_t\} + \varepsilon_{isc} \tag{3.2}$$

I leave out the year before the policy was implemented so that the average treatment effect I estimate is the change in college attendance relative to the year prior to the policy.

3.4.2 Heterogeneity: Triple Difference

I estimate the above models using all ACS individuals who are out of high school. However, I am interested in the marginal student who is moved into attending college by the policy. The mechanism I have in mind is that students who would otherwise be constrained financially are able to take the exam because it is free and offered at a convenient time and location. To focus on the group for whom mandating ACT or SAT testing most likely relaxed a binding budget constraint, I first estimate Equation 3.1 on three different high impact sample defined by mother's education being below college, family income being in the first quartile of that state-year income distribution, and Supplemental Nutrition Assistance Program (SNAP) participation. To further explore the interaction between the policy introduction and the likelihood of impact, I implement a triple difference-in-differences approach. I interact exam access with a measure of the likelihood of being impacted by the policy in Equation 3.3.

$$
\begin{aligned}
\texttt{attendance}_i = \alpha &+ \theta_1 \texttt{access}_{sc} + \theta_2 \texttt{likely}_i + \delta \texttt{access}_{sc} * \texttt{likely}_i \\
&+ \eta_s + \lambda_c + \phi_1 \texttt{likely}_i * \eta_s + \phi_2 \texttt{likely}_i * \lambda_c \\
&+ \{\texttt{X}_i \beta + \texttt{S}_{sc} \zeta + \gamma_t\} + \varepsilon_{isc}
\end{aligned} \tag{3.3}
$$

where `likely` is an indicator for being more likely impacted by the policy. I proxy likelihood in the same three ways using mother's education below college, family income in the first quartile of the state-year income distribution, and food stamp receipt. The coefficient of interest δ now represents the change in the probability of attending college for a high-likelihood individual who has full access to the exams compared to a low-likelihood individual who has full access to the exam. To verify the specification of these groups, I expect the coefficient θ_1 on `access` to be indistinguishable from zero because it represents the effect of exam access on college attendance for those who are least likely to be moved by the policy.

3.4.3 Identification

While being unbalanced on levels would not undermine the identification strategy, being unbalanced on trends would bias the comparison between states over time. I examine trends in both test taking and college attendance, plotting separately for each policy-year treatment group. The trends are parallel and describe an increase over time in both outcomes for both groups.

Control and treatment groups show similar college attendance trends. 3.3 shows raw trends for policy-year groups 2001 and 2015, depicting a long series of parallel pre-trends (2015) and post-trends (2001). Attendance has increased similarly over time for both treatment and control groups. These parallel trends also hold when looking at the first stage, test taking. Figure C.3 depicts the trends in the number of test takers for two different policy years. There is a general upward trend in test taking which is similar for treatment and control groups. Other than a stark jump exactly at the policy year, the trends are parallel.

Another concern might be that states differ in the type of test they implement. The difference would only bias my estimates to the extent that characteristics of states which are correlated with college attendance, exam scores, and test taking are also correlated with states' choice of which test to implement. Nevertheless, Table C.2 in the Appendix presents results from an OLS regression of the type of test implemented on state characteristics. Average poverty rate, average race, average food stamp receipt, proportion of women, proportion married, and proportion with a college degree do not have a statistically significant impact on the implementation of the policy. Conditional on the policy being implemented, these characteristics also do not explain the choice between ACT and SAT.

In summary, the difference-in-differences identification strategy hinges on exogenous adoption of the policy by the states. The condition is likely to hold in this setting not only because the motive for adoption was to adhere to federal standards, but also because, as shown, observable state characteristics do not predict treatment status. In addition, to further address the exogeneity of the policy adoption over time, I control for specific time-varying state trends in macroeconomic conditions and guidance counseling access. The results from analysis with and without those control are very similar.

3.4.4 Recent Advances in Difference-in-Differences

The recent literature on dynamic difference-in-difference models with staggered adoption of treatment has highlighted the importance of properly considering the heterogeneity introduced by the variation in policy adoption over time (see, for example, Borusyak et al. (2024); Callaway and Sant'Anna (2021); De Chaisemartin and d'Haultfoeuille (2020); Goodman-Bacon (2021); Sun and Abraham (2021); Wooldridge (2021)). Since the SAT and ACT policies were implemented at different times in different states, I carry out several alternative specifications to verify that my results are not biased by "forbidden comparisons" between newly-treated and earlier-treated groups. As my main robustness check, I apply the Borusyak et al. (2024) estimator. Their approach is intuitively imputation-based, using untreated observations to impute untreated potential outcomes for the treated observations. The results presented in this paper are robust to the Borusyak et al. (2024) approach. They are also robust to an approach in which I stack cohorts of treated and control units and estimate a standard difference-in-differences model on the stacked dataset, controlling for cohort fixed effects.

3.5 Results

3.5.1 Testing

I first verify that offering the exams for free during the school day to all juniors at public high schools does indeed result in significantly more students taking the ACT or SAT in that state. Figure 3.4 plots the number of test takers as a ratio of graduates in ACT policy states against the number of years since policy implementation. After the policy begins, there is an increase of about 70 percentage points in the ratio of test takers to graduates. Test taking trends in SAT states look very similar, as shown in Figure C.4. Table 3.2 reports results from an OLS regression of the combined number of ACT and SAT takers in a state on state policy status. Across all specifications, the estimated coefficient is statistically significant at the 1% level and positive, indicating that implementing an ACT policy is correlated with an increase in test taking of about

0.1 standard deviations (for the SAT policy, an increase of 0.15 standard deviations). Notice that the outcome is the total number of test takers. The mandatory exam policies are not just shifting test takers from one exam to another, but also encouraging new students to take an exam.

Next I investigate how the policy affected the average test scores in these states. I expect that mandatory testing will capture many students who have no interest in taking the exam and either are not intellectually prepared for the material or have little motivation to put in effort for an exam from which they see no benefit. Expanding testing to all students captures much more of the left tail of the test score distribution than it does of the right tail. Figure 3.4 shows that indeed, the average test scores in ACT states plummet after the introduction of the policy. Test scores in SAT states respond similarly, as shown in Figure C.4.

3.5.2 College Attendance

Table 3.3 presents the results from the main model. The outcome variable is a binary choice to attend college one or more years after graduating high school. Although the initial regression uses ordinary least squares, I find very similar marginal effects from probit (Column 5) and logit (Column 6) models. I begin with only state and cohort fixed effects to isolate the variation within state between graduating class cohorts which is my source of identification. In subsequent models I add individual-level demographic controls (family income adjusted to 2018 dollars using the CPI-U, marital status, sex, race, ethnicity, poverty, number of children, English speaking, citizenship, and mother's education), interview year fixed effects, time-varying state characteristics (employment, GDP, and student-counselor ratio), and cohort-year interactions. Cohort-year interactions control for differential interview year effects by age.

All of the estimates of the effect of exam access on subsequent college enrollment are statistically insignificant and near zero in magnitude. My preferred specification with the full set of controls reports an estimated change in the probability of attending college by 0.1 percentage points for those who had full access to the ACT or SAT, with standard error 0.3 percentage points.

I check the robustness of the lack of an effect to alternative definitions in Table C.3 in the Appendix. First, I test whether the estimates could be biased by control states that were implementing similar policies at the district level. As Table C.1 describes, several states with no state-level policy nevertheless began offering a free ACT or SAT in certain districts. For example, New York has no state policy, but New York City has implemented free exams in some districts since 2016. In Column 1, I remove partially treated states and find no change in my results – if anything, the magnitude of the (statistically insignificant) effect is closer to zero. Next, in Column 2, I limit the sample to those likely impacted, as proxied by mother's education below college and family income in the first quartile of the state's income distribution. In Column 3, I limit the sample to individuals interviewed exactly one year after their graduating class cohort year, which allows me to define their high school state with the greatest accuracy. In Column 4, I use state of birth instead of high school state to define where each individual would have taken the exam. No estimate is significant and all are close to 0 in magnitude with standard errors around 0.4 percentage points. I find similar results when I define college attendance two years after senior year (Column 5) and when I define college attendance using variables for being currently in school in addition to education level (Column 6).

Figure 3.5 depicts the average treatment effect by year relative to the start of the policy. Around the first year of exposure to the policy (year zero defines the first year that the exams were administered during the school day in that state), there is absolutely no change in college attendance. Although the estimates become noisier several years in the future, there continues to be no statistically significant change in college attendance due to the policy. These estimates are normalized such that each year is compared to the year immediately before policy introduction. The confidence intervals rule out any effect larger than about 2 percentage points.

In order to understand these results, I calculate the minimum effect size I have the power to detect. With 51 clusters, at the standard 95% significance level and power of 80%, I can detect a minimum effect of 0.0168. That effect size corresponds to college entrance exam access increasing the probability of college attendance by about 1.7 percentage points. For comparison, Bettinger et al. (2012) find that providing support in filing federal financial aid forms to low-income students increased their likelihood of attending college by 8 percentage points. Given that studies of similar policies have found effects ranging from 2 to 10 percentage points, were there an effect of a similar magnitude, I would have detected it.

3.5.3 Likely Impacted Students

Table 3.4 presents the results from the triple difference-in-differences model. I use three different proxies for the group of students that was most likely impacted. I look at those whose mother's education is a high school degree or below, those whose family income is in the first quartile of the income distribution, and those who participated in the federal food stamp program, SNAP. I prefer the first proxy, mother's education, because that variable is more reliably time-constant. The variables are measured in ACS interview which occurs after the student graduates high school, sometimes by many years. SNAP use later in life is therefore a rough proxy for SNAP use during high school which is a proxy for being likely impacted by the freely available exams. As expected, growing up in a family that is less educated, lower income, or using food stamps decreases the likelihood of attending college by 6 to 17 percentage points. This result is highly statistically significant. For example, having a mother with less than a college education means the student is 15.5 percentage points less likely to attend college. However, all of the coefficients on the interaction between policy and impact group are near zero in magnitude, ranging from 0 to -0.4 percentage points, and statistically insignificant. A null effect here means that the students most likely to be impacted do not respond to the policy differently than those less likely to be impacted. The results of the triple difference suggest that even among those who were most likely to be moved by the policy, college attendance did not increase.

3.5.3.1 Type of College

Thus far I have explored the average treatment effect of offering mandatory exams on whether college attendance. However, the margin on which this policy operates might not be the decision to attend college, but instead the choice of where to attend. Students who were already going to attend some type of college might be induced to apply to higher quality schools. Perhaps the policy moves students to apply to more selective schools; Goodman (2016) found that selective enrollment rose by 20% following the mandates in early-adopting states. Although I do not have information on the selectivity of the college, I can explore whether students are attending four year colleges (those more likely to require the ACT or SAT) instead of attending two year colleges. I also investigate heterogeneity in whether the school is a public or private institution. Table 3.5 presents estimates of Equation 3.1 in which the outcome variable, college attendance, has been replaced by an indicator for a four- versus two- year college or an indicator for a public versus private college. These estimates are conditional on college attendance. My results do not provide evidence for any changes in the type of college that students attend. Estimates are all close to zero in magnitude and not statistically significant with standard errors around 0.6 percentage points. However, the mandatory exam policy may still be relevant for other dimensions of higher quality, as Goodman's analysis suggests.

3.5.4 Heterogeneity

In this section I explore various subsamples in which we might expect the policy to have different effects. First, I treat the two exams as separate policies. Next, I consider early- and late- adopting states separately. Finally, I allow for different effects across various student demographic characteristics. In all cases, I find no impact of the policy on future college attendance.

I first explore the two exams separately in order to allow for any potential differences in the preference of students for a certain exam, the reliance of colleges on a certain exam, or the implementation of the policy. I characterize states by whether their mandatory exam policy was related to the SAT or to the ACT. Another dimension in which the policies could differ is in the timing of adoption. In order to evaluate whether early-adopting states differ from late-adopting states in terms of the effectiveness of the policy, I split the sample into states that adopted a policy in or before 2012 and those that adopted a policy later than 2012. I also consider dimensions of individual variation. I explore whether the effect of mandatory exam policies on college attendance differed for male and female students and for students of different races and ethnicities.

Figure 3.6 presents the results from the heterogeneity analysis. Estimates of the effect of access to college entrance exams on college attendance range from -0.010 (confidence interval

[-0.041, 0.021]) to 0.004 (confidence interval [-0.012, 0.019]). This rules out effect sizes of a decrease larger than 4 percentage points and an increase larger than 2 percentage points. In addition, none of the effects are statistically distinguishable from each other. The evidence from this analysis suggests that the lack of effect in the overall sample is not masking the presence of an effect among certain groups of individuals or certain types of policies.

3.5.5 Cost-Benefit Analysis

In what follows, I calculate cost efficiency as the ratio of the cost per person over the change in attendance due to the policy. Lower costs and larger effects both result in a smaller cost efficiency estimate, which is desirable from a policy perspective. From Table 3.3 Column 7, I estimate the average treatment effect at 0.1 percentage points (standard error 0.3 percentage points). The registration fee for taking either the ACT or SAT is $50.95.[14] Assuming any savings from a state contract with the testing company offset the costs of administering the exam, that amounts to a cost efficiency of $50,950.

Intensive programs such as Head Start and Project STAR increase the likelihood of attending college, but at a high financial cost. Head Start, which provides preschool to low-income students, has a cost efficiency of $161,556 (Deming, 2009). The Project STAR experiment, which decreased class sizes, has a cost efficiency of $484,667 (Dynarski et al., 2013). In comparison to these programs, if I were confident that the administration of mandatory exams had even a 0.1 percentage point effect on college attendance, the relatively low cost would suggest that offering the exams is a beneficial option for policymakers seeking to reduce barriers to college.

On the other hand, similar less-cost-intensive policies are much more cost efficient and

[14] All dollar amounts in this section are given in 2018 dollars.

bring about a positive effect on college attendance. The H&R Block intervention to provide aid in filling out the federal financial aid application increased college attendance by 8 percentage points leading to a cost efficiency of $1,333 (Bettinger et al., 2012). The experiment implemented by Hoxby and Turner (2015) to mail customized application information and waivers led to a 12 percentage point increase in the number of colleges to which students were admitted and a 19 percentage point increase in the probability of enrolling at a peer institution. This corresponds to a cost efficiency of $53.92 and $34.05, respectively. Because those two programs are less costly per person and have larger (and statistically significant) average treatment effects, policymakers should consider such more focused interventions. My results have shown that offering the ACT or SAT for free in high schools is not a reliable policy alone in terms of efficiently moving students to attend college.

3.6 Conclusion

Capitalizing on variation in policy implementation across state and time, I find that increased access to college entrance exams causes no change in subsequent college attendance or the type of college chosen. I evaluate policies in 30 different states implemented between 2001 and 2017 which made college entrance exams more available to students, and conclude that the policies did not significantly increase college attendance. Even when looking at the groups most expected to be helped by free college entrance exams, I find no effect. These results indicate the potential benefit of shifting away from these entrance exam policies and toward other modest financial and informational interventions that have been more successful in improving college attendance.

Though previous studies looking at just one state's policy conclude with suggestive evidence of an increase in college applications and enrollment, one concern with their approach is that other policies may have been enacted simultaneously that would bias the results upward. For example, Arkansas enacted the No Limits test prep campaign a year after implementing their mandatory ACT policy. In addition, if more students apply, supply-side constraints may prevent them from actually attending college. In the presence of supply-side constraints, the general equilibrium effect could be null. The policy could instead be changing who goes to college rather than increasing the total amount of students attending, but I do not find strong evidence even for a shift in the makeup of college attenders. I should have seen an increase in attendance among low-income students who have stronger qualifications but are constrained by exam costs. I do not find this either by limiting my sample to that group or by implementing a triple difference using variation in likelihood of impact. The ACT and SAT policies seemed a promising intervention which were ultimately not successful on their own at nudging low-income students to attend college.

