THE UNIVERSITY OF CHICAGO

THE LONG-RUN EFFECTS OF PROGRAM INTERACTIONS IN THE SOCIAL SAFETY NET: EVIDENCE FROM FOOD STAMPS AND HEAD START

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
MICHELE CARTER

CHICAGO, ILLINOIS
JUNE 2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to my advisors, Manasi Deshpande, Derek Neal, and Alessandra Voena, for their guidance, patience, and encouragement. I also wish to thank Joel Carter, Sharon Carter, Amanda Carter, Jason Carter, Katie Carter, Gabriel Carter, Theo Carter, and Jacob Jaffe for their love and support.

ABSTRACT

This dissertation contains two research papers.

In Chapter 2, I estimate the effect of program interactions between Food Stamps and Head Start during childhood on labor market and health outcomes in adulthood. Using a nationally representative panel data set and exploiting variation in the exact timing of program introduction for Food Stamps and Head Start across counties during the 1960s and 1970s, I compare outcomes for children who were differentially exposed to each program during childhood by virtue of the county and year in which they were born. Estimates from event study and difference-in-differences models indicate that (1) accessing Food Stamps during childhood improves labor market and health outcomes in adulthood and (2) Head Start availability at preschool age improves health outcomes in adulthood. However, I find no evidence that the long-run effect of accessing Food Stamps during childhood depends on the availability of Head Start at preschool age.

In Chapter 3, I study the effect of Head Start on children's long-run outcomes. Jackson and Johnson (2019) exploit quasi-random variation in the timing of Head Start introduction across counties to estimate the effect of preschool-age Head Start spending on outcomes in adulthood. Their event study estimates imply that, for poor children who reside in high Head Start spending counties, preschool-age exposure to Head Start increases educational attainment. I perform a replication exercise using Jackson and Johnson's analytical sample, and I am unable to reproduce these event study estimates. Instead, I find no statistically significant effect of preschool-age exposure to Head Start on poor children's educational attainment, in either high or low Head Start spending counties.

CHAPTER 1

INTRODUCTION

In this dissertation, I use quasi-experimental methods to identify the causal effect of childhood access to the social safety net on labor market and health outcomes in adulthood. In the United States, the social safety net consists of a wide array of programs that provide, among other benefits, human capital investments to children from participating households. Eligibility requirements for safety net programs are typically set to target benefits to low-income households and, as a result, many children participate in more than one safety net program during childhood. Determining optimal policy in a setting where participation in multiple programs is common requires two steps: first, identifying the impact of each individual safety net program on children's long-run outcomes and, second, identifying how safety net programs interact with each other to affect children's long-run outcomes. I carry out each of these steps in the context of two central safety net programs in the United States, Food Stamps and Head Start.

This dissertation proceeds as follows. In Chapter 2, I ask how childhood access to Food Stamps and Head Start, respectively, interact to affect labor market and health outcomes in adulthood. I find no evidence that the return to accessing Food Stamps during childhood depends on the availability of Head Start at preschool-age. In Chapter 3, I review a set of estimates from Jackson and Johnson (2019) for the individual program effect of Head Start on educational attainment in adulthood. I perform a replication exercise where I show that I am unable to reproduce Jackson and Johnson's estimates.

CHAPTER 2

THE LONG-RUN EFFECTS OF PROGRAM INTERACTIONS IN THE SOCIAL SAFETY NET: EVIDENCE FROM FOOD STAMPS AND HEAD START

2.1 Introduction

In the United States, safety net programs are an important source of human capital investment in children. The social safety net - comprised of programs like Food Stamps, Head Start, Medicaid, Supplemental Security Income and WIC - provides children in disadvantaged households with access to a variety of economic, health, nutritional and educational benefits[1]. Because eligibility requirements are similar across programs, many children participate in multiple safety net programs during the course of their childhoods. However, existing studies tend to consider the impact of each safety net program on children's outcomes in isolation. These studies typically find that exposure to individual safety net programs during childhood improves labor market and health outcomes in adulthood[2]. Less is known about how safety net programs accessed at different points during childhood may interact to affect children's outcomes in the long run.

In this paper, I study program interactions between two major components of the social safety net: the Food Stamps program and the Head Start program. The Food Stamps program provides low income households with vouchers that can be redeemed for food, while the Head Start program offers early childhood education services to children from low income households. For each stage of childhood τ, I ask whether the impact from accessing Food

1. The Food Stamps program is also known as the Supplemental Nutrition Assistance Program. WIC is also known as the Special Supplemental Nutrition Program for Women, Infants, and Children.

2. Examples include Ludwig and Miller (2007) for Head Start; Brown, Kowalski and Lurie (2015) for Medicaid; Aizer et. al. (2016) for Mothers' Pension program; Hoynes, Schanzenbach and Almond (2016) for Food Stamps; Goodman-Bacon (2018) for Medicaid; Jackson and Johnson (2019) for Head Start.

Stamps beginning at stage τ on labor market and health outcomes in adulthood depends on the availability of Head Start at preschool age. To answer this question, I exploit exogenous variation in access to each program during childhood, induced by the independent, overlapping, county-level roll-outs of Food Stamps and Head Start during the 1960s and 1970s. Children born during this time period were differentially exposed to each program by virtue of the county and year in which they were born. I estimate event study and difference-in-differences models based on this variation using a nationally representative panel data set that links affected children to their outcomes in adulthood. These models identify causal interaction effects between Food Stamps and Head Start by comparing long-run outcomes for children who were quasi-randomly exposed to one, both, or neither program at each stage of childhood.

I produce two sets of empirical results. In the first set of results, I estimate the effect of childhood exposure to Food Stamps and Head Start, respectively, on labor market and health outcomes in adulthood, ignoring any potential interaction effects between the two programs. This paper builds directly on two prior studies that use the same data set and a similar program roll-out based empirical strategy to estimate the individual program effects of Food Stamps and Head Start. Hoynes, Schanzenbach and Almond (2016) find that childhood exposure to Food Stamps improves labor market and health outcomes in adulthood, while Jackson and Johnson (2019) find that preschool-age exposure to Head Start improves labor market outcomes in adulthood. I compare my estimates to the estimates reported in these papers. The purpose of this exercise is twofold: to establish the credibility of the results reported in the remainder of the paper by identifying the source of any discrepancy between my findings and those reported in prior literature, and to summarize the individual program effects of Food Stamps and Head Start before taking into account interaction effects between the two programs.

In the second set of results, I estimate the interaction effect between childhood exposure to Food Stamps and preschool-age exposure to Head Start on labor market and health outcomes in adulthood. The interaction effect between Food Stamps and Head Start measures the extent to which the long-run return to Food Stamps is higher (or lower, if negative) for children with access to Head Start at preschool age, compared to children with no access to Head Start at preschool age. In other words, the interaction effect captures the degree of complementarity between Food Stamps and Head Start, which may vary depending on the stage of childhood at which exposure to Food Stamps begins.

The key identifying assumption for the event study and difference-in-differences models in this paper is that, in the absence of Food Stamps and Head Start, outcomes would have evolved similarly across birth cohorts between children who are exposed to both programs and children who are exposed to only one or neither program. For this assumption to hold, two conditions must be satisfied. First, the timing of program introduction for Food Stamps and Head Start should be approximately randomly assigned across counties. Second, the timing of program introduction for Food Stamps cannot be perfectly correlated with the timing of program introduction for Head Start. To address potential concerns about the validity of the first condition, I include a county of birth fixed effect and a set of time-varying county of birth characteristics in my empirical models. To establish that the second condition is satisfied, I show that the raw and regression-adjusted correlations between the timing of program introduction for Food Stamps and the timing of program introduction for Head Start are less than one in absolute value.

I produce my results using a sample of children born between 1950 and 1976 whose parents did not graduate high school. The outcome variables of interest in my analysis include income, employment status, educational attainment, and health conditions experienced during adulthood (including high blood pressure, diabetes, heart disease, and obesity). Because I consider a wide range of outcome variables in my analysis, I follow the approach used in Hoynes, Schanzenbach and Almond (2016), and I construct an economic self-sufficiency index to summarize labor market outcomes in adulthood and a metabolic syndrome index to summarize health outcomes in adulthood.

I find that exposure to Food Stamps during childhood improves labor market and health outcomes in adulthood. This result is consistent with the findings in Hoynes, Schanzenbach and Almond (2016). Children who are fully exposed to Food Stamps between the in-utero period and age eleven experience a 0.482 standard deviation increase in the economic self-sufficiency index and a statistically insignificant 0.107 standard deviation reduction in the metabolic syndrome index, relative to children with no exposure to Food Stamps over the same age range[3]. These composite effects on the economic self-sufficiency index and the metabolic syndrome index are driven, in part, by a 25 percentage point (29 percent, relative to the sample mean) increase in the likelihood of graduating high school and a 24 percentage point (67 percent, relative to the sample mean) reduction in the incidence of obesity among children who are fully exposed to Food Stamps between the in-utero period and age eleven.

I find that preschool-age exposure to Head Start improves health outcomes and negatively affects labor market outcomes in adulthood. Children who are fully exposed to Head Start between age three and age six experience a 0.15 standard deviation reduction in the economic self-sufficiency index and a statistically insignificant 0.13 standard deviation reduction in the metabolic syndrome index, compared to children with no exposure to Head Start over the same age range. These composite effects on the economic self-sufficiency index and the metabolic syndrome index result, in part, from a statistically insignificant 4 percentage point (5 percent, relative to the sample mean) decrease in the likelihood of graduating high school and a 16 percentage point (43 percent, relative to the sample mean) reduction in the incidence of obesity among children who are fully exposed to Head Start between age three and age six.

Lastly, I extend my analysis to allow for interaction effects between Food Stamps and Head Start. My results do not indicate that the long-run return to Food Stamps beginning at any stage of childhood depends on the availability of Head Start at preschool age. Event

3. For the remaining discussion, "statistically insignificant" means statistically insignificant at a 10 percent significance level.

study and difference-in-differences estimates show that the differential effect of exposure to Food Stamps beginning at age τ on labor market and health outcomes (as measured by the economic self-sufficiency index and the metabolic syndrome index, respectively) between children with any preschool-age exposure to Head Start and children with no preschool-age exposure to Head Start is not statistically different from zero, for any age $\tau < 12$. It is important to note, however, that my estimates are statistically insignificant at a 10% level, and I am unable to rule out sizable, non-zero interaction effects between Food Stamps and Head Start with any reasonable degree of certainty.

2.1.1 Contribution to the Literature

This paper contributes to three strands of the economics literature.

First, my work is related to an active literature that examines the long-run effects of early childhood exposure to the social safety net. The existing work on this topic tends to focus on the long-run effects of individual safety net programs accessed during childhood, and does not consider how the return to one program may vary depending on the availability of another program. Safety net programs evaluated in this manner include Medicaid (Brown, Kowalski and Lurie 2019; Goodman-Bacon 2017), the Earned Income Tax Credit (Bastian and Michelmore 2018), cash welfare (Aizer et. al. 2016), Food Stamps (Hoynes, Schanzenbach and Almond 2016; Bailey et. al. 2019), and Head Start (Ludwig and Miller 2007). These papers typically find positive impacts of childhood exposure to individual safety net programs on labor market and health outcomes in adulthood.

Unlike most of the previous research on this topic, this paper studies interactions between safety net programs accessed during childhood. I ask how the long-run return to Food Stamps exposure during childhood depends on the availability of Head Start at preschool age. My work is closely related to two recent papers. Jackson and Johnson (2019) find that the return to Head Start at age four is higher for children with greater exposure to school finance reform-induced increases in school spending between ages five and eighteen.

Bailey et. al (2018) estimate the long-run return to Head Start exposure at preschool age. In a set of supplemental results, they show that the return to preschool-age Head Start exposure is positively related to an aggregate measure of Medicaid exposure between ages zero and eighteen and negatively related to an aggregate measure of Food Stamps exposure between ages zero and eighteen. I expand on the existing literature by documenting how the interaction effects between Food Stamps and Head Start vary dynamically depending on the age at which exposure to each program begins, showing that the interaction effects are likely to be causally identified in my setting, and considering a wide range of outcome variables in my analysis (including health).