One other study has found null effects of a similar policy. Foote et al. (2015) use a regression discontinuity at the college-readiness ACT score and conclude that students who receive information that they are college-ready are not more likely to attend college. It is quite possible that students do not know the exact college-readiness benchmark scores well enough for this to have been a meaningful intervention. In a similar way, decreasing the barriers to taking exams has clearly shown to increase test taking, but this paper suggests that other constraints not addressed by the mandatory exam policies remain binding in the college application process. My findings suggest a role for future work in determining whether and which supply side factors or additional demand side constraints are at play.

3.7 Tables

Table 3.1: Treated and Control State Characteristics, Full Sample

Variable	Control	Treatment	Difference
Age	29.005	28.882	-0.124**
	(6.700)	(6.696)	(0.057)
Female	0.495	0.498	0.003
	(0.500)	(0.500)	(0.002)
Non-White	0.490	0.306	-0.184***
	(0.500)	(0.461)	(0.052)
Marital Status	0.376	0.395	0.020
	(0.484)	(0.489)	(0.013)
Number of Children	0.762	0.821	0.060*
	(1.153)	(1.175)	(0.033)
Mother's Years of Education	12.311	12.852	0.541*
	(3.654)	(2.880)	(0.271)
Citizenship Status	0.853	0.931	0.078***
	(0.354)	(0.253)	(0.020)
Speaks English	1.912	1.964	0.053***
	(0.352)	(0.221)	(0.015)
On SNAP	0.135	0.149	0.014
	(0.342)	(0.356)	(0.010)
% of Poverty Line	285.386	272.789	-12.596*
	(170.471)	(167.615)	(6.503)
Family Earnings	81,241.086	71,849.508	-9,391.584***
	(83,305.391)	(70,943.930)	(3,287.628)
Observations	6,839,705	4,648,547	11,488,252

Note: This table shows differences in means for demographic characteristics among treated and control states. Treatment states are less white, have more citizens, speak English better, are more impoverished, and have lower family incomes. The treatment group is all states that administered a mandatory college entrance exam at any point over the years 2001 to 2019; the control group is all states that never implemented a mandatory exam policy. Treatment and control standard deviations in parentheses; standard errors, clustered at the state level, reported in parentheses for the difference. Average characteristics computed from 17-47 year olds in the 2000-2018 ACS.

Table 3.2: Effect of Mandatory College Entrance Exams on Number of Test Takers

	(1) Test-Takers	(2) Test-Takers	(3) Test-Takers
Policy State	0.117***	0.141***	
	(0.014)	(0.013)	
ACT Policy			0.100***
			(0.012)
SAT Policy			0.146***
			(0.014)
Average Poverty Rate	–0.000	–0.000**	–0.000*
	(0.000)	(0.000)	(0.000)
Average Race	0.315***	0.359***	0.365***
	(0.049)	(0.047)	(0.046)
Average Food Stamps	–0.536***	0.357**	0.199
	(0.143)	(0.179)	(0.176)
Average Female	0.805***	0.413**	0.427**
	(0.203)	(0.198)	(0.194)
Average Married	–0.624***	–0.746***	–0.584***
	(0.075)	(0.138)	(0.141)
Average English Speakers	1.089**	1.474***	1.701***
	(0.501)	(0.489)	(0.479)
Average Education	–0.000	0.027***	0.023***
	(0.002)	(0.004)	(0.004)
Year Fixed Effects		X	X
Region	X	X	X
Observations	955	955	955
Adjusted R^2	0.344	0.418	0.440

Note: Results from a regression of the state-year proportion of test takers on policy status. There is a strong positive correlation between policy existence and amount of test takers. Test takers is defined as the sum of ACT and SAT test takers. The proportion of test takers is calculated as the number of testers divided by the number of graduates, where graduate counts come from the NCES and are projected after the 2012-2013 school year. Average state characteristics are calculated from 17 to 47 year olds from the 2000-2018 ACS. Standard errors, clustered at the state level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.3: Effect of Mandatory Testing Policies on College Attendance

	College						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Policy	–0.004	–0.006	–0.007	–0.008	–0.007	–0.007	0.001
	(0.008)	(0.008)	(0.008)	(0.006)	(0.005)	(0.005)	(0.003)
State Fixed Effects	X	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X	X
Demographic Controls		X	X	X	X	X	X
Year Fixed Effects			X	X	X	X	X
State-Year Controls				X	X	X	X
Cohort-Year Interactions							X
Model	LPM	LPM	LPM	LPM	Probit	Logit	LPM
Observations	11,438,734	10,553,014	10,553,014	6,321,533	6,321,533	6,321,533	6,321,533
Adjusted R^2	0.008	0.182	0.187	0.173			0.176

Note: Results from a two-way fixed effects regression of college attendance on policy status. Marginal effects are reported for probit and logit models. College attendance is defined as an indicator for whether an individual in cohort t attended college one year or more after graduating high school. Cohort is defined as the year the individual would have been a senior using standard grade ages. Unless noted, sample is 17 to 47 year olds from the 2000-2018 ACS. Standard errors, clustered at the state level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.4: Effects of Mandatory Testing Policies on College Attendance, by Likelihood of Impact

	College		
	(1)	(2)	(3)
Policy	–0.003	–0.009	–0.008
	(0.005)	(0.010)	(0.009)
Likely Impacted	–0.155***	–0.057***	–0.167***
	(0.044)	(0.004)	(0.005)
Policy x Likely Impacted	–0.004	–0.000	–0.001
	(0.005)	(0.006)	(0.006)
Proxy for Impact	Mother's Ed	Low Family Income	Food Stamps
State Fixed Effects	X	X	X
Cohort Fixed Effects	X	X	X
Controls	X	X	X
Year Fixed Effects	X	X	X
Observations	10,758,782	11,232,966	10,553,014
Adjusted R^2	0.189	0.186	0.187

Note: Results from a two-way fixed effects regression of college attendance on policy status interacted with a proxy for the likelihood of being impacted by the policy. This proxy is measured in three ways: by mother's education being high school degree or less, by family income being in the first quartile of that state's income distribution, and by the individual receiving food stamps. College attendance is defined as an indicator for whether an individual in cohort *t* attended college one year or more after graduating high school. Cohort is defined as the year the individual would have been a senior using standard grade ages. Sample is 17 to 47 year olds from the 2000-2018 ACS. Standard errors, clustered at the state level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

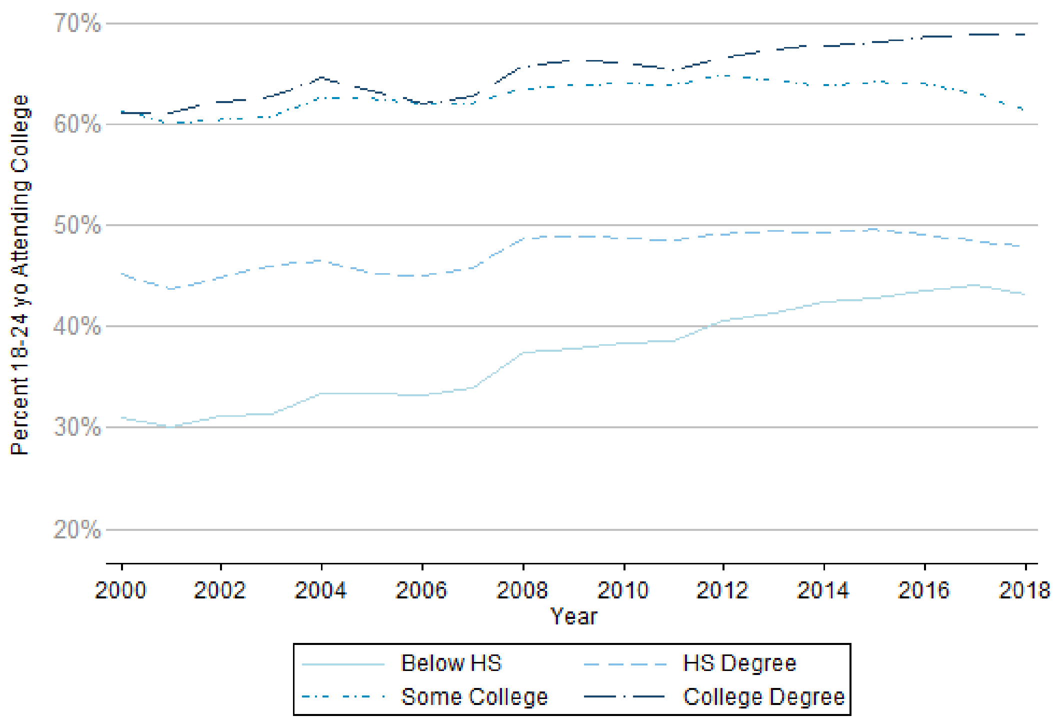
Table 3.5: Effects of Mandatory Testing Policies on Type of College

	Two or Four Year		Public or Private	
	(1)	(2)	(3)	(4)
Policy	0.002	0.002	−0.008	−0.003
	(0.006)	(0.004)	(0.008)	(0.004)
State Fixed Effects	X	X	X	X
Cohort Fixed Effects	X	X	X	X
Controls		X		X
Year Fixed Effects		X		X
Observations	6,319,674	3,460,522	2,003,981	1,347,110
Adjusted R^2	0.006	0.017	0.046	0.055

Note: Results from a two-way fixed effects regression of type of college attended on policy status. Type of college is differentiated in two ways: two-year versus four-year and public versus private. Cohort is defined as the year the individual would have been a senior using standard grade ages. Sample is 17 to 47 year olds who attended college from the 2000-2018 ACS. Standard errors, clustered at the state level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

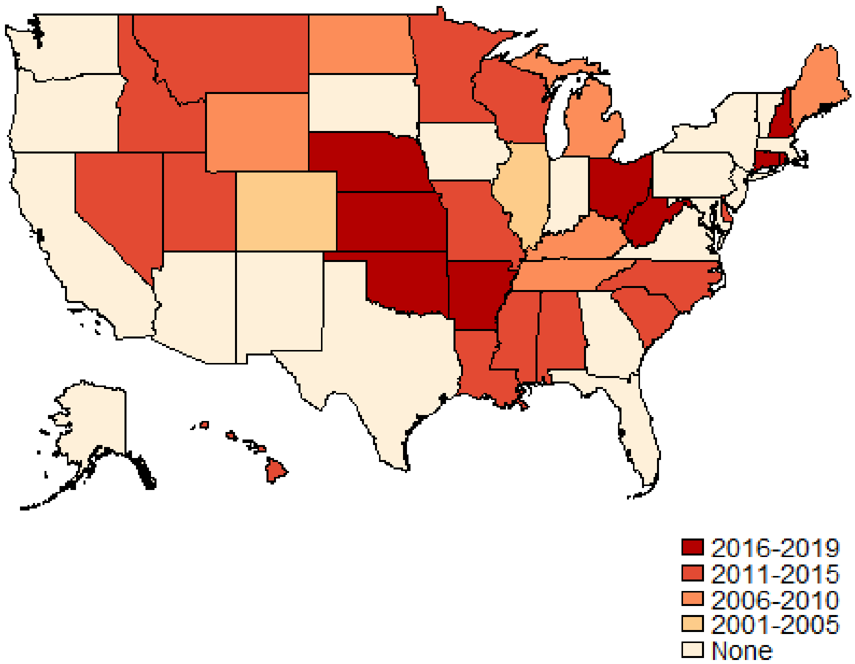
3.8 Figures

Figure 3.1: College Attendance by Mother's Education



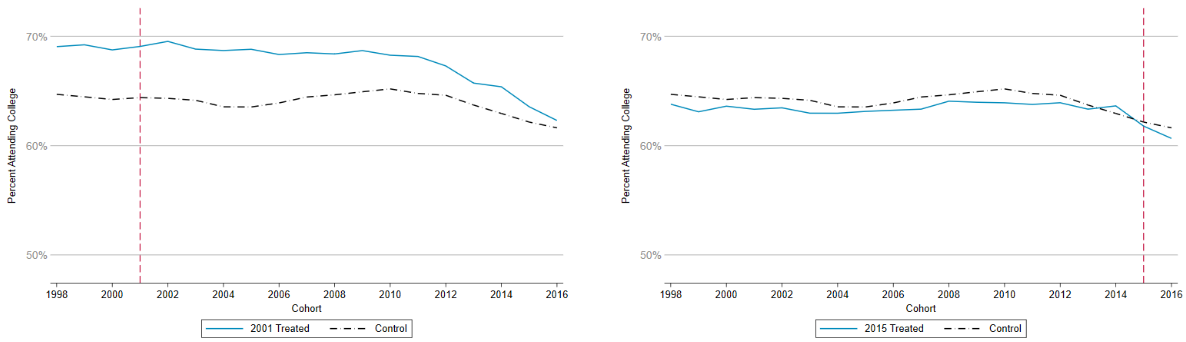
Note: This figure shows college attendance by mother's education. Attendance trends are similar, but levels are much higher for individuals whose mothers have more education. Mother's education is grouped into four categories corresponding to no high school degree, high school degree, some college, and college degree or more. College attendance is defined as an indicator for whether an individual in cohort *t* attended college one year or more after graduating high school. Cohort is defined as the year the individual would have been a senior using standard grade ages. Sample is cohorts 1990 to 2020 from the 2000-2018 ACS.

Figure 3.2: Geographic Variation in Year Policy Implemented



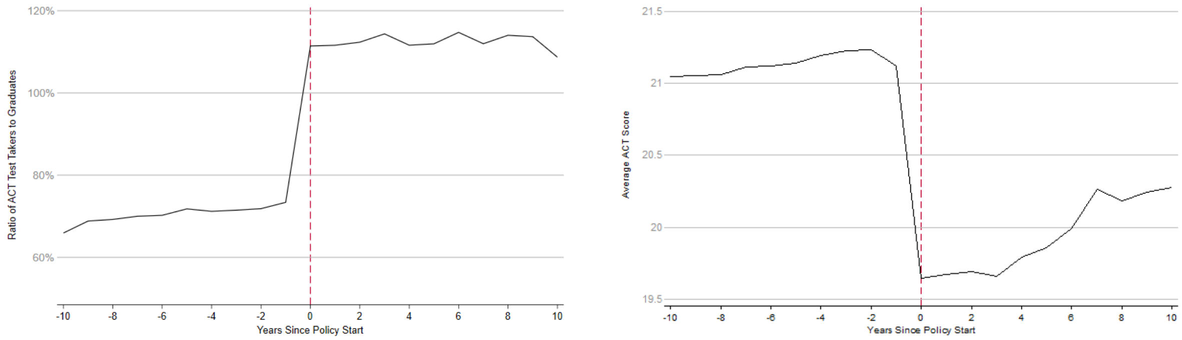
Note: This map shows the years states implemented a mandatory college entrance exam policy. Years are divided evenly into 4 groups. The earliest policy implementation is 2001 and the latest is 2019. I am able to evaluate only policies implemented before 2018 (this amounts to 30 states, of 32 with policies through 2019).