Second, my work contributes to a recent empirical literature that studies how a broader set of human capital investments (or shocks) received during childhood interact to affect outcomes later in life (e.g., Malamud et. al. 2016; Adhvaryu et. al. 2018; Rossin-Slater and Wust 2019; Gunnsteinsson et. al. 2019). The scarcity of work on this topic can be attributed to the difficulty of (1) finding valid, overlapping identification strategies for two human capital investments (Almond and Mazumder 2013), and (2) linking these identification strategies to a data set that contains information about individuals' childhood circumstances and adult outcomes. In the existing literature, estimates for the interaction effect between human capital investments made during childhood vary in sign and magnitude across different settings. For example, Malamud et. al. (2016) find no evidence of an interaction effect between access to abortion and school quality on human capital in Romania. In contrast, Rossin-Slater and Wust (2019) estimate a negative interaction effect between a nurse home visiting program and a preschool program on human capital in Denmark. This heterogeneity in the estimated interaction effect between human capital investments made during childhood across different settings could be explained by several factors, including differences in the underlying sample population or differences in the timing and type of human capital investments undertaken. Some combination of these factors could explain why I find no evidence of causal interaction effects between Food Stamps and Head Start, while other papers estimate sizable interaction effects between human capital investments made during early childhood.

Third, my work is related to a theoretical literature that hypothesizes dynamic complementarities in skill formation. The process of skill formation exhibits dynamic complementarities if skills acquired in previous time periods raise the productivity of skill investments made in subsequent time periods (Heckman and Cunha 2007; Heckman and Mosso 2014). A positive relationship between the return to a later-period skill investment and the receipt of an earlier-period skill investment implies dynamic complementarities in skill formation, only if there are no additional investment responses (e.g., by parents) to the receipt of the earlier-period investment. In my setting, I do not observe the complete set of public and private investments that each child receives. As a result, I cannot "rule out" or "rule in" dynamic complementarities in skill formation based on the sign of the interaction effect between Food Stamps and Head Start that I estimate. Instead, I provide evidence on the degree of reduced-form complementarity between Food Stamps and Head Start. My results do not indicate that, in a reduced-form sense, Food Stamps and Head Start are either complements or substitutes in the production of human capital.

2.1.2 Organization

This paper proceeds as follows. Section 2 provides historical background on the two programs studied in this paper, Food Stamps and Head Start. Section 3 describes the data set and my empirical strategy. Section 4 presents my estimates for the individual program effects of childhood exposure to Food Stamps and Head Start, and compares these estimates to similar estimates reported in the literature. Section 5 presents my estimates for the interaction effects between childhood exposure to Food Stamps and Head Start. Section 6 concludes.

2.2 Background

2.2.1 Background on Food Stamps Program

The Food Stamps program (also known as the Supplemental Nutrition Assistance Program, or SNAP) is a federally-funded, means-tested nutritional assistance program. Participating households receive vouchers that can be used to purchase eligible food items from authorized stores[4]. To qualify for the program, households must satisfy income and asset limits[5]. The average monthly benefit from Food Stamps for an individual was $126.78 in 2018 (USDA 2019).

Previous research has found that the amount most households spend on food prior to receiving Food Stamps exceeds the value of the Food Stamps benefits they receive once they begin participating in the program. Providing in-kind Food Stamps benefits to these inframarginal households has the same effect as providing a cash transfer of equivalent value, to the extent that households treat income from different sources as fungible[6]. Importantly for interpreting the estimates in the sections that follow, this implies that Food Stamps may impact children's outcomes through channels other than nutrition - for example, due to the changes in household consumption and labor supply that are associated with the receipt of a cash transfer.

4. Today, vouchers are delivered electronically to participants via debit-like electronic benefit transfer cards. At the program's inception, vouchers were issued as physical paper coupons. Items that cannot be purchased with Food Stamps vouchers include non-food items, prepared foods, alcohol, and tobacco products.

5. For example, in 2019, gross monthly income could not exceed 130 percent of the poverty line and net monthly income (after taking into account allowable deductions) could not exceed 100 percent of the poverty line (USDA 2019).

6. For evidence on the impact of Food Stamps receipt on food consumption, see Hoynes and Schanzenbach (2009). For a discussion of the fungibility of Food Stamps benefits, see Hastings and Shapiro (2018).

2.2.2 Background on Head Start Program

The Head Start program is an early childhood education program that operates with the stated goal of promoting school readiness among children from low income households. The program serves children between the ages of three and six, and in recent years, eighty percent of enrollees were between the ages of three and four[7]. To enroll in Head Start, children must belong to a household whose income falls below a maximum allowable threshold that varies across Head Start centers[8]. Head Start enrollment for children between the ages of three and six totaled 750,000 in 2018 (HHS 2019).

Head Start provides a wide range of benefits to enrolled children and their families. Children who attend Head Start receive educational services, health screenings, and daily meals, and parents of Head Start participants are encouraged to engage with their children outside of the program. As a result, the Head Start program can be understood as a bundle of educational, health, nutritional, and parental investments provided to participating children (and, to the extent that investments spill over within families, to their siblings).

2.2.3 Introduction of Food Stamps and Head Start

President Lyndon B. Johnson declared a War on Poverty in 1964 that, in subsequent years, resulted in the dramatic expansion of the social safety net in the United States. Between 1962 and 1980, spending on social safety net programs nearly doubled as a share of federal outlays[9] (OMB 2018). Many of the current programs that serve the poor and the elderly were introduced during this time period, including Food Stamps, Head Start, Medicare, Medicaid,

7. The introduction of the Early Head Start program in 1995 expanded the program to include children between the ages of zero and three.

8. Typically, the maximum allowable income threshold is 100 to 130 percent of the federal poverty line.

9. For this tabulation, spending on social safety net programs includes the following spending categories: unemployment compensation, housing assistance, food and nutrition assistance, other income security spending, health care services (excluding Medicare), and health research and training (excluding Medicare). Spending rose from 6.9 percent to 12.6 percent of federal outlays.