Figure 3.3: Parallel Pre-trends in College Attendance



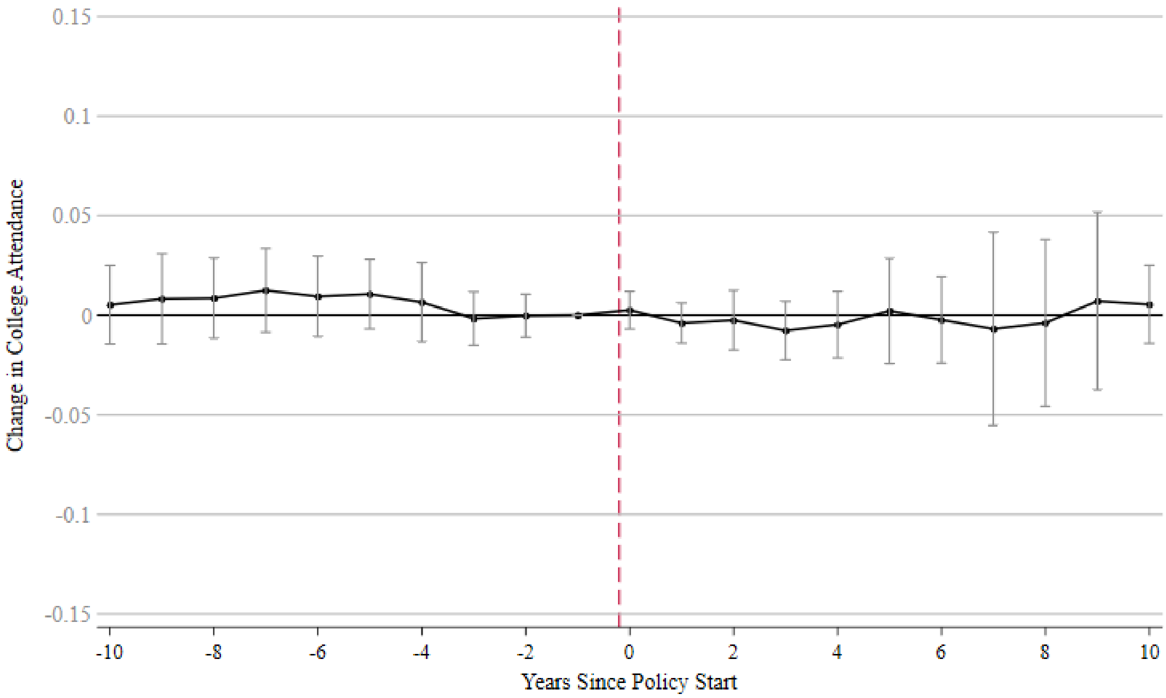
Note: This figure shows college attendance trends in treatment and control states for the 2001 and 2015 policy-year groups. Trends are parallel far into the future after the 2001 policy implementation. Similarly, trends in college attendance are parallel for many years prior to policy implementation in the 2015 states. Treatment states in 2001: Colorado and Illinois. Treatment states in 2015: Minnesota, Mississippi, Missouri, Nevada, South Carolina, and Wisconsin. Control states are defined as those who never implemented a mandatory college entrance exam. Sample is 17-47 year olds from the 2000-2018 ACS.

Figure 3.4: Sharp Testing Changes in ACT Policy States



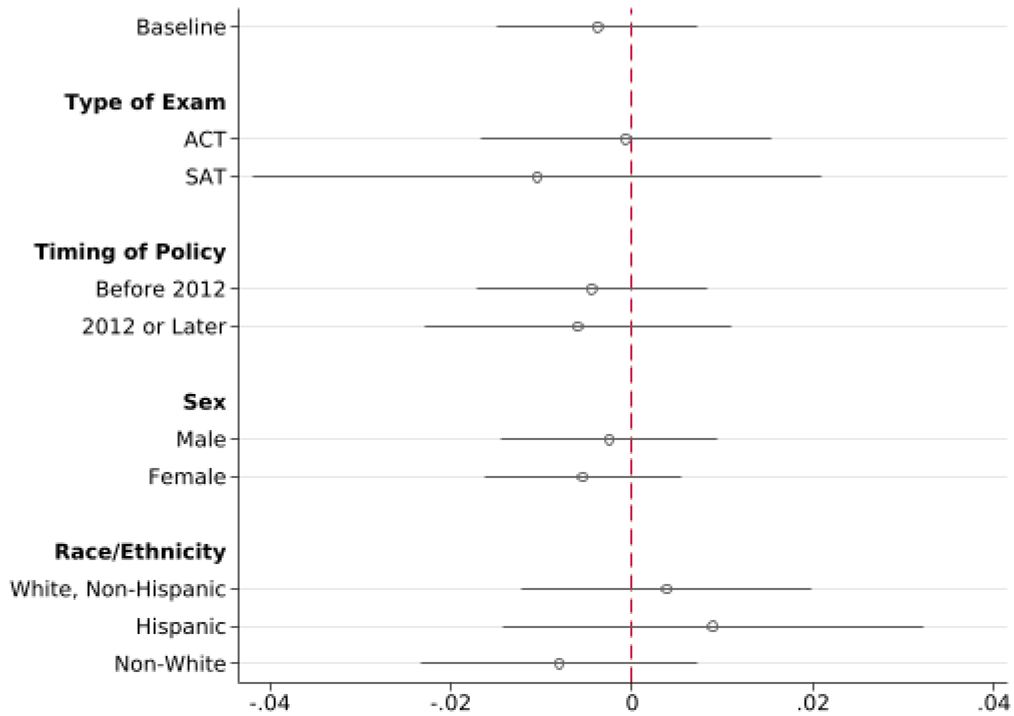
Note: This figure shows the large jump in ACT test takers and the large drop in average ACT scores exactly at the year of mandatory ACT implementation for policy states. The proportion of ACT testers is calculated as the number of testers divided by the number of graduates, where graduate counts come from the NCES and are projected after the 2012-2013 school year. The percent of test takers can be greater than 100 because students are able to take the exam multiple times.

Figure 3.5: Regression-Adjusted Change in College Attendance



Note: This figure shows the results of the difference-in-differences event study specification. Attendance remains very smooth across the policy implementation. College attendance is defined as an indicator for whether an individual in cohort *t* attended college one year or more after graduating high school. Cohort is defined as the year the individual would have been a senior using standard grade ages. Sample is 17 to 47 year olds from the 2000-2018 ACS.

Figure 3.6: Heterogeneity in the Effect of Exam Policies on College Attendance



Note: This figure presents the results from a two-way fixed effects regression of college attendance on policy status. I estimate the effect separately for groups defined by four characteristics: the type of exam, the timing of the policy, sex, and race/ethnicity. College attendance is defined as an indicator for whether an individual in cohort *t* attended college one year or more after graduating high school. Cohort is defined as the year the individual would have been a senior using standard grade ages. Sample is 17 to 47 year olds from the 2000-2018 ACS.

Appendix A: Appendix to Chapter 1

A.1 Cost Benefit Calculations

This appendix details the calculations and sources that make up the cost-benefit estimates discussed in Section 1.6. Because there are a plethora of sources on various costs of crime, and each source approaches the accounting differently, I endeavor to create high and low estimates whenever possible, and report the average of those estimates in Section 1.6. As discussed in that section, I also endeavor to use costs specific to type of crime whenever available. I convert all monetary values to 2022 dollars using the BLS CPI inflation calculator.[1]

A.1.1 Costs of Mental Health Treatment

The main difficulty in estimating the costs of the policy comes from the large variation in the possible components of treatment. For this analysis I define treatment as the combination of an initial evaluation, weekly meetings with a therapist for the duration of the probation sentence, and a 50% chance of taking antidepressants. For the low estimate I use the average community sentence length (4.8 months), the lower bound cost of a monthly supply of antidepressants ($4.60, Cherney (2020)), and the facility cost of therapy and evaluation ($80.17 and $127.12, DHHS (2013)). For the high estimate I use the average intermediate sentence length (9.6 months), the

[1] `https://data.bls.gov/cgi-bin/cpicalc.pl`

upper bound cost of a monthly supply of antidepressants ($149.50, Cherney (2020)), and the non-facility cost of therapy and evaluation ($87.53 and $161.75, DHHS (2013)).[2] In both cases, I adjust the cost of medication based on the Medicaid rebate, estimated at 0.231 by the Congressional Budget Office (CBO, 2022). The combination of initial evaluation and weekly therapy and medication for the duration of the sentence results in a baseline estimate within [$1,846.92, $4,237.64].

I make two adjustments to that baseline estimate of the costs. First, I address the likely individuals who are currently using Medicaid's services. In North Carolina, which did not expand Medicaid with the Affordable Care Act, eligibility for coverage comes predominantly from disability or responsibility for minors paired with an income requirement DHHS (2016). Using estimates from Glaze and Maruschak (2008) for the number of offenders who are parents of and financially responsible for children and were employed before prison, along with the distribution of earnings among employed offenders before prison from Garin et al. (2023), I calculate that %17 to %23 of probationers would be eligible for Medicaid due to being parents. Using estimates from Wang (2022) for the number of disabled prisoners and BLS (2014) for the employment rate among people with a disability, along with the distribution in earnings from Garin et al. (2023), I calculate that %38 of probationers would be eligible for Medicaid due to disability. I combine these and scale by the take-up rate of %61 among eligibles from Sommers et al. (2012) resulting in an estimated %34 to %38 of probationers who would have already been using Medicaid for help with medical costs. Subtracting the costs of providing medication and eight sessions of therapy (the maximum covered in a year by Medicaid) for those already-covered pro-

[2]Community and intermediate are the two types of probation sentences in North Carolina; intermediate involves an assigned probation officer, while community can be less regimented.

bationers, I arrive at my main estimate of the costs of treatment, \$2,705.65 (interval: [\$1,641.20, \$3,770.09]).

The second adjustment addresses the potential behavioral response of taxpayers if the policy were paid for by an increase in taxes. I adjust the cost estimate by adding the deadweight loss due to funding the policy by increasing federal income taxes. I use Saez et al.'s estimate of the marginal excess burden per tax dollar at \$0.195 (2012). Combining the effect of deadweight loss through taxation and the treatment that would have already been covered for about a third of probationers yields a final estimate of the costs of mental health treatment to the government of \$3,233.25 per probationer (interval: [\$1,961.24, \$4,505.25]).

A.1.2 Benefits of Mental Health Treatment

The benefits of mental health treatment among probationers likely include more than just the reduction in costs associated with lower crime. However, this paper's focus is on the effects of treatment on future criminal involvement, and lacks information on other potential benefits. I therefore proceed by quantifying only the decrease in costs associated with the decrease in crime, and consider these estimates to be a lower bound of the true monetary valuation of the benefits of mental health treatment.

Even limiting to the reduction in costs associated with lower crime, quantifying the costs proves a daunting task. While there have been several attempts to assign monetary values to the criminal justice system, different methods of valuation can produce strikingly different cost estimates. In addition, the costs vary enormously by the type of crime and the punishment. For example, the direct operating costs associated with a year of prison average \$42,862 in 2022 dollars, whereas the operating costs for a probationer average only $3,973 (Hunt et al., 2017). Similar variation exists in the social costs associated with different crimes. With any crime that involves valuing a human life, estimates of the social cost are over $8.6 million; in contrast, selling or possessing stolen property has an average social cost of $686 per crime, and other crimes might have zero social cost. In order to incorporate that variation, I use cost estimates that are specific to crime types whenever possible. I also synthesize the existing literature into high and low estimates to address the sometimes large differences in estimates for ostensibly the same category of costs.

I consider four categories of costs associated with crime. The first is judicial costs, which include the operating costs of police forces, courts, jails, prisons, and probation and parole. The second is loss of tax revenue to the government due to prisoners not being employed. The third is impacts on offenders, such as through loss of earnings, court fines, and increased health issues due to incarceration. The fourth is the social costs, which are also called the costs of victimization or the intangible costs. These include medical bills and destroyed property of victims, as well as psychological effects due to fear of crime.

When sources break costs apart by crime, they often do so only for the FBI Part I crimes. Those include homicide, rape, robbery, aggravated assault, burglary, larceny, motor vehicle theft, and arson. Some sources provide partial estimates of total costs to more aggregated groups of Part II crimes, which are all crimes outside of the Part I crimes. To estimate costs for Part II crimes, I use the minimum cost from those partial estimates as a lower bound estimate and the average cost without disaggregating by crime type as the upper bound estimate. I assign costs to all Part I and Part II crime types, then aggregate those costs to the five offense groups used throughout this analysis (violent and property, financial and fraud, traffic and public order, drugs and alcohol, and miscellaneous). The "cost to offenders" category is not available disaggregated into crime types. Because most of the costs to offender are associated with incarceration, I weight the low and high estimates by the likelihood of incarceration within each offense group to form crime-type-specific estimates. Table A.1 describes the sources used for each category of costs.

Table A.1: Sources for Costs of Crime

Type of Cost	Source
Judicial Costs	Hunt et al. (2017), McCollister et al. (2010)
Loss of Tax Revenue	McCollister et al. (2010)
Costs to Offenders	Binswanger et al. (2007), Garin et al. (2023), McLaughlin and Stokes (2002), McLaughlin et al. (2016), NIMH (2023), Rafael (2023), The Annie E. Casey Foundation (2023)
Social Costs	Heaton (2010), Hunt et al. (2017), McCollister et al. (2010)

Note: This table lists the sources for various costs associated with crime, used to quantify the benefits from the reduction in crime due to mental health treatment.

The judicial, social, and lost tax revenue costs come directly from the listed sources, but the costs to offenders involve some additional calculations. The base estimates come from McLaughlin et al. (2016) and include lost wages while incarcerated, sexual abuse in prison, suicide risk, mortality risk, homelessness, moving, eviction, and adverse health. I also include reduced lifetime earnings from Garin et al. (2023) and conviction-related fines from Rafael (2023). Costs associated with the risks of homelessness, moving, eviction, and adverse health are given as annual costs over all offenders in the criminal justice system; I divide by the number of prisoners, probationers, and parolees in 2009 (5.8 million) to form a per-offense cost. The suicide and mortality costs are determined by comparing rates in prison to rates outside of prison; I adjust this by comparing to rates outside of prison among individuals with similar demographic characteristics.

I use the suicide rate of Black individuals (13 out of 100,000) from NIMH (2023). I use the mortality rates from Binswanger et al. (2007) but adjust the comparison group's rate using the association between decreases in income and increases in mortality from McLaughlin and Stokes (2002), the median household income for Black individuals from The Annie E. Casey Foundation (2023), and the average income (conditional on employment) before incarceration from Garin et al. (2023). Because these costs to offenders are based on incarceration, I adjust the estimates by the likelihood of a prison sentence within each crime type.

Table A.2: Benefits from Reduction in Crime

Offense Type	Fraction of Sample	Judicial Costs	Costs to Offenders	Costs to Society	Loss of Revenue
Violent/Property	.28	$1,181, $14,497	$37,720, $77,583	$113,632, $217,089	$4,204, $7,235
Financial/Fraud	.24	$4,227, $5,162	$10,764, $22,140	$1,382, $35,617	$627, $7,232
Drugs/Alcohol	.22	$1,719, $4,910	$20,486, $42,137	$17,444, $86,247	$877, $11,836
Traffic/Public Order	.25	$8,649, $10,383	$16,782, $34,517	$3,341, $54,766	$492, $10,423
Miscellaneous	.01	$14, $14,741	$39,587, $81,424	$686, $91,543	$228, $18,301

Note: This table presents the lower and upper bound estimates for the costs of each crime, broken apart by four categories of costs and reported separately for five offense types. The costs are used to quantify the benefits from the reduction in crime due to mental health treatment. The sources for the costs are listed in Table A.1.

To form the final estimate of the benefits of mental health treatment, I combine the costs of each crime type with the decrease in that crime type associated with treatment, then aggregate the resulting estimates based on the relative prevalence of each future crime type in the sample. Table A.2 lists the low and high estimates for each crime in each category of costs. The variation in estimates comes from the different methods used by the various studies I rely on for cost information. Incorporating the estimated effect of mental health treatment on recidivism adds an additional source of uncertainty via the standard errors on the point estimates. Using the low estimates of crime costs, I estimate total benefits from the reduction in crime due to mental health treatment of about $9,185 (confidence interval: [$4,390, $13,985]). Using the high estimates of crime costs, I estimate the total benefits at about $23,077 (confidence interval: [$10,690, $35,484]).

A.2 Data Preparation

A.2.1 ACIS

Data from the North Carolina Administrative Office of the Courts (AOC)'s Automated Criminal Infractions System (ACIS) is provided in a series of text files. I link those files using the case and charge keys. The result is a case-charge level data set. There are several entries per case for two reasons. First, one arrest may lead to several charges which are addressed in the same trial date. Second, the data include several entries per charge in order to describe additional sentence details. I collapse the data to the case level. I define most characteristics of the case using the most severe offense from among the charges. For example, the punishment (prison versus probation) is determined from the most severe charge. However, for special conditions of probation and certain aspects of the case, I use information from all charges. For example, a defendant received a mental health treatment mandate if any of the charges associated with the case involve a mental health treatment mandate. Similarly, a case is characterized as a domestic violence case if any of the component charges are related to domestic violence. When the same trial results in different disposition dates, I base my analysis on the earliest disposition date and associated judge. Later dates could represent future judgements based on the same case, such as probation violations.

I assign points following the Structured Sentencing Training and Reference Manual. Offenders are given more points if they have more prior sentences or if they are charged with a more severe offense. These points determine the class of sentence that offenders could receive – either active, intermediate, or community. Probation sentences are either community or intermediate, whereas prison sentences are active. Appendix Figures A.1 and A.2 depict how the points map to sentences for felonies and misdemeanors. Among felonies, offense class I (the lowest) with 0, 1-4, or 5-8 prior points result in C or I punishments. Misdemeanors of classes 1 and 2 with prior point levels 0 and 1-4, and class 3 with 0 prior points, all result in C or I punishments.

Identification of mental health treatment comes from the additional sentence details. I define offenders as being treated if any of the phrases "mental," "mntl," "mntal," "eval," "exam," "exm," "asses," "couns," "cnsl," "therap," "trt," "trea," "psy," "behavioral," "trmnt," "prescribed medicine," "prescribed meds," "psd meds," "mental health med," "mental med," "depress," "anger," "stress," "anxi," or "mood disord" appear in the detail field. I identify substance use disorder (SUD) treatment from the same additional sentence details, characterized by the phrases "drug trt," "drg trt," "drug trea," "drg trea," "drug eval," "drg eval," "alcohol," "DART," "subs," "sub abus," or "TASC." In both cases, I do not characterize the offender as being treated if the phrase "court does not recommend" also appears in the sentence field.

As my main definition of treatment, I make the additional restriction that phrases qualified with 'substance abuse' or 'drug' - for example, if the phrase 'treatment' appears but is preceded by 'drug' - do not contribute toward mental health treatment cases. Specifically, I remove cases from the treatment definition if I have also categorized them as SUD treatment *and* they do not include any of "mental," "mntl," "mntal," "couns," "therap," "psy," "behavioral," "mental health med," "mental med," "depress," "anxi," and "mood disord."

I consider several alternate definitions of treatment, including one based only on the above phrases without removing likely SUD-only cases. I broaden the definition further by including offenders likely in some sort of program, characterized by “program,” “medical issu,” “medical eval,” “medical prob,” “medical trea,” “residential,” “inpatient,” or “rehab.” This represents the most broad definition of treatment. I also narrow it in several ways. From the baseline measure, I remove all cases that are also categorized as SUD treatment. In another more narrow definition, I remove cases that might have been sent to a mental health court, characterized by “S.T.E.P.,” “mental health court,” “by mhc,” “community resource court,” “to crc,” “in crc,” “crc court,” “crc prog,” “complete crc,” “by crc,” “completed crc,” or “attend crc.” These represent the most narrow definition.

I create longitudinal links using name and birth date in order to track offenders’ cases and identify recidivism. The data contain inconsistencies in the spelling of names. I correct errors in which similar names have the same other key characteristics. For example, I treat “Vasquez, Jose Perrero Arturo” and “Vasquez, Jose Perrero” as the same person if they have the same birth date, race, and sex. Similarly, I correct birth date inconsistencies when other key characteristics are the same. For example, I treat “011754” and “101754” as the same person when the name, race, and zip code exactly match. When birth dates disagree as in this example, I calculate the defendant’s age using the most frequent birth date among that defendant’s cases.

A.2.2 Judge Schedules and District Maps

I create a list of judges and their tenures using yearly documents from the NC Judicial Department website.[3] The same website provides judicial district maps, which I use to determine the superior court circuits and for verification of judge districts. Spring and Fall schedules are also available, from which I draw the exact judges who were presiding in a given court at a given

[3] `https://www.nccourts.gov/documents/publications`

time of day and day of week.

I combine the supplemental information with the ACIS in several stages. First, I add information about the District and Superior districts and the Superior circuit based on the county and year associated with the trial. Next, I add information about the schedules of the judges based on the judge, district, circuit, court, year, and season associated with the case. When doing so, I correct obvious spelling errors in the ACIS judge initials, such as cases where the "O" was typed as a "0." In addition, the judge is only identified by a two- or three-letter initial in the ACIS, whereas my supplemental information gives the full name of the judge. I create both two- and three-letter versions of the judge initials from the full name in order to combine files. When the ACIS is missing information about court, circuit, and district, I instead add the judge schedule information using only judge, year, and season.

A.2.3 Sample Restrictions

Table A.3: Sample Reductions

	Observations	% Dropped	% Remaining of Original
Start	2108124	0.0	100.0
Key variables	1612161	23.5	76.5
Adults	1572485	2.5	74.6
Struct. Sent. Period	1440145	8.4	68.3
No Drug Court	1439406	0.1	68.3
Judge Cases	1375395	4.4	65.2
Probation Charge	1039994	24.4	49.3
Enough Cases per Judge	1027754	1.2	48.8

Note: This table walks through each sample restriction, starting from the raw data and ending with a sample of individuals who were only at risk of a probation sentence (no chance of a prison sentence) due to the Structured Sentencing Laws in place from October 1, 1994 to December 1, 2009.

Table A.3 displays the sample restrictions made. After collapsing to the case level, there are 2,108,141 offender-trial observations in the raw data.[4] I limit the sample to offenders with key information, including an individual identifier, a judge, and the judge's district or court. I require offenders in my sample to be adults, which in North Carolina means 16 or older, and remove individuals with unrealistic ages (older than 100). I limit the time period to criminal court cases from October 1, 1994 to December 1, 2009, during the first era of structured sentencing. I remove those in "A" and "I/A" punishment classes based on their charged offense, losing 20 percent compared to the step before. This leaves only individuals whose only available sentence is probation. I also remove judges with fewer than 10 cases in a year in order to have enough variation to construct the instrument. After implementing these restrictions, the sample includes 1,028,087 cases, 51,404 (5 percent) of which involved a sentence to seek mental counseling.

[4] The original sample included several entries per case in order to describe additional sentence details.

A.3 Additional Tables and Figures

Table A.4: Offender Characteristics Before Limiting to Probation Cases

	All	Limit to Probation
Age	30.94	30.80
	(10.41)	(10.37)
Female	0.18	0.20
	(0.38)	(0.40)
Black	0.47	0.45
	(0.50)	(0.50)
Hispanic	0.02	0.03
	(0.16)	(0.16)
Violent/Property Crime	0.34	0.27
	(0.47)	(0.44)
Drug/Alcohol Crime	0.18	0.19
	(0.38)	(0.39)
Financial/Fraud Crime	0.13	0.15
	(0.33)	(0.35)
Traffic/Public Order Crime	0.32	0.36
	(0.47)	(0.48)
Sex Offender	0.01	0.01
	(0.12)	(0.11)
First Time Offender	0.58	0.59
	(0.49)	(0.49)
Private Attorney	0.28	0.30
	(0.45)	(0.46)
Assigned MH Treatment	0.05	0.05
	(0.22)	(0.21)
Assigned SUD Treatment	0.05	0.04
	(0.21)	(0.20)
Recidivism within 3 Years	0.33	0.34
	(0.47)	(0.47)
Observations	1,439,406	1,039,994

Note: This table presents summary statistics for the sample before and after the restriction to probation cases is made. The second column is a subset of the sample in the first column. I designate First-Time Offender based on when offenders first appear within the sample period beginning in 1994, which could misclassify offenders who were arrested in years prior to 1994. "SUD" stands for Substance Use Disorder. Source: 1994-2009 North Carolina ACIS data.

Table A.5: Effect of Mental Health Treatment on Recidivism, by Years Since Trial

	1 Year	2 Years	3 Years	4 Years	5 Years
Assigned MHT	–0.058	–0.087**	–0.120***	–0.123**	–0.095*
	(0.037)	(0.044)	(0.044)	(0.051)	(0.055)
1st Stage F-Stat	970	970	970	829	690
Mean of Dependent Var	0.20	0.28	0.34	0.38	0.41
Observations	727,184	727,184	727,184	632,772	547,732
Adj. R^2	0.045	0.056	0.066	0.074	0.083

Note: This table presents results from OLS and IV regressions of recidivism on being mandated to seek mental health treatment. Recidivism is defined first looking only one year out from the trial, then two years, and so on up to five years. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Robustness: Comparison of Different Instrument Cluster Levels

	Base	Remove Time Variables	Remove Circuit
Assigned MHT	–0.121***	–0.139***	–0.152***
	(0.044)	(0.053)	(0.055)
1st Stage F-Stat	970	428	402
Mean of Outcome	0.342	0.342	0.342
Observations	727,184	727,184	727,184
Adj. R^2	0.066	0.067	0.067

Note: This table presents results from OLS and IV regressions of three-year recidivism on being mandated to seek mental health treatment. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects, but the columns vary in how the court-time effects are defined. Column 1 presents the baseline specification. Column 2 removes time variables, using only the district or circuit-district to form the groups. Column 3 removes circuit as well. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Robustness: Comparison of Different Instrument Time Horizons

	Base	First Year	Three Years	By Year	Omit Cluster
Assigned MHT	–0.121***	–0.052	–0.118***	–0.122***	–0.127***
	(0.044)	(0.115)	(0.039)	(0.047)	(0.047)
1st Stage F-Stat	970	104	937	574	732
Mean of Outcome	0.342	0.339	0.342	0.343	0.342
Observations	727,184	620,892	725,925	723,679	726,822
Adj. R^2	0.066	0.067	0.066	0.066	0.065

Note: This table presents results from OLS and IV regressions of three-year recidivism on being mandated to seek mental health treatment. The columns vary by the time horizons used to create the instrument. Column 1 presents the baseline specification. Column 2 uses only cases from the judge's first year to form the instrument. Column 3 creates the instrument within groups of three years. Column 4 creates the instrument by each year. Column 5 omits cases within the same court-time group. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Robustness: Full Saturation in Covariates

	Regular Controls	Saturated Controls	Saturated Instrument	Saturated Controls & Instrument
Assigned MHT	–0.121***	–0.100**	–0.122***	–0.092**
	(0.044)	(0.045)	(0.044)	(0.044)
1st Stage F-Stat	970	930	220	214
Mean of Outcome	0.342	0.342	0.342	0.342
Observations	727,184	713,730	727,184	713,730
Adj. R^2	0.066	0.054	0.066	0.054

Note: This table presents results from OLS and IV regressions of three-year recidivism on being mandated to seek mental health treatment. Columns 2-4 fully saturate the model with single, double, and triple combinations of the controls and the instruments. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Robustness: Convictions

	Convicted Sample		Control for Judge Conviction	
	OLS	IV	OLS	IV
Assigned MHT	−0.051***	−0.138***	−0.047***	−0.142***
	(0.003)	(0.046)	(0.003)	(0.044)
1st Stage F-Stat		884		961
Mean of Outcome	0.352	0.352	0.342	0.342
Observations	644,915	644,915	727,204	727,204
R^2	0.098	0.064	0.095	0.065

Note: This table presents results from OLS and IV regressions of three-year recidivism on being mandated to seek mental health treatment. The first two columns limit the sample to only those cases which resulted in a conviction. The second two columns use the main analysis sample, but control for the judge's tendency toward conviction. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Robustness: Definitions of Mental Health Treatment

	Less Restrictive	Base	More Restrictive	Most Restrictive
Assigned MHT	–0.152***	–0.141***	–0.134***	–0.165***
	(0.044)	(0.041)	(0.047)	(0.059)
1st Stage F-Stat	997	1126	851	762
Mean of Outcome	0.342	0.342	0.342	0.342
Observations	727,184	727,184	727,184	727,184
Adj. R^2	0.065	0.065	0.065	0.065

Note: This table presents results from OLS and IV regressions of three-year recidivism on being mandated to seek mental health treatment. The columns vary by which key words were included when defining mental health treatment. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11: Robustness: Definitions of SUD (Substance Use Disorder) treatment

	Least Restrictive	Less Restrictive	Base	Most Restrictive
Assigned MHT	–0.096**	–0.093**	–0.120***	–0.127***
	(0.045)	(0.045)	(0.044)	(0.041)
1st Stage F-Stat	984	927	970	1090
Mean of Outcome	0.342	0.342	0.342	0.342
Observations	727,184	727,184	727,184	727,184
Adj. R^2	0.066	0.066	0.066	0.066

Note: This table presents results from OLS and IV regressions of three-year recidivism on being mandated to seek mental health treatment. The columns vary by which key words were included when defining substance use disorder treatment. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.12: Combining Treatment: Either Mental Health or SUD Treatment

	OLS		IV	
MHT or SUDT	–0.093***	–0.045***	–0.135***	–0.129***
	(0.003)	(0.002)	(0.033)	(0.032)
1st Stage F-Stat			1470	1345
Mean of Outcome	0.342	0.342	0.342	0.342
Controls		X		X
Observations	727,184	727,184	727,184	727,184
Adj. R^2	0.033	0.095	0.022	0.065

Note: This table presents results from OLS and IV regressions of three-year recidivism on being mandated to seek either mental health (MH) or substance use disorder (SUD) treatment. The instrument is the judge's tendency toward mandating either MH or SUD treatment. All specifications condition on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.13: Separating Treatment: Disentangling Mental Health from SUD Treatment

	MHT		MHT & SUDT		SUDT		All Three	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
MHT	–0.052***	–0.162***					–0.055***	–0.151***
	(0.003)	(0.052)					(0.003)	(0.054)
MHT & SUDT			–0.034***	–0.158*			–0.039***	–0.010
			(0.004)	(0.086)			(0.004)	(0.090)
SUDT					–0.037***	–0.111*	–0.041***	–0.098
					(0.004)	(0.061)	(0.004)	(0.060)
1st Stage F-Stat		266		114		108		97
Mean of Outcome	0.342	0.342	0.342	0.342	0.342	0.342	0.342	0.342
Observations	727,184	727,060	727,184	727,060	727,184	727,060	727,184	727,060
Adj. R^2	0.095	0.065	0.095	0.065	0.095	0.066	0.095	0.065

Note: This table presents results from OLS and IV regressions of three-year recidivism on being mandated to seek four mutually exclusive treatments: mental health (MH) alone, substance use disorder (SUD) treatment alone, MH and SUD treatment together, and neither (which is the left-out category). Each treatment is instrumented with the judge's tendency toward assigning that specific treatment among other offenders in the sample. All specifications condition on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data. Standard errors, clustered at the court-time level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.1: Felony Punishment Classes