Supplemental Security Income (SSI) and Supplemental Nutrition for Women, Infants and Children (WIC).

The War on Poverty was a unique effort that encouraged local-level organizations and city- and county-level government agencies to implement and operate federally-funded anti-poverty programs. This approach was partly an attempt to tailor anti-poverty programs to the specific needs of local communities[10]. Food Stamps and Head Start are two examples of programs that fit this pattern. At the time that Food Stamps and Head Start were introduced, the Food Stamps program was administered by government agencies at the county level, while Head Start centers were operated by local community organizations.

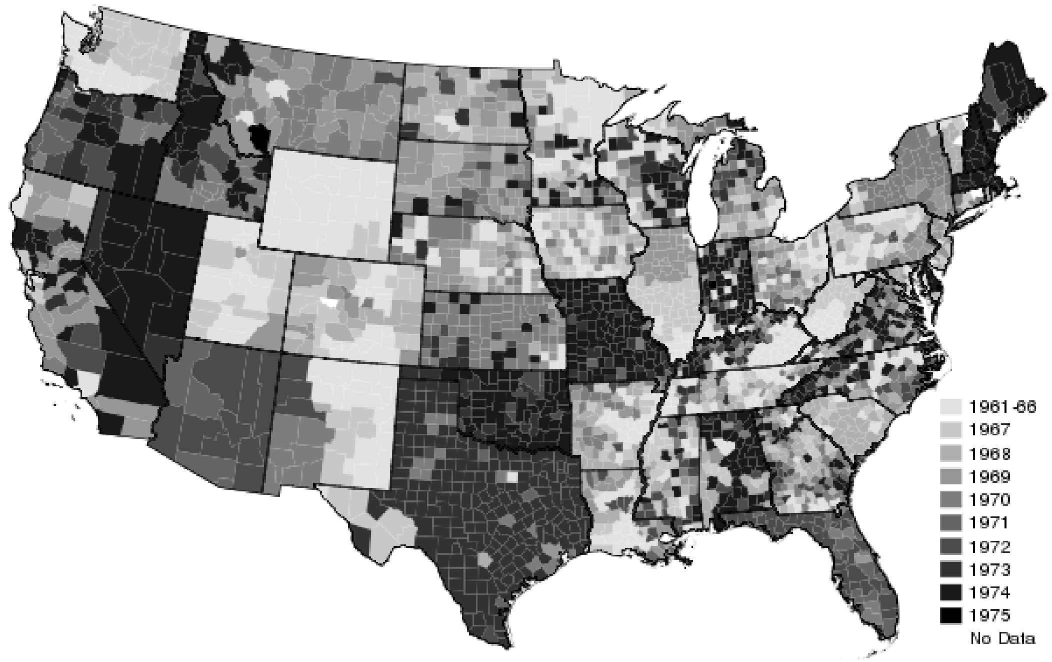
Local-level program administration produced substantial geographic variation in the timing of program introduction for Food Stamps and Head Start across counties, even within states. Figure 2.1 depicts this variation, separately for Food Stamps and Head Start. Panel (a) of Figure 2.1 shows the staggered roll-out of Food Stamps across counties during the 1960s and 1970s. The Food Stamps program began as a pilot program in a small number of counties in 1961. The program began expanding across counties following the passage of the 1964 Food Stamps Act. By 1975, the program was available in every county in the United States[11]. Panel (b) of Figure 2.1 shows the staggered roll-out of Head Start across counties during the same time period. The first Head Start center opened in 1965 with federal funding from the executive branch's Office of Economic Opportunity. The program spread across counties over subsequent years as new grants were issued by the federal government to local operators of Head Start centers. By 1980, eighty percent of the population in the United States resided in a county where a Head Start center was present[12].

10. See Bailey and Duquette (2014) for a detailed discussion of how War on Poverty programs were funded and implemented.

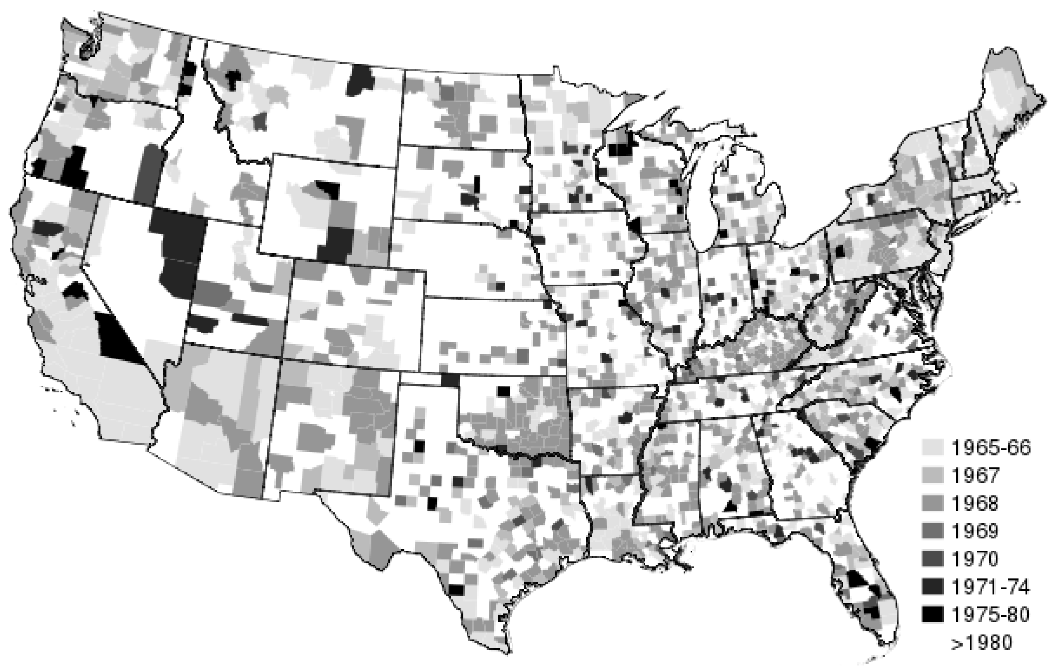
11. This is a brief summary of the history of the early years of the Food Stamps program. This history is adapted from the discussion in Hoynes, Schanzenbach and Almond (2016). See their paper for more details.