PRIOR RECORD LEVEL

OFFENSE CLASS	**I** **0 Points**	**II** **1-4 Points**	**III** **5-8 Points**	**IV** **9-14 Points**	**V** **15-18 Points**	**VI** **19+ Points**
A	**Death or Life Without Parole**					
B1	**A** *240 - 300* **192 - 240** *144 – 192*	**A** *288 - 360* **230 - 288** *173 – 230*	**A** *336 - 420* **269 - 336** *202 – 269*	**A** *384 - 480* **307 - 384** *230 – 307*	**A** *Life Without Parole* **346 - 433** *260 – 346*	**A** *Life Without Parole* **384 - 480** *288 - 384*
B2	**A** *157 - 196* **125 - 157** *94 - 125*	**A** *189 - 237* **151 - 189** *114 - 151*	**A** *220 - 276* **176 - 220** *132 - 176*	**A** *251 - 313* **201 - 251** *151 - 201*	**A** *282 - 353* **225 - 282** *169 - 225*	**A** *313 - 392* **251 - 313** *188 - 251*
C	**A** *73 – 92* **58 - 73** *44 - 58*	**A** *100 – 125* **80 - 100** *60 - 80*	**A** *116 – 145* **93 - 116** *70 - 93*	**A** *133 - 167* **107 - 133** *80 - 107*	**A** *151 - 188* **121 - 151** *90 - 121*	**A** *168 - 210* **135 - 168** *101 - 135*
D	**A** *64 - 80* **51 - 64** *38 - 51*	**A** *77 - 95* **61 - 77** *46 - 61*	**A** *103 - 129* **82 - 103** *61 - 82*	**A** *117 - 146* **94 - 117** *71 - 94*	**A** *133 - 167* **107 - 133** *80 - 107*	**A** *146 - 183* **117 - 146** *88 - 117*
E	**I/A** *25 - 31* **20 - 25** *15 - 20*	**I/A** *29 - 36* **23 - 29** *17 - 23*	**A** *34 - 42* **27 - 34** *20 - 27*	**A** *46 - 58* **37 - 46** *28 - 37*	**A** *53 - 66* **42 - 53** *32 - 42*	**A** *59 - 74* **47 - 59** *35 - 47*
F	**I/A** *16 - 20* **13 - 16** *10 - 13*	**I/A** *19 - 24* **15 - 19** *11 - 15*	**I/A** *21 - 26* **17 - 21** *13 - 17*	**A** *25 - 31* **20 - 25** *15 - 20*	**A** *34 - 42* **27 - 34** *20 - 27*	**A** *39 - 49* **31 - 39** *23 - 31*
G	**I/A** *13 - 16* **10 - 13** *8 - 10*	**I/A** *15 - 19* **12 - 15** *9 - 12*	**I/A** *16 - 20* **13 - 16** *10 - 13*	**I/A** *20 - 25* **16 - 20** *12 - 16*	**A** *21 - 26* **17 - 21** *13 - 17*	**A** *29 - 36* **23 - 29** *17 - 23*
H	**C/I/A** *6 - 8* **5 - 6** *4 - 5*	**I/A** *8 - 10* **6 - 8** *4 - 6*	**I/A** *10 - 12* **8 - 10** *6 - 8*	**I/A** *11 - 14* **9 - 11** *7 - 9*	**I/A** *15 - 19* **12 - 15** *9 - 12*	**A** *20 - 25* **16 - 20** *12 - 16*
I	**C** *6 - 8* **4 - 6** *3 - 4*	**C/I** *6 - 8* **4 - 6** *3 - 4*	**I** *6 - 8* **5 - 6** *4 - 5*	**I/A** *8 - 10* **6 - 8** *4 - 6*	**I/A** *9 - 11* **7 - 9** *5 - 7*	**I/A** *10 - 12* **8 - 10** *6 - 8*

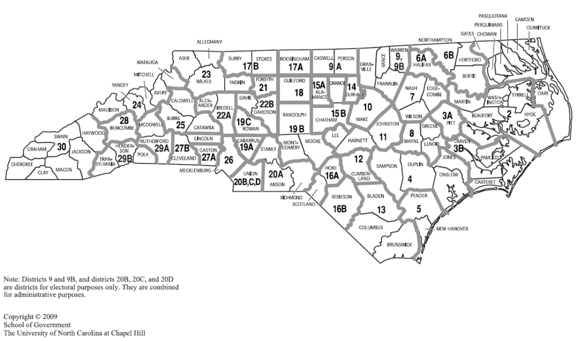
Note: This chart shows how the class of the current offense and the number and severity of prior convictions determine different levels of punishment and sentence lengths. A = Active, I = Intermediate, C = Community. The numbers in the cells represent sentence minimums in months. The majority of felony classes involve an active punishment (incarceration). Source: North Carolina Sentencing and Policy Advisory Commission's Structured Sentencing Training and Reference Manual.

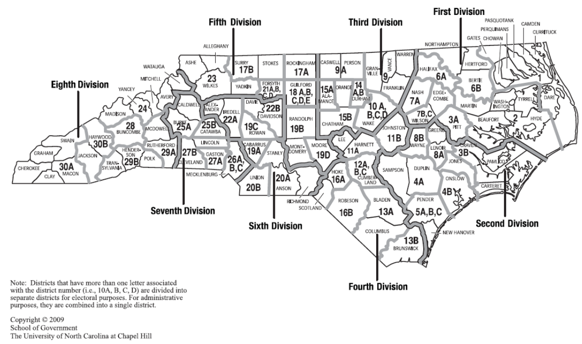
Figure A.2: Misdemeanor Punishment Classes

CLASS	PRIOR CONVICTION LEVEL		
	I **No Prior Convictions**	**II** **One to Four Prior Convictions**	**III** **Five or More Prior Convictions**
A1	**C/I/A** 1 - 60 days	**C/I/A** 1 - 75 days	**C/I/A** 1 - 150 days
1	**C** 1 - 45 days	**C/I/A** 1 - 45 days	**C/I/A** 1 - 120 days
2	**C** 1 - 30 days	**C/I** 1 - 45 days	**C/I/A** 1 - 60 days
3	**C** 1 - 10 days	**C/I** 1 - 15 days	**C/I/A** 1 - 20 days

Note: This chart shows how the class of the current offense and the number and severity of prior convictions determine different levels of punishment and sentence lengths. A = Active, I = Intermediate, C = Community. All misdemeanors classes offer options for non-active punishments. Source: North Carolina Sentencing and Policy Advisory Commission's Structured Sentencing Training and Reference Manual.

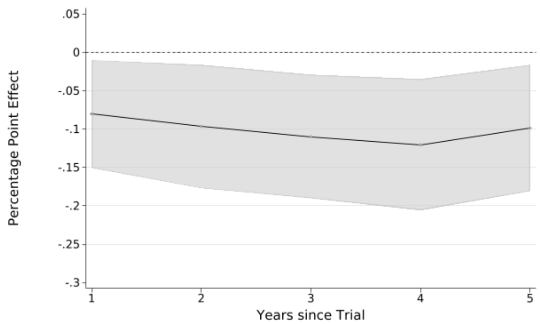
Figure A.3: District and Superior Court Boundaries in 2009



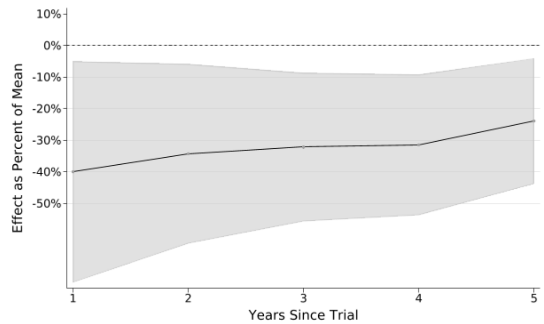


Note: These maps depict the geographic makeup of the two North Carolina court systems. The District Court (left) consists of 43 districts in which judges are randomly assigned to individual cases. The Superior Court (right) consists of 8 circuits (labeled "divisions" in this map), within which judges are rotated every 6 months to a new district. Source: North Carolina Courts Statistical and Operational Summary 2008-09.

Figure A.4: The Effect of Mental Health Treatment on Recidivism over Time in the Balanced Sample



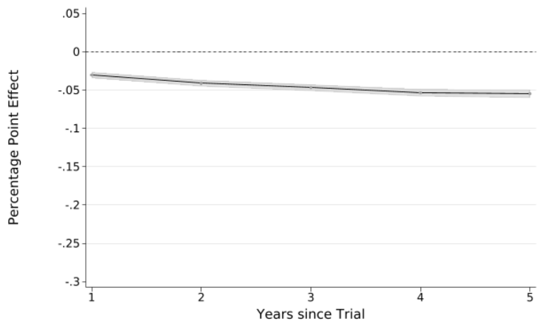
(a) Percentage Point



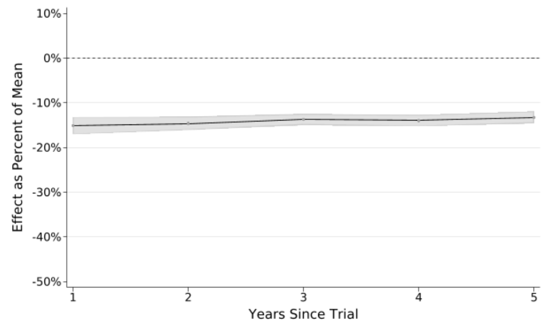
(b) Percent of Mean

Note: This figure plots the percentage point (left) and the percent (right) decline in recidivism due to mental health treatment, estimated from the main IV regression with full controls. Recidivism is defined first looking only within one year out from the trial, then between one and two years, then between two and three years, and so on up to five years. The estimation is carried out in a balanced sample that restricts to observations in the years 2004 and earlier. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data.

Figure A.5: The Effect of Mental Health Treatment on Recidivism over Time, OLS Estimates



(a) Percentage Point



(b) Percent of Mean

Note: This figure plots the percentage point (left) and the percent (right) decline in recidivism due to mental health treatment, estimated from the OLS regression with full controls. Recidivism is defined first looking only within one year out from the trial, then between one and two years, then between two and three years, and so on up to five years. All specifications condition on the judge's tendency toward substance use disorder treatment as well as on court-time effects. The included controls are race, ethnicity, sex, age cubic, first-time offender, region of the United States, shift, day of week, month, year, attorney type, offense types, whether the offender had a prior offense in the past year, and county-level descriptors. Source: 1994-2009 North Carolina ACIS data.

Appendix B: Appendix to Chapter 2

B.1 Measurement Details

B.1.1 Identifying Same-Sex Couples

The ACS collects detailed relationship information relative to Person 1 for up to 4 other people in the household through question 2. Up through 2018 in our sample, question 2 has 15 checkbox responses. Starting in 2019, question 2 has 16 responses. Figure A1 below shows the question 2 for both 2013 and 2019 and provides the instructions from 2013. To summarize, the following changes occur between 2013-2018 and 2019: (1) Husband-wife is disaggregated into opposite-sex and same-sex; (2) Unmarried partner is disaggregated into opposite-sex and same-sex and appears earlier in responses; (3) Roomer or boarder is dropped; and (4) ordering of phrase "Housemate or roommate" is switched.

A number of papers address the measurement of same-sex couples in available U.S. data. Starting with survey data, Badgett et al. (2021) compare results from the ACS to results from the National Health Interview Survey (NHIS) and Behavioral Risk Factor Surveillance System (BRFSS) and note consistencies across the sources. The NHIS and BRFSS are notable in that they have a question asking about sexual orientation, whereas in the ACS same-sex status can only be inferred by the sexes of the individuals in a couple until 2019 when it is explicitly collected. The

NHIS (2013-2018) identifies 1.4% of women as same-sex and 1.8% of men; in comparison the ACS (2008-2018) identifies 1.2% of women as same-sex and 1.1% of men. When limiting the ACS to years 2013 and beyond, the share of all couples who identify as same-sex is between 1.2% and 1.6%; in comparison the BRFSS (2014-2018) identifies 1.7% of all couples as same-sex.

Turning to administrative data, Fisher et al. (2018) combine tax records from 2013 to 2015 with the SSA Numident for information on sex and find that 0.48% of joint tax filers are same-sex, or about 59% of the number of same-sex partners identified in the ACS. There are many possible reasons for this mismatch, including misreporting on the survey, inaccuracies in the sex reported in the Social Security information, and differences in the likelihood to file jointly. Fisher et al. argue that the most compelling reason for the mismatch is the difference in how marriage is defined for tax versus survey purposes. In particular, in states where same-sex marriage had just been legalized, couples might not have been legally married on their tax returns but might identify as being in a marriage-like relationship, or might have been able to legally marry by the time of the survey.[1]

B.1.2 Identifying Child Timing

Note that we do not have birth records, unlike Moberg (2016) and Andresen and Nix (2022), who use data from several Scandinavian countries. Thus we cannot determine which of two female same-sex partners gave birth. In a similar fashion, we do not know if one of the partners in a same-sex male couple is genetically related to the child. In the ACS data, all relationships

[1]The counts tabulated from the three surveys and the administrative tax data discussed above pertain only to adults in relationships. For an overall tabulation of individuals identifying as same-sex in the United States, Gates (2010) finds that about 3.5% of adults identify as lesbian, gay, or bisexual. He uses the National Epidemiological Survey on Alcohol and Related Conditions (2004-2005), National Survey of Family Growth (2006-2008), General Social Survey (2008), California Health Interview Survey (2009), and National Survey of Sexual Health and Behavior (2009) for his calculations.

are defined relative to the reported head of household. In our main results we combine three categories of children: biological children, adopted children, and stepchildren, and do not attempt to separately identify the birth and non-birth parent.

Several studies have provided information on the differences between same-sex couples and opposite-sex couples in terms of the timing, amount, and likelihood of having children. Using the 2019 ACS, Walker and Taylor (2021) find that same-sex couples have far fewer children than opposite-sex couples (about 59% of opposite-sex couples are childless, while about 86% of same-sex couples are childless). Badgett et al. (2021) come to similar conclusion using both the 2008-2018 ACS and 2013-2018 NHIS. From the ACS they calculate that about 44% of women in opposite-sex couples have children in the household (38% of men), whereas only 22% of women in same-sex couples have children (10% of men). Using the NHIS they can identify same-sex individuals even when they are not in a couple, so their calculations are for all adults rather than only those in couples. They find that 33% of straight women have children (31% of men), while 20% of lesbian women have children (5% of gay men).

Since adoption is an important component of family formation for same-sex couples, we needed to correct for potential timing differences between the child's date of birth and the date the child joined the ACS household we observe them in. We first identified the population of adopted children in the 2013-2019 ACS using relationship type 3 (adopted child) to the head of household. We only kept those who could be matched to Census data, or had a Protected Identity Key (PIK) for both the parent (either household head or spouse) and the child. We then cleaned up duplicates, which mostly seemed to consist of the same household being surveyed in the ACS multiple times in the 2013-2019 span. In cases of duplicate parent-PIK and child-PIK pairs, we kept the earliest ACS record. But we otherwise did not deduplicate on child or adult PIK, so duplicate child PIKs remain when they match to more than one parent PIK and duplicate parent PIKs remain when they match to more than one adopted child. We used records from up to 6 children in each ACS household, starting with the eldest.

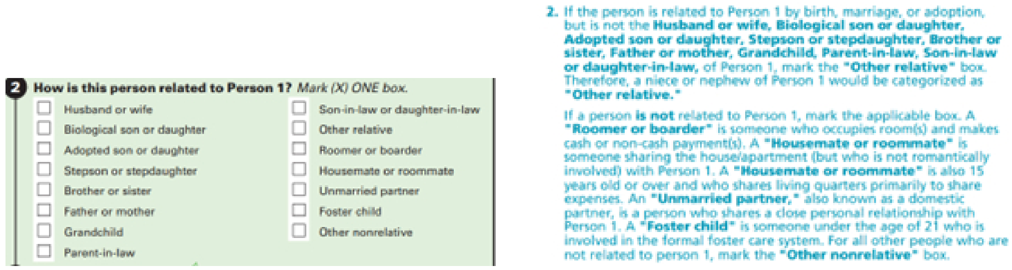
We then merged the parent PIKs from the ACS adopted children file to the IRS 1040 data for years 2000-2018. We did this as a many-to-many match, since there were duplicates on both sides. On the ACS side, duplicates occurred because there were duplicates parent PIKs when parents had more than one adopted child; on the IRS side, duplicates occurred when the parent PIK had multiple 1040 records.

For those records where a parent PIK matched, we then did a match to the child PIK, creating an indicator if the child PIK matched any of the (up to) 6 PIKs on the 1040 record. We then deduplicated the records within year. If there were 1040 records for a PIK that both included the child and did not include the child, we chose the ones with the child PIK. We also chose records where the spouse PIK appeared rather than ones where the spouse PIK did not. After those two targeted deduplications, we randomly dropped the remaining duplicates of parent-PIK, child-PIK pairs within year.

We then merged all of the IRS years 2000-2018 together, assigned a first year observed to each child-PIK, parent-PIK pair to get a year from the IRS data. To find an estimated age at adoption we combined the child-birth year from the Census Numident, if available (or the ACS date of birth if not) with this first year observed in the IRS. If the parent PIKs never appear in the IRS 1040 data, then both the parent and the child will not appear in this final file, and the ACS birth date is then used as a default, instead of the estimated adoption date.