12. By author's calculation. See Figure 2.3 for details.

Figure 2.1: Timing of Program Introduction for Food Stamps and Head Start



(a) Food Stamps



(b) Head Start

Notes: Figure indicates year of program introduction for each county. Data on year of program introduction collected from Jackson and Johnson (Head Start) and Hoynes, Schanzenbach and Almond (Food Stamps).

My empirical strategy requires that the timing of program introduction for both Food Stamps and Head Start be effectively randomly assigned across counties. To assess the plausibility of this assumption, I review the factors identified in prior literature that explain why some counties are earlier-adopters of each program than others. Hoynes and Schanzenbach (2009) and Hoynes, Schanzenbach and Almond (2016) identify two predictors of the timing of program introduction for Food Stamps. First, limited funding during the program's roll-out period meant that some counties were placed on waiting lists when they applied to participate in the Food Stamps program. For these counties, the exact timing of program introduction depended on when federal funding became available. Second, some rural counties operated a commodity distribution program that provided low-income households with food that had been purchased by the federal government through its agricultural price support program. Counties were not allowed to operate both the commodity distribution program and the Food Stamps program simultaneously, so these rural counties tended to be later-adopters of the Food Stamps program. Ludwig and Miller (2007) identify a predictor of the timing of program introduction for Head Start. During the early years of the program, the federal government provided free assistance with completing Head Start grant applications to the poorest 300 counties in the United States. As a result, these counties were more likely to be earlier-adopters of Head Start.

Table 2.1 identifies additional predictors of the timing of program introduction for Food Stamps and Head Start. In this table, I regress the year of program introduction for either Food Stamps (Column 1) or Head Start (Column 2) on a set of 1960 county characteristics. The results in Column 1 indicate that poorer, more populous, and less rural counties were more likely to introduce Food Stamps earlier during the roll-out period. The results in Column 2 show that more populous counties with larger populations of preschool-age children tended to be earlier-adopters of the Head Start program. To address the potential concerns about non-random timing of program introduction for Food Stamps and Head Start that are raised by these findings, I include a county of birth fixed effect and a set of time-varying county of birth characteristics in all of my empirical models[13].

Table 2.1: Predictors of the Timing of Program Introduction

	(1) Food Stamps	(2) Head Start
Log of Population	-0.764***	-0.371***
	(0.0626)	(0.0822)
Percent of Land in Farming	0.0120***	0.000893
	(0.00212)	(0.00270)
Percent of Population with Income Less than 3K	-0.0444***	0.00973
	(0.00757)	(0.0102)
Percent of Population in Urban Area	-0.00478	0.00110
	(0.00350)	(0.00470)
Percent of Population Black	-0.0171**	0.0261***
	(0.00582)	(0.00781)
Percent of Population Under 5	-0.189***	-0.555***
	(0.0513)	(0.0666)
Percent of Population Over 65	-0.0861**	-0.188***
	(0.0320)	(0.0413)
Number of Counties	3059	1561
R-squared	0.557	0.205

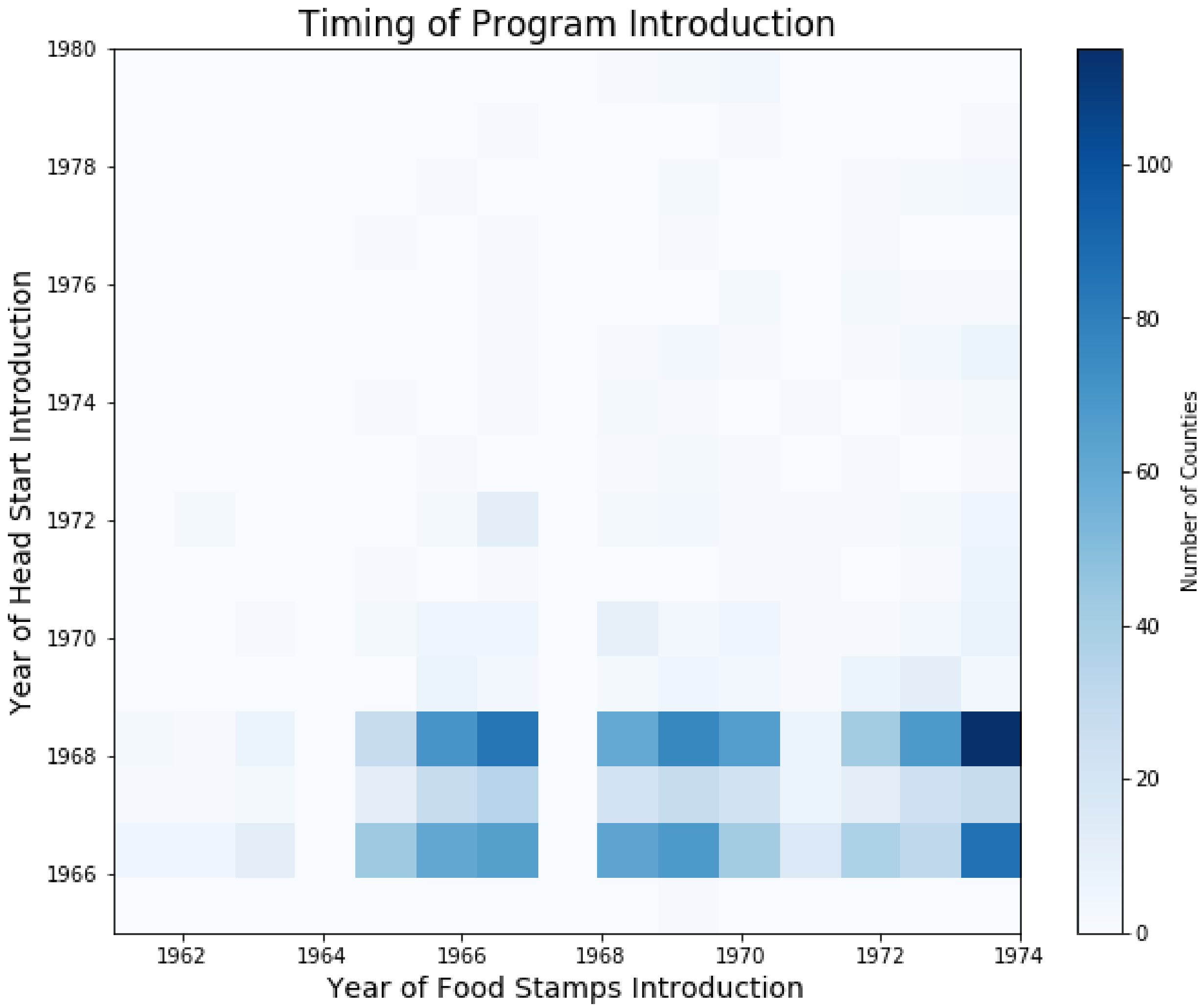
Notes: Table presents estimates from regressing the year of program introduction on a set of 1960 county characteristics and a state fixed effect, separately for Food Stamps (Column 1) and Head Start (Column 2). Year of program introduction is normalized to equal 0 in the first year of the program's existence. Regressions are weighted using 1960 county population. Data on year of program introduction collected from Jackson and Johnson (Head Start) and Hoynes, Schanzenbach and Almond (Food Stamps). Data on 1960 county characteristics collected from the City and County Databook. Standard errors are in parentheses. Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