B.2 Additional Tables and Figures

Figure B.1: Relationship Question in the ACS 2013 and 2019



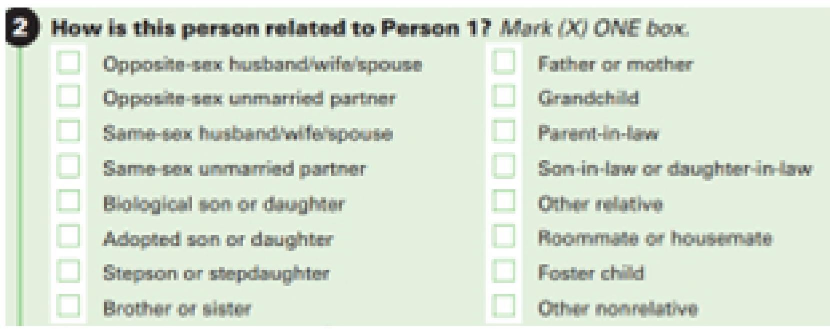
2 How is this person related to Person 1? *Mark (X) ONE box.*

- ☐ Husband or wife
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roomer or boarder
- ☐ Housemate or roommate
- ☐ Unmarried partner
- ☐ Foster child
- ☐ Other nonrelative

2. If the person is related to Person 1 by birth, marriage, or adoption, but is not the **Husband or wife, Biological son or daughter, Adopted son or daughter, Stepson or stepdaughter, Brother or sister, Father or mother, Grandchild, Parent-in-law, Son-in-law or daughter-in-law,** of Person 1, mark the **"Other relative"** box. Therefore, a niece or nephew of Person 1 would be categorized as **"Other relative."**

If a person **is not** related to Person 1, mark the applicable box. A **"Roomer or boarder"** is someone who occupies room(s) and makes cash or non-cash payment(s). A **"Housemate or roommate"** is someone sharing the house/apartment (but who is not romantically involved) with Person 1. A **"Housemate or roommate"** is also 15 years old or over and who shares living quarters primarily to share expenses. An **"Unmarried partner,"** also known as a domestic partner, is a person who shares a close personal relationship with Person 1. A **"Foster child"** is someone under the age of 21 who is involved in the formal foster care system. For all other people who are not related to person 1, mark the **"Other nonrelative"** box.

(a) 2013



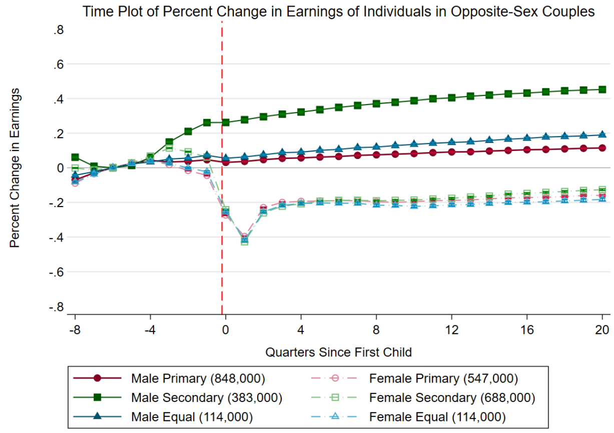
2 How is this person related to Person 1? *Mark (X) ONE box.*

- ☐ Opposite-sex husband/wife/spouse
- ☐ Opposite-sex unmarried partner
- ☐ Same-sex husband/wife/spouse
- ☐ Same-sex unmarried partner
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roommate or housemate
- ☐ Foster child
- ☐ Other nonrelative

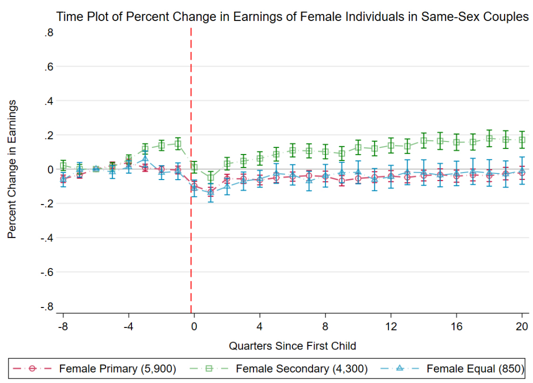
(b) 2019

Note: This figure shows how the relationship question was asked in the ACS in years 2013 (Panel A) and 2019 (Panel B). Source: ACS website `https://www.census.gov/programs-surveys/acs/about/forms-and-instructions.html`

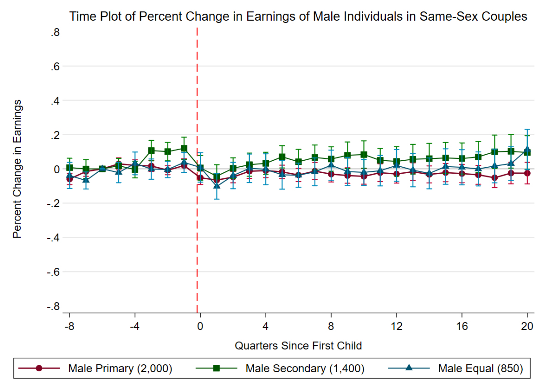
Figure B.2: Individual Earnings after Entry of First Child, Without Propensity Weights



(a) Individuals in Opposite-Sex Couples



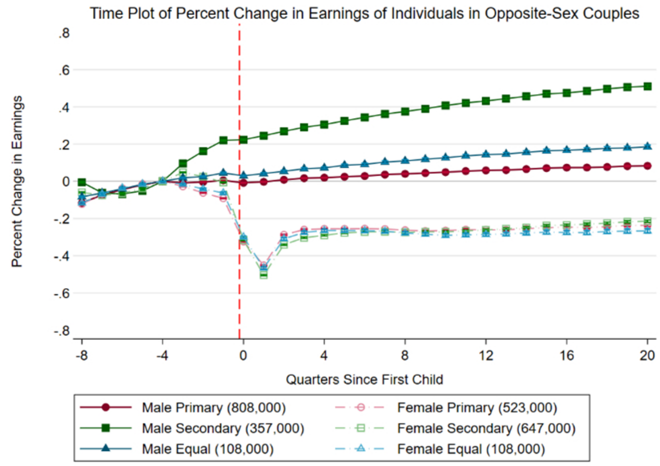
(b) Individuals in Female Same-Sex Couples



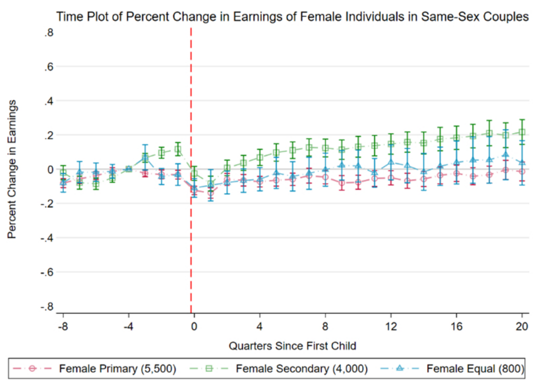
(c) Individuals in Male Same-Sex Couples

Note: This figure presents the main event studies of earnings around child entry, but the results use ACS sample weights alone rather than also incorporating propensity weights. For easier interpretation, earnings changes are presented as percent changes following Kleven et al. (2019). When the gap in pre-birth earnings is 10% or less, both members of the couple are designated as equal earners; when the gap in pre-birth earnings is larger than 10%, earners are designated as primary if they have higher earnings than their partner and secondary if they have lower. Individuals in the analysis sample are required to have nonzero earnings in at least one quarter in the two years before the child enters the household. Source: ACS, Numident, and IRS data.

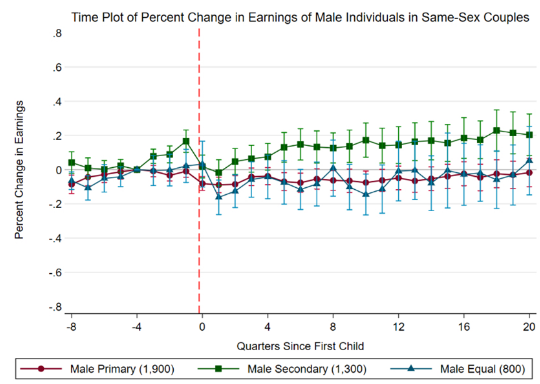
Figure B.3: Individual Earnings after Entry of First Child, Parents-to-be As Controls



(a) Individuals in Opposite-Sex Couples



(b) Individuals in Female Same-Sex Couples



(c) Individuals in Male Same-Sex Couples

Note: This figure presents the main event studies of earnings around child entry, but the results only use earnings up to 2014. Parents whose children enter the household between 2014-2019 then serve as the control group, comparing with parents whose children enter the household before 2014. For easier interpretation, earnings changes are presented as percent changes following Kleven et al. (2019). When the gap in pre-birth earnings is 10% or less, both members of the couple are designated as equal earners; when the gap in pre-birth earnings is larger than 10%, earners are designated as primary if they have higher earnings than their partner and secondary if they have lower. Individuals in the analysis sample are required to have nonzero earnings in at least one quarter in the two years before the child enters the household. Source: ACS, Numident, and IRS data.

Appendix C: Appendix to Chapter 3

C.1 Additional Tables and Figures

Table C.1: State Mandatory Test Policies

ACT States

State	Policy Begin
Kentucky	2008
Wyoming	2008
Tennessee	2009
North Dakota	2010
Utah	2011
Louisiana	2012
North Carolina	2012
Hawaii	2013
Montana	2013
Alabama	2014
Minnesota	2015
Mississippi	2015
Missouri	2015
Nevada	2015
Wisconsin	2015
Arkansas	2016
Oklahoma	2016
Nebraska	2017
Ohio	2017
Kansas	2019

Control States

State	Some Districts
Alaska	2014
Arizona	2018
California	2017
Florida	2020
Georgia	
Indiana	
Iowa	2019
Maryland	2018
Massachusetts	
New Jersey	
New Mexico	2020
New York	2016
Oregon	2015
Pennsylvania	
South Dakota	
Texas	2019
Vermont	
Virginia	2019
Washington	2020

SAT States

State	Policy Begin
Maine	2006
Delaware	2011
Idaho	2012
District of Columbia	2013
Connecticut	2016
New Hampshire	2016
Rhode Island	2017
West Virginia	2018

ACT to SAT States

State	Policy Begin	Policy Switch
Colorado	2001	2017
Illinois	2001	2015
Michigan	2007	2016
South Carolina	2015	2017

Note: 24 states required the ACT and 12 required the SAT at some point during the sample years. Of those, 4 switched from requiring the ACT to requiring the SAT and 2 ended their policies after a short trial period (Minnesota in 2018 and Missouri in 2017). In several control states, a couple districts started offering free in-school ACT or SAT exams during the sample period; I drop those states from my estimation sample in a robustness check with no impact on my results.

Table C.2: Predictive Power of Covariates for States' Policy Implementation

	(1) Policy Implemented	(2) ACT Policy	(3) SAT Policy
Policy Implemented		0.721***	0.435***
		(0.076)	(0.083)
Average Poverty Rate	–0.001	0.001	–0.001
	(0.001)	(0.001)	(0.001)
Average Race	–0.350	0.375	–0.368
	(0.314)	(0.300)	(0.286)
Average Food Stamps	–0.543	–1.293	1.675
	(0.680)	(0.997)	(1.062)
Average Female	0.535*	–0.843	0.430
	(0.305)	(0.571)	(0.635)
Average Married	0.720	2.001	–2.496**
	(0.604)	(1.253)	(1.156)
Average English Speakers	1.024	1.795	–2.877
	(1.017)	(1.946)	(2.146)
Average Education	–0.003	–0.032*	0.037**
	(0.012)	(0.017)	(0.014)
Year Fixed Effects	X	X	X
Observations	1,577	969	969
Adjusted R^2	0.424	0.601	0.326

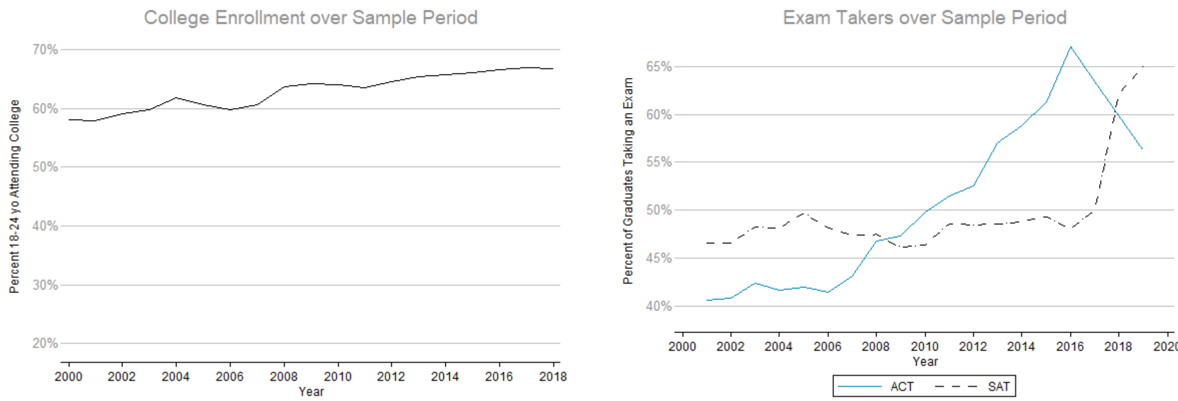
Note: Results from a regression of an indicator for whether states instituted any mandatory college entrance exam policy during the years 2001 to 2019 on various demographic characteristics of the states. The results suggest that average characteristics like race, poverty, and gender are for the most part not significantly correlated with policy implementation. Earnings converted to 2018 dollars using the CPI-U. Average state characteristics calculated from 17 to 47 year olds from the 2000-2018 ACS. Standard errors, clustered at the state level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: Effect of Mandatory Testing Policies on College Attendance (Robustness Checks)

	College					
	(1) Strict Control Group	(2) Likely Impacted	(3) Strict One Year	(4) State of Birth	(5) College Two Years	(6) Alternative College Def
Policy	0.000	0.004	–0.004	0.002	0.001	0.000
	(0.004)	(0.005)	(0.006)	(0.004)	(0.003)	(0.004)
State Fixed Effects	X	X	X	X	X	X
Cohort Fixed Effects	X	X	X	X	X	X
Controls	X	X	X	X	X	X
Year Fixed Effects	X	X	X	X	X	X
Cohort-Year Interactions	X	X	X	X	X	X
Observations	3,744,120	1,337,066	424,947	5,436,998	5,896,586	6,321,533
Adjusted R^2	0.160	0.133	0.109	0.140	0.177	0.176