13. I collect data on time-varying county characteristics for Census years and I interpolate this data for intermediate, non-Census years.

My empirical strategy also requires that the timing of Food Stamps introduction be less-than-perfectly correlated with the timing of Head Start introduction. During this time period, Food Stamps and Head Start were funded at the federal level by independently-operated government agencies (the Department of Agriculture and the Office of Economic Opportunity, respectively). This fact is reassuring, as idiosyncratic differences in application review processes and funding availability across agencies would tend to reduce the correlation between the timing of program introduction for Food Stamps and Head Start.

I present additional empirical evidence to support the less-than-perfect correlation assumption. Figure 2.2 plots the empirical joint probability distribution between the year of Food Stamps introduction and the year of Head Start introduction. Figure 2.2 shows no visual evidence of a strong correlation (positive or negative) between the year of Food Stamps introduction and the year of Head Start introduction. I also calculate the raw and regression-adjusted correlations (after controlling for a set of 1960 county characteristics and a state fixed effect) between the year of Food Stamps introduction and the year of Head Start introduction. The results are shown in Table 2.2. The raw correlation equals 0.096 and is statistically significant at a 1% level, while the regression-adjusted correlation equals 0.0555 and is statistically significant at a 10% level. The results in Table 2.2 prove that the timing of Food Stamps introduction is not perfectly correlated with the timing of Head Start introduction.

Figure 2.2: Empirical Joint Probability Distribution for the Timing of Program Introduction



Notes: Figure shows the empirical joint probability distribution for the year of Head Start introduction and the year of Food Stamps introduction. Data on year of program introduction collected from Jackson and Johnson (Head Start) and Hoynes, Schanzenbach and Almond (Food Stamps).

Table 2.2: Correlation between Timing of Program Introduction for Food Stamps and Head Start

	Year of Head Start Introduction	Residualized Year of Head Start Introduction
Year of Food Stamps Introduction	0.0961***	
Residualized Year of Food Stamps Introduction		0.0555*
Number of Counties	1,563	1,560

Notes: Table shows correlation between timing of program introduction for Food Stamps and timing of program introduction for Head Start. Column 1 presents raw correlation. Column 2 presents residualized correlation (after controlling for a set of 1960 county characteristics [log population, percent of land in farming, percent of population with income less than $3,000, percent of population in urban area, percent of population black, percent of population under 5, and percent of population over 65] and a state fixed effect). Data on year of program introduction collected from Jackson and Johnson (Head Start) and Hoynes, Schanzenbach and Almond (Food Stamps). Data on 1960 county characteristics collected from the City and County Databook. Significance levels: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

2.3 Data and Empirical Strategy

2.3.1 Data

I use data from a restricted-use version of the Panel Study of Income Dynamics (PSID). The PSID is a longitudinal survey that began in 1968 with a sample of 5,000 households. Members of the original sample and their descendants are re-interviewed every one to two years. The PSID collects information on survey respondents' earnings, household income, labor force participation, educational attainment, and demographic characteristics. In recent survey waves, the PSID has also collected information on the health conditions experienced by household heads and their spouses. Throughout my analysis, I use PSID sample weights to adjust for sample attrition and to produce nationally-representative estimates.

The restricted-use version of the PSID used in this paper includes the county of residence for each survey respondent in each survey wave. This information on county of residence allows me to construct a county of birth variable, equal to the respondent's county of residence in the year of birth (or, if the respondent was born prior to the first survey wave in 1968, equal to the respondent's county of residence in 1968)[14]. I merge the county of birth variable to a data set containing the year of program introduction for Food Stamps and Head Start in each county. This data was originally collected by Hoynes, Schanzenbach and Almond (2016) and Jackson and Johnson[15] (2019). I determine whether Food Stamps and Head Start are available in the respondent's county of birth during each stage of childhood[16]. I then exploit the longitudinal structure of the PSID to link each respondent's childhood exposure to Food Stamps and Head Start to her outcomes in adulthood.

I merge the PSID sample to a set of county-level variables collected from various sources. These variables include Census year demographic and labor market characteristics from ICPSR's County and City Databook series; poverty rates from the Census Bureau; safety net program expenditures from the Bureau of Economic Analysis Regional Economic Information System; annual Head Start expenditures from the National Archives Community Action Program Grant and Grantees Files and the Federal Outlays System Files; and annual Food Stamps caseloads from Hoynes, Schanzenbach and Almond (2016).

My analysis sample consists of the set of survey respondents in the PSID who were born between 1950 and 1976. These respondents were born during or just prior to the program rollouts for Food Stamps and Head Start. I include in the sample all observations from survey waves 1968 through 2017 for which the respondent was at least 24 years old and no more

14. I fill in remaining missing values using the respondent's household's county of residence in the 8 years prior to the year of birth.

15. I verified that the data on the year of program introduction for Head Start provided by Jackson and Johnson was consistent with data on the year of program introduction for Head Start provided by Bailey and Goodman-Bacon (2015). See Figure A.1.

16. By determining program availability for Food Stamps and Head Start using the respondent's county of birth instead of her county (or counties) of residence during childhood, I address concerns about endogenous mobility as a response to program introduction.

than 54 years old during the survey year. My results focus on the subsample of respondents whose parents did not complete high school, and whose households were therefore most likely to be income-eligible to participate in the Food Stamps and Head Start programs.

2.3.2 Empirical Strategy

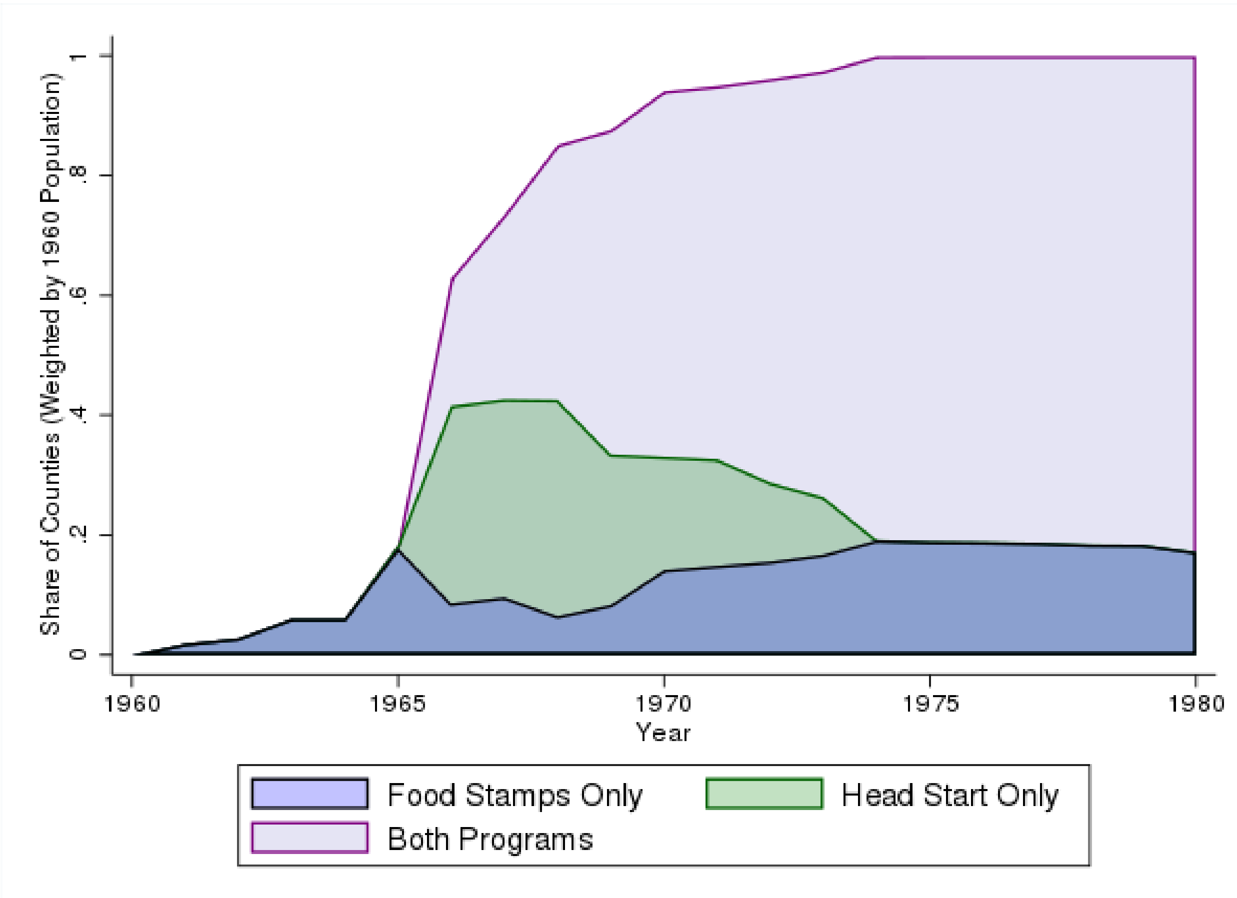
I present two sets of results in this paper. First, I use event study and difference-in-differences models to identify the effect of childhood access to Food Stamps and Head Start, respectively, on outcomes in adulthood. Second, I augment these models to include interaction effects, and I ask whether the long-run effect of access to Food Stamps at each stage of childhood depends on the availability of Head Start at preschool age.

The event study and difference-in-differences models estimated in this paper exploit variation in program exposure during childhood across individuals born during the time period that Food Stamps and Head Start were initially rolled out across counties. Figure 2.3 illustrates the source of this variation. The height of each shaded region in Figure 2.3 indicates the population-weighted share of counties that had introduced Food Stamps only, Head Start only or both programs by the indicated year, for each year between 1960 and 1980. Several features of the program roll-outs for Food Stamps and Head Start should be highlighted: the staggered roll-out of Food Stamps spanned the years 1961 to 1975, the staggered roll-out of Head Start spanned the years 1965 to 1980, and the order in which the two programs were introduced varied across counties[17]. These features produced two types of variation in program exposure across individuals that I take advantage of in my empirical design. First, because the exact timing of introduction for each program varied across counties, some individuals received more cumulative years of program exposure during childhood by virtue of their county and year of birth. Second, because the program roll-outs for Food Stamps and

17. Note that, unlike Food Stamps, some counties had not introduced Head Start by the time the roll-out period for the program ended. This is because Congress required that all counties introduce the Food Stamps program by 1975, but no such requirement was imposed for the Head Start program.