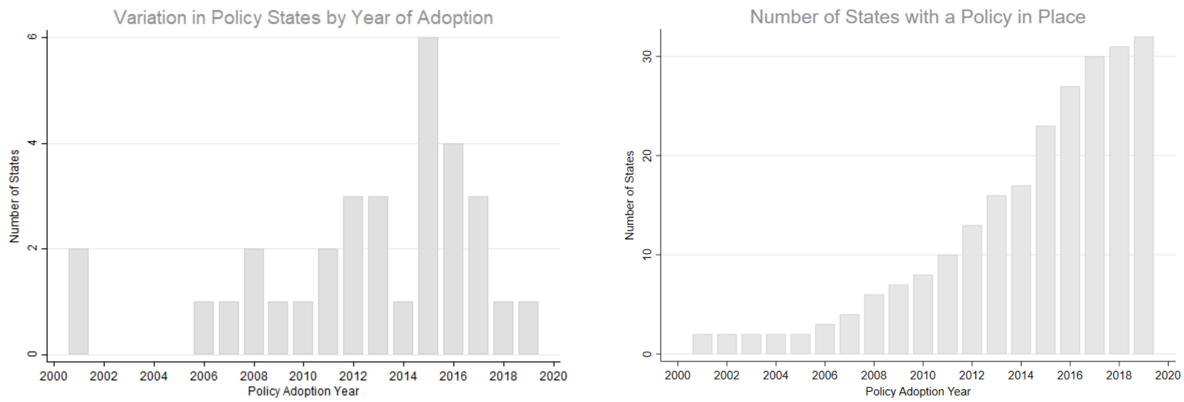
Note: Results from a two-way fixed effects regression of college attendance on policy status. Various robustness checks were conducted. In column (1), I drop control states with some district implementation of exams. In column (2), I restrict the sample to those likely impacted (defined as mother's education below college and family income below first quartile). In column (3), I restrict the sample to only those one year after they graduate high school (to avoid issues with migration). In column (4), I define state as the state where the individual was born instead of their current state. In column (5), I measure college two or more years after high school instead of one or more. In column (6), I define college attendance using the ACS variables "school" and "gradeattd" instead of "educd." Unless otherwise noted, college attendance is defined as an indicator for whether an individual in cohort t attended college one year or more after graduating high school. Cohort is defined as the year the individual would have been a senior using standard grade ages. Unless otherwise noted, sample is 17 to 47 year olds from the 2000-2018 ACS. Standard errors, clustered at the state level, in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure C.1: General Trends in College-Going and Test-Taking



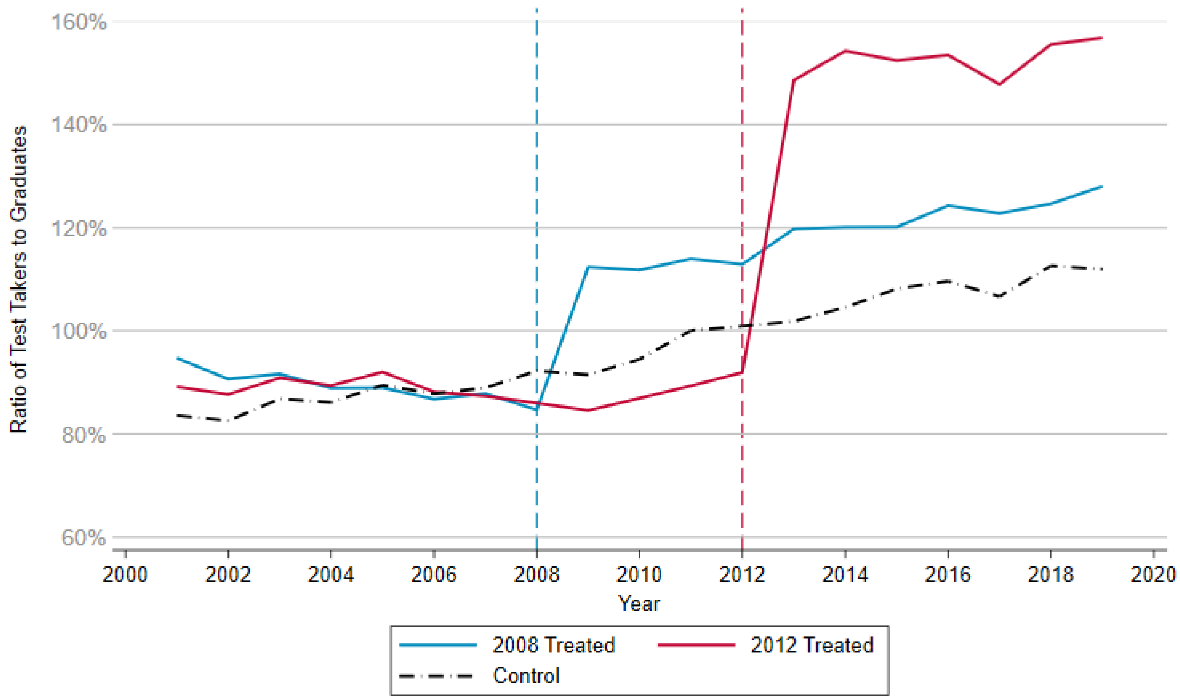
Note: This figure shows general trends in test taking and college attendance over the sample period. College attendance increases slowly, while ACT testing increases by a large amount and SAT testing remains flat for most of the sample, until the later years. The proportion of ACT and SAT testers is calculated as the number of testers divided by the number of graduates, where graduate counts come from the NCES and are projected after the 2012-2013 school year. Sample is 17-47 year olds from the 2000-2018 ACS.

Figure C.2: Time Variation in Policy Implementation



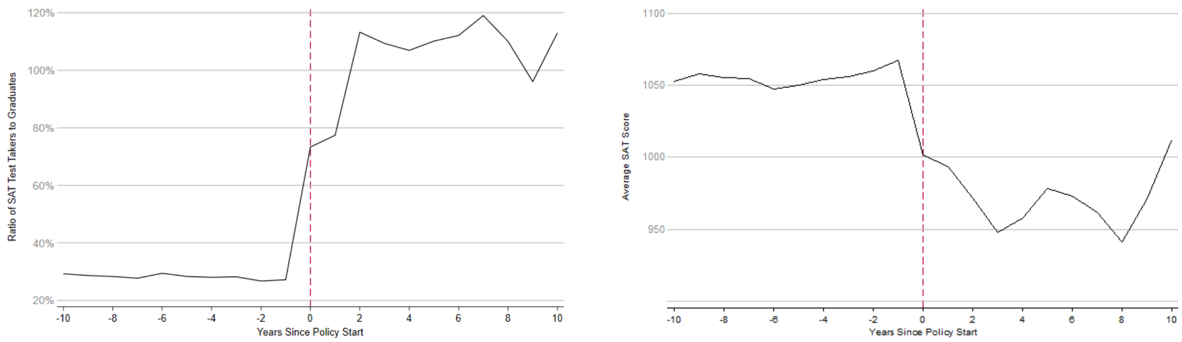
Note: This figure shows variation in policy implementation over time across states. The left graph shows the number of states who first implemented a policy in each year. The right graph shows the cumulative proportion of states with a policy in each year.

Figure C.3: Parallel Pre-Trends in Test Taking



Note: This figure shows test-taking trends in treatment and control states for the 2008 and 2012 policy-year groups. We can see parallel trends other than the stark jump in testers that occurs around the policy implementation. Control states are defined as those who never implement a mandatory college entrance exam. Treatment states in 2008: Kentucky and Wyoming. Treatment states in 2012: Louisiana, Idaho, and North Carolina. Control states are defined as those who never implemented a mandatory college entrance exam.

Figure C.4: Sharp Testing Changes in SAT Policy States



Note: This figure shows the large jump in SAT test takers and the large drop in average SAT scores in the year of mandatory SAT implementation for policy states. The proportion of SAT testers is calculated as the number of testers divided by the number of graduates, where graduate counts come from the NCES and are projected after the 2012-2013 school year. The percent of test takers can be greater than 100 because students are able to take the exam multiple times.

Bibliography

J. M. Abowd, B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock. The lehd infrastructure files and the creation of the quarterly workforce indicators. In *Producer Dynamics: New Evidence from Micro Data*, pages 149–230. University of Chicago Press, 2009.

M. E. Andresen and E. Nix. What causes the child penalty? evidence from adopting and same-sex couples. *Journal of labor economics*, 40(4):971–1004, 2022.

N. Angelov, P. Johansson, and E. Lindahl. Parenthood and the gender gap in pay. *Journal of labor economics*, 34(3):545–579, 2016.

H. Antecol and M. D. Steinberger. Labor supply differences between married heterosexual women and partnered lesbians: A semi-parametric decomposition approach. *Economic Inquiry*, 51(1):783–805, 2013.

W. Arbour. Can recidivism be prevented from behind bars? evidence from a behavioral program. *Working Papers tecipa-683*, 2021.

D. H. Autor and S. N. Houseman. Do temporary-help jobs improve labor market outcomes for low-skilled workers? evidence from "work first". *American Economic Journal: Applied Economics*, 2(3):96–128, 2010.

C. Avery and T. J. Kane. Student perceptions of college opportunities: The boston coach program. In *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, pages 355–394. University of Chicago Press, 2004.

M. L. Badgett, C. S. Carpenter, and D. Sansone. Lgbtq economics. *Journal of Economic Perspectives*, 35(2):141–170, 2021.

E. Bagalman and A. S. Cornell. *Prevalence of Mental Illness in the United States: Data Sources and Estimates*. Congressional Research Service, 2018.

G. Bauer. Gender roles, comparative advantages and the life course: the division of domestic labor in same-sex and different-sex couples. *European Journal of Population*, 32:99–128, 2016.

G. S. Becker. A theory of the allocation of time. *The economic journal*, 75(299):493–517, 1965.

G. S. Becker. *A treatise on the family: Enlarged edition*. Harvard University Press, 1991.

A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429, 2012.

E. P. Bettinger, B. T. Long, P. Oreopoulos, and L. Sanbonmatsu. The role of application assistance and information in college decisions: Results from the h&r block fafsa experiment. *The Quarterly Journal of Economics*, 127(3):1205–1242, 2012.

I. A. Binswanger, M. F. Stern, R. A. Deyo, P. J. Heagerty, A. Cheadle, J. G. Elmore, and T. D. Koepsell. Release from prison — a high risk of death for former inmates. *New England Journal of Medicine*, 356(2):157–165, 2007.

D. A. Black, S. G. Sanders, and L. J. Taylor. The economics of lesbian and gay families. *Journal of economic perspectives*, 21(2):53–70, 2007.

C. Blandhol, J. Bonney, M. Mogstad, and A. Torgovitsky. When is tsls actually late? *National Bureau of Economic Research Working Paper*, 2022.

C. Blattman, J. C. Jamison, and M. Sheridan. Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review*, 107(4):1165–1206, 2017.

C. Blattman, S. Chaskel, J. C. Jamison, and M. Sheridan. Cognitive behavior therapy reduces crime and violence over 10 years: Experimental evidence. *National Bureau of Economic Research Working Paper*, 2022.

BLS. Persons with a disability: Labor force characteristics news release. Technical report, Bureau of Labor Statistics, Washington, DC, 2014.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event study designs: Robust and efficient estimation. *forthcoming at Review of Economic Studies*, 2024.

W. G. Bowen, M. M. Chingos, and M. S. McPherson. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton University Press, 2009.

J. Breslau, M. Lane, N. Sampson, and R. C. Kessler. Mental disorders and subsequent educational attainment in a us national sample. *Journal of Psychiatric Research*, 42(9):708–716, 2008.

J. Bronson and M. Berzofsky. Indicators of mental health problems reported by prisoners and jail inmates, 2011–12. Technical report, Bureau of Justice Statistics, 2017.

M. Bubonya, D. Cobb-Clark, and M. Wooden. Mental health and productivity at work: Does what you do matter? *Labour Economics*, 46:150–165, 2017.

G. Bulman. The effect of access to college assessments on enrollment and attainment. *American Economic Journal: Applied Economics*, 7(4):1–36, 2015.

M. Burns and L. Dague. In-kind welfare benefits and reincarceration risk: Evidence from medicaid. *National Bureau of Economic Research Working Paper*, 2023.

W. Burton, A. Schultz, C.-Y. Chen, and D. Edington. The association of worker productivity and mental health: A review of the literature. *International Journal of Workplace Health Management*, 2008.

T. Byker. The opt-out continuation: Education, work, and motherhood from 1984 to 2012. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(4):34–70, 2016.

B. Callaway and P. H. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230, 2021.

P. Carneiro, J. J. Heckman, and E. J. Vytlacil. Estimating marginal returns to education. *American Economic Review*, 101(6):2754–2781, 2011.

CBO. Prescription drugs: Spending, use, and prices. Technical report, Congressional Budget Office, 2022.

Census. General population characteristics. Technical report, US Census Bureau, 1990.

Census. Profiles of general demographic characteristics. Technical report, US Census Bureau, May 2001.

K. Cherney. How much does depression cost? `https://www.healthline.com/health/depression/how-much-does-depression-cost`, 2020.

V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Y. Chung, B. Downs, D. H. Sandler, R. Sienkiewicz, et al. The parental gender earnings gap in the united states. Technical report, Center for Economics Studies (WP 17-68), 2017.

E. Ciscato, A. Galichon, and M. Goussé. Like attract like? a structural comparison of homogamy across same-sex and different-sex households. *Journal of Political Economy*, 128(2):740–781, 2020.

E. E. Cook and S. Turner. Missed exams and lost opportunities: Who could gain from expanded college admission testing? *AERA Open*, 5(2), 2019.

C. J. Cronin, M. P. Forsstrom, and N. W. Papageorge. What good are treatment effects without treatment? mental health and the reluctance to use talk therapy. *National Bureau of Economic Research Working Paper*, 2020.

C. De Chaisemartin and X. d'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996, 2020.

D. Deming. Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, 1(3):111–34, 2009.

E. P. Deschenes, S. Turner, and P. W. Greenwood. Drug court or probation? an experimental evaluation of maricopa county's drug court. *Justice System Journal*, 18(1):55–73, 1995.

N. DHHS. Cpt service rates for specialty 113. `https://medicaid.ncdhhs.gov/documents/providers/bulletins/archives/behavioral-health/other-bhs/mhfee1-010113/open`, 2013.

N. DHHS. Basic medicaid eligibility chart. `https://medicaid.ncdhhs.gov/documents/files/basic-medicaid-eligibility-chart-2016-09-19`, 2016.

E. W. Dillon and J. A. Smith. Determinants of the match between student ability and college quality. *Journal of Labor Economics*, 35(1):45–66, 2017.

W. Dobbie, J. Goldin, and C. S. Yang. The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2):201–240, 2018.

J. L. Doleac. Encouraging desistance from crime. *Journal of Economic Literature*, 61(2):383–427, 2023.

S. Dynarski, J. Hyman, and D. W. Schanzenbach. Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management*, 32(4):692–717, 2013.

L. L. Espinosa, J. M. Turk, M. Taylor, and H. M. Chessman. Race and ethnicity in higher education: A status report. Technical report, American Council on Education, 2019.

ESSA. Assessments under title i, part a and title i, part b: Summary of final regulations. Fact sheet, US Department of Education, December 2016.

S. Fazel, P. Lichtenstein, M. Grann, G. M. Goodwin, and N. Långström. Bipolar disorder and violent crime: New evidence from population-based longitudinal studies and systematic review. *Archives of General Psychiatry*, 67(9):931–938, 2010.

T. E. Feucht and J. Gfroerer. Mental and substance use disorders among adult men on probation or parole: Some success against a persistent challenge. Technical report, Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, 2011.

R. Fisher, G. Gee, and A. Looney. Same-sex married tax filers after windsor and obergefell. *Demography*, 55(4):1423–1446, 2018.

A. Foote, L. Schulkind, and T. M. Shapiro. Missed signals: The effect of act college-readiness measures on post-secondary decisions. *Economics of Education Review*, 46:39–51, 2015.

B. Frandsen, E. Leslie, and S. McIntyre. Cluster jackknife instrumental variables estimation. *Working Paper*, 2023.

R. G. Frank and T. G. McGuire. Mental health treatment and criminal justice outcomes. In *Controlling Crime: Strategies and Tradeoffs*, pages 167–207. University of Chicago Press, 2010.

GAO. Adult drug courts: Evidence indicates recidivism reductions and mixed results for other outcomes. Technical report, US Government Accountability Office, 2005.

A. Garin, D. Koustas, C. McPherson, S. Norris, M. Pecenco, E. K. Rose, Y. Shem-Tov, and J. Weaver. The impact of incarceration on employment, earnings, and tax filing. Technical report, University of Chicago, Becker Friedman Institute for Economics Working Paper 2023-108, 2023.

G. J. Gates. Same-sex couples in us census bureau data: Who gets counted and why. Technical report, The Williams Institute, 2010.

K. R. Genadek, S. M. Flood, and J. G. Roman. Same-sex couples' shared time in the united states. *Demography*, 57(2):475–500, 2020.

L. Giddings, J. M. Nunley, A. Schneebaum, and J. Zietz. Birth cohort and the specialization gap between same-sex and different-sex couples. *Demography*, 51(2):509–534, 2014.

R. Glauber. Trends in the motherhood wage penalty and fatherhood wage premium for low, middle, and high earners. *Demography*, 55(5):1663–1680, 2018.

L. E. Glaze and T. P. Bonczar. Probation and parole in the united states, 2009. Technical report, Bureau of Justice Statistics, US Department of Justice, Washington, DC, 2010.

L. E. Glaze and L. M. Maruschak. Parents in prison and their minor children. Technical report, Bureau of Justice Statistics, US Department of Justice, Washington, DC, 2008.

S. Goodman. Learning from the test: Raising selective college enrollment by providing information. *Review of Economics and Statistics*, 98(4):671–684, 2016.

A. Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.

D. C. Gottfredson, S. S. Najaka, and B. Kearley. Effectiveness of drug treatment courts: Evidence from a randomized trial. *Criminology & Public Policy*, 2(2):171–196, 2003.

P. A. Grady and L. L. Gough. Self-management: A comprehensive approach to management of chronic conditions. *American Journal of Public Health*, 104(8):e25–e31, 2014.

E. Green. Mental illness and violence: Is there a link? Technical report, Illinois Criminal Justice Information Authority, 2020.

J. A. Groen, M. J. Kutzbach, and A. E. Polivka. Storms and jobs: The effect of hurricanes on individuals' employment and earnings over the long term. *Journal of Labor Economics*, 38(3): 653–685, 2020.

J. Guydish, M. Chan, A. Bostrom, M. A. Jessup, T. B. Davis, and C. Marsh. A randomized trial of probation case management for drug-involved women offenders. *Crime & Delinquency*, 57 (2):167–198, 2011.

T. Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica, Journal of the Econometric Society*, pages 1–12, 1943.

T. Haavelmo. The probability approach in econometrics. *Econometrica: Journal of the Econometric Society*, pages iii–115, 1944.

C. Hakulinen, P. Böckerman, L. Pulkki-Råback, M. Virtanen, and M. Elovainio. Employment and earnings trajectories before and after sickness absence due to major depressive disorder: A nationwide case–control study. *Occupational and Environmental Medicine*, 78(3):173–178, 2021.

E. A. Hall, M. L. Prendergast, and U. Warda. A randomized trial of the effectiveness of using incentives to reinforce parolee attendance in community addiction treatment: Impact on post-treatment outcomes. *Journal of Drug Issues*, 47(2):148–163, 2017.

M. E. Hansen, M. E. Martell, and L. Roncolato. A labor of love: The impact of same-sex marriage on labor supply. *Review of Economics of the Household*, 18:265–283, 2020.

P. Heaton. Hidden in plain sight: What cost-of-crime research can tell us about investing in police. *RAND Occasional Paper*, 2010.

J. Heckman and R. Pinto. Causal analysis after haavelmo. *Econometric Theory*, 31(1):115–151, 2015.

J. J. Heckman and R. Pinto. Unordered monotonicity. *Econometrica*, 86(1):1–35, 2018.

S. B. Heller, A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, and H. A. Pollack. Thinking, fast and slow? some field experiments to reduce crime and dropout in chicago. *The Quarterly Journal of Economics*, 132(1):1–54, 2017.

N. Hendren and B. Sprung-Keyser. A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3):1209–1318, 2020.

V. A. Hiday, M. S. Swartz, J. W. Swanson, R. Borum, and H. R. Wagner. Impact of outpatient commitment on victimization of people with severe mental illness. *American Journal of Psychiatry*, 159(8):1403–1411, 2002.

T. Hofmarcher and E. Plug. Specialization in same-sex and different-sex couples. *Labour Economics*, 77:101995, 2022.

J. L. Hotchkiss, M. M. Pitts, and M. B. Walker. Impact of first birth career interruption on earnings: evidence from administrative data. *Applied Economics*, 49(35):3509–3522, 2017.

C. Hoxby and S. Turner. What high-achieving low-income students know about college. *American Economic Review*, 105(5):514–17, 2015.

C. M. Hoxby and C. Avery. Low-income high-achieving students miss out on attending selective colleges. *Brookings Papers on Economic Activity, Spring*, 10, 2013.

J. E. Humphries, A. Ouss, K. Stavreva, M. T. Stevenson, and W. van Dijk. Conviction, incarceration, and recidivism: Understanding the revolving door. *Working Paper (SSRN 4507597)*, 2023.

P. Hunt, J. Anderson, and J. Saunders. The price of justice: New national and state-level estimates of the judicial and legal costs of crime to taxpayers. *American Journal of Criminal Justice*, 42 (2):231–254, 2017.

M. Hurwitz, J. Smith, S. Niu, and J. Howell. The maine question: How is 4-year college enrollment affected by mandatory college entrance exams? *Educational Evaluation and Policy Analysis*, 37(1):138–159, 2015.

J. Hyman. Act for all: The effect of mandatory college entrance exams on postsecondary attainment and choice. *Education Finance and Policy*, 12(3):281–311, 2017.

G. W. Imbens and J. D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

E. Jácome. Mental health and criminal involvement: Evidence from losing medicaid eligibility. *Princeton University Working Paper*, 2020.

C. Jepsen and L. K. Jepsen. Labor-market specialization within same-sex and difference-sex couples. *Industrial Relations: A Journal of Economy and Society*, 54(1):109–130, 2015.

E. Kalish. Foster family characteristics and children's educational outcomes. *Working Paper*, 2023.

A. Killewald. A reconsideration of the fatherhood premium: Marriage, coresidence, biology, and fathers' wages. *American sociological review*, 78(1):96–116, 2013.

D. Klasik. The act of enrollment: The college enrollment effects of state-required college entrance exam testing. *Educational Researcher*, 42(3):151–160, 2013.

F. Kleibergen and R. Paap. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126, 2006.

H. Kleven. The geography of child penalties and gender norms: Evidence from the united states. Technical report, National Bureau of Economic Research, 2022.

H. Kleven, C. Landais, and J. E. Søgaard. Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, 11(4):181–209, 2019.

H. Kleven, C. Landais, and J. E. Søgaard. Does biology drive child penalties? evidence from biological and adoptive families. *American Economic Review: Insights*, 3(2):183–198, 2021.

J. R. Kling. Incarceration length, employment, and earnings. *American Economic Review*, 96(3): 863–876, 2006.

R. M. Kreider, N. Bates, and Y. Mayol-García. Improving measurement of same-sex couple households in census bureau surveys: Results from recent tests. In *PAA 2017 Annual Meeting*. PAA, 2017.

L. A. Kurdek. The allocation of household labor in gay, lesbian, and heterosexual married couples. *Journal of Social Issues*, 49(3):127–139, 1993.

Kurtz-Blum. Criminal defense trials in north carolina: What you can expect when you go to trial. `https://www.kurtzandblum.com/criminal-defense-attorneys/trials/`, 2021. Accessed: 2021-03-24.

I. Kuziemko, J. Pan, J. Shen, and E. Washington. The mommy effect: Do women anticipate the employment effects of motherhood? Technical report, National Bureau of Economic Research, 2018.

L. Larson. Federal and state testing requirements for k-12 public school students. Information brief, Minnesota House of Representatives, November 2002.

R. Layard. Mental health: The new frontier for labour economics. *IZA Journal of Labor Policy*, 2(1):1–16, 2013.

S. Lee and B. Salanié. Identifying effects of multivalued treatments. *Econometrica*, 86(6):1939–1963, 2018.

O. T. Lenning and E. J. Maxey. Act versus sat prediction for present-day colleges and students. *Educational and Psychological Measurement*, 33(2):397–406, 1973.

E. Leslie and N. G. Pope. The unintended impact of pretrial detention on case outcomes: Evidence from new york city arraignments. *The Journal of Law and Economics*, 60(3):529–557, 2017.

J. E. Louden and J. L. Skeem. Parolees with mental disorder: Toward evidence-based practice. *Bulletin of the Center for Evidence-Based Corrections*, 7(1):1–9, 2011.

G. Mari. Is there a fatherhood wage premium? a reassessment in societies with strong male-breadwinner legacies. *Journal of Marriage and Family*, 81(5):1033–1052, 2019.

M. E. Martell and P. Nash. For love and money? earnings and marriage among same-sex couples. *Journal of Labor Research*, 41(3):260–294, 2020.

K. E. McCollister, M. T. French, and H. Fang. The cost of crime to society: New crime-specific estimates for policy and program evaluation. *Drug and Alcohol Dependence*, 108(1-2):98–109, 2010.

D. K. McLaughlin and C. S. Stokes. Income inequality and mortality in us counties: Does minority racial concentration matter? *American Journal of Public Health*, 92(1):99–104, 2002.

M. McLaughlin, C. Pettus-Davis, D. Brown, C. Veeh, and T. Renn. The economic burden of incarceration in the us. *Institute for Advancing Justice Research and Innovation Working Paper*, 2016.

C. Mitton, L. Simpson, L. Gardner, F. Barnes, and G. McDougall. Calgary diversion program: A community-based alternative to incarceration for mentally ill offenders. *The Journal of Mental Health Policy and Economics*, 10(3):145–151, 2007.

Y. Moberg. Does the gender composition in couples matter for the division of labor after childbirth? Technical report, Working Paper, 2016.

M. Mogstad, A. Torgovitsky, and C. R. Walters. The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review*, 111(11):3663–3698, 2021.

R. Mojtabai, E. A. Stuart, I. Hwang, W. W. Eaton, N. Sampson, and R. C. Kessler. Long-term effects of mental disorders on educational attainment in the national comorbidity survey ten-year follow-up. *Social Psychiatry and Psychiatric Epidemiology*, 50:1577–1591, 2015.

J. Moody. Act vs. sat: How to decide which test to take. Technical report, US News, June 2020.

J. Mountjoy. Community colleges and upward mobility. *American Economic Review*, 112(8): 2580–2630, 2022.

M. Mueller-Smith. The criminal and labor market impacts of incarceration. *American Economic Review (Revise and Resubmit)*, 2023.

C. Neumeier, T. Sørensen, and D. Webber. The implicit costs of motherhood over the lifecycle: Cross-cohort evidence from administrative longitudinal data. *Southern Economic Journal*, 84 (3):716–733, 2018.

NIMH. Figure 1: Suicide rates in the united states (2000-2020). `https://www.nimh.nih.gov/health/statistics/suicide`, 2023. data courtesy of CDC's National Center for Health Statistics.

I. Pal and J. Waldfogel. The family gap in pay: New evidence for 1967 to 2013. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(4):104–127, 2016.

A. Pallais and S. Turner. Opportunities for low-income students at top colleges and universities: Policy initiatives and the distribution of students. *National Tax Journal*, 59(2):357–386, 2006.

V. Patel and A. Kleinman. Poverty and common mental disorders in developing countries. *Bulletin of the World Health Organization*, 81:609–615, 2003.

E. S. Paykel, J. Scott, J. D. Teasdale, A. L. Johnson, A. Garland, R. Moore, A. Jenaway, P. L. Cornwall, H. Hayhurst, R. Abbott, and M. Pope. Prevention of relapse in residual depression by cognitive therapy: A controlled trial. *Archives of General Psychiatry*, 56(9):829–835, 1999.

J. Pearl. Trygve haavelmo and the emergence of causal calculus. *Econometric Theory*, 31(1): 152–179, 2015.

F. S. Pearson, D. S. Lipton, C. M. Cleland, and D. S. Yee. The effects of behavioral/cognitive-behavioral programs on recidivism. *Crime & Delinquency*, 48(3):476–496, 2002.

M. L. Prendergast, E. A. Hall, J. Grossman, R. Veliz, L. Gregorio, U. S. Warda, K. Van Unen, and C. Knight. Effectiveness of using incentives to improve parolee admission and attendance in community addiction treatment. *Criminal Justice and Behavior*, 42(10):1008–1031, 2015.

K. C. Prickett, A. Martin-Storey, and R. Crosnoe. A research note on time with children in different-and same-sex two-parent families. *Demography*, 52(3):905–918, 2015.

C. Prins, K. Officer, E. L. Einspruch, K. L. Jarvis, M. S. Waller, J. R. Mackin, and S. M. Carey. Randomized controlled trial of measure 57 intensive drug court for medium-to high-risk property offenders: Process, interviews, costs, and outcomes. Technical report, Technical Report, Oregon Criminal Justice Commission and NPC Research, 2015.

M. A. Project. Equality maps: Foster and adoption laws. `https://www.lgbtmap.org/equality-maps/foster_and_adoption_laws`. Accessed: 2022-07-05.

M. Rafael. The burden of court debt on washingtonians. Technical report, Vera Institute of Justice, 2023.

S. F. Reardon. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, 1(1):91–116, 2011.

M. Ridley, G. Rao, F. Schilbach, and V. Patel. Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science*, 370(6522), 2020.

E. Saez, J. Slemrod, and S. H. Giertz. The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, 50(1):3–50, 2012.

SAMHSA. Key substance use and mental health indicators in the united states: Results from the 2021 national survey on drug use and health. NSDUH Series H-57 HHS Publication No. PEP22-07-01-005, Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, 2022.

D. Sandler, N. Szembrot, et al. Maternal labor dynamics: participation, earnings, and employer changes. Technical report, Center for Economics Studies (WP 19-33), 2019.

C. K. Scott and M. L. Dennis. The first 90 days following release from jail: Findings from the recovery management checkups for women offenders (rmcwo) experiment. *Drug and Alcohol Dependence*, 125(1-2):110–118, 2012.

B. S. Silveira. Bargaining with asymmetric information: An empirical study of plea negotiations. *Econometrica*, 85(2):419–452, 2017.

F. Sloan, A. Platt, L. Chepke, and C. Blevins. Deterring domestic violence: Do criminal sanctions reduce repeat offenses? *Journal of Risk and Uncertainty*, 46(1):51–80, 2013.

F. Sloan, L. Eldred, S. McCutchan, and A. Platt. Deterring rearrests for drinking and driving. *Southern Economic Journal*, 83(2):416–436, 2016.

B. Sommers, R. Kronick, K. Finegold, R. Po, K. Schwartz, and S. Glied. Understanding participation rates in medicaid: Implications for the affordable care act. *Public Health*, 93(1):67–74, 2012.

L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199, 2021.

K. C. D. C. The Annie E. Casey Foundation. Median family income by race and ethnicity (5-year average) in washington. `https://datacenter.aecf.org/data/tables/4682-median-family-income-by-race-and-ethnicity-5-year-average#detailed/2/any/false/1983,1692,1691,1607,1572,1485,1376,1201,1074,880/437,172,133,12,4100,826,816,13,438/10944`, 2023.

E. Vytlacil. Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1):331–341, 2002.

L. Walker and D. Taylor. Same-sex couple households, 2019. Technical report, US Census Bureau, Report Number ACSBR-005, 2021.

L. Wang. Chronic punishment: The unmet health needs of people in state prisons. Technical report, Prison Policy Initiative, 2022.

K. A. Weeden, Y. Cha, and M. Bucca. Long work hours, part-time work, and trends in the gender gap in pay, the motherhood wage penalty, and the fatherhood wage premium. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(4):71–102, 2016.

H. C. West, W. J. Sabol, and S. J. Greenman. Prisoners in 2009. Technical report, Bureau of Justice Statistics, US Department of Justice, Washington, DC, 2010.

J. M. Wooldridge. Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Available at SSRN 3906345*, 2021.

L. E. Wüsthoff, H. Waal, and R. W. Gråwe. The effectiveness of integrated treatment in patients with substance use disorders co-occurring with anxiety and/or depression - a group randomized trial. *BMC Psychiatry*, 14(1):1–12, 2014.

W.-h. Yu and Y. Hara. Motherhood penalties and fatherhood premiums: effects of parenthood on earnings growth within and across firms. *Demography*, 58(1):247–272, 2021.

V. Ziemek. The effects of mental health treatment in correctional facilities. *Undergraduate Research*, 2017.