Head Start overlapped, some individuals were exposed to one, both, or neither program at each stage of childhood by virtue of their county and year of birth.

Figure 2.3: Overlapping Program Roll-outs for Food Stamps and Head Start



Notes: Figure depicts the program roll-outs for Food Stamps and Head Start. The height of each shaded region in the figure indicates the share of counties (weighted by 1960 population) that had introduced Food Stamps only, Head Start only, or both programs by the indicated year. Data on year of program introduction collected from Jackson and Johnson (Head Start) and Hoynes, Schanzenbach and Almond (Food Stamps).

The outcome variables of interest in my analysis consist of a set of labor market and health outcomes in adulthood. I follow the approach used by Hoynes, Schanzenbach and Almond (2016) and construct two summary indices of these variables. First, I construct a metabolic syndrome index by taking the average of the z-scores for whether an individual experienced each of the following conditions in adulthood: high blood pressure, diabetes, heart attack, heart disease and obesity[18]. Second, I construct an economic self-sufficiency index by taking the average of the z-scores for personal income, household income, years of education and whether an individual: was employed, graduated high school, and resided in a household whose income was above the federal poverty line in adulthood. When constructing the z-scores, I use as a reference group the set of birth cohorts born prior to 1961 (i.e., born prior to initial roll-outs for Food Stamps and Head Start). This index-based approach increases the statistical power of my estimates and reduces concerns about multiple hypothesis testing that arise when the empirical analysis involves many outcome variables (Kling, Liebman and Katz 2007; Hoynes, Schanzenbach and Almond 2016).

Empirical Models: Individual Program Effects

To measure the effect of childhood access to Food Stamps and Head Start, respectively, on outcomes in adulthood, I estimate the following event study equation:

$$y_{icbt} = \gamma_c + \lambda_b + \psi_t + \sum_{\tau} \alpha_\tau Prog^{\tau}_{cb} + \eta X_{icbt} + \epsilon_{icbt} \tag{2.1}$$

where α_τ are the event study estimates for the effect on outcome y of being exposed to the program beginning at age τ, instead of at omitted age ten[19]. γ_c is a county of birth fixed effect, λ_b is a year of birth fixed effect, and ψ_t is a survey year fixed effect. $Prog^{\tau}_{cb}$ is an indicator variable equal to one if the age at initial program exposure for an individual born in county c and year b is equal to τ. X_{icbt} is a vector of control variables that includes sex, race, and a set of time-varying county of birth characteristics (log population, share of population that is black, share of population that lives in an urban area, share of population

18. The z-score is constructed by subtracting the mean and dividing by the standard deviation. Hence, the z-score is measured in standard deviation units.

19. In other words, α_τ measures the difference in outcome y between individuals who are exposed to the program beginning at age τ and individuals who are exposed to the program beginning at age $\tilde{\tau} > \tau$, relative to the difference in outcome y between individuals who are exposed to the program beginning at age 10 and individuals who are exposed to the program beginning at age $\tilde{\tau} > 10$.

with annual income less than \$3000 in 1960 dollars, share of land that is used for farming). I cluster standard errors at the county of birth-by-county of residence level and I weight the regression using PSID survey weights. I estimate the model on a panel data set that includes multiple observations per survey respondent (where t indexes the survey year from which each observation is collected). In the results that follow, I plot the event study estimates α_τ by age at initial program exposure τ.

For table estimates, I estimate the following difference-in-differences equation:

$$y_{icbt} = \gamma_c + \lambda_b + \psi_t + \alpha ShareProg_{cb}^{\tau_0,\tau_1} + \eta X_{icbt} + \epsilon_{icbt} \tag{2.2}$$

where α is the difference-in-differences estimate for the effect on outcome y of full exposure to the program between ages τ_0 and τ_1[20]. $ShareProg_{cb}^{\tau_0,\tau_1}$ measures the share of years between ages τ_0 and τ_1 for which an individual born in county c in year b was exposed to the program. In the results that follow, I report estimates for α.

Empirical Models: Interaction Effects

To measure how the long-run effect of access to Food Stamps beginning at age $A \in [-6, 11]$ depends on the availability of Head Start between ages three and six, I estimate the following

20. For a county c that introduces the program in $YearIntro_c$, the difference-in-differences design compares the difference between the treatment and control groups in outcome y for cohorts born before $YearIntro_c$ to the difference between the treatment and control groups in outcome y for cohorts born after $YearIntro_c$. Here, the treatment group consists of the set of individuals born in a county c. The control group consists of the set of individuals born in counties that introduced the program after $YearIntro_c$ and who were not exposed to the program between ages τ_0 and τ_1.

augmented version of event study model 2.1[21]:

$$y_{icbt} = \gamma_c + \lambda_b + \psi_t + \sum_{\nu} \theta_\nu HS_{cb}^{\nu} + \sum_{\tau} \alpha_\tau FS_{cb}^{\tau} + \sum_{\tau} \beta_\tau ShareHS_{cb}^{3,6} \times FS_{cb}^{\tau} + \eta X_{icbt} + \epsilon_{icbt} \tag{2.3}$$

where β_τ is the added return to outcome y from exposure to Food Stamps beginning at age τ instead of at omitted age ten that results from having full access to a Head Start center between ages three and six. γ_c is a county of birth fixed effect, λ_b is a year of birth fixed effect, and ψ_t is a survey year fixed effect. HS_{cb}^{ν} is an indicator variable equal to one if the age at initial Head Start exposure for an individual born in county c in year b equals ν, and FS_{cb}^{τ} is an indicator variable equal to one if the age at initial Food Stamps exposure for an individual born in county c in year b equals τ. The share variable $ShareHS_{cb}^{3,6}$ measures the share of years between ages 3 and 6 for which an individual born in county c in year b was exposed to Head Start. I specify the interaction term as $\sum_{\tau} \beta_\tau ShareHS_{cb}^{3,6} \times FS_{cb}^{\tau}$ instead of as $\sum_{\tau,\nu} \beta_{\tau,\nu} HS_{cb}^{\nu} \times FS_{cb}^{\tau}$ because I lack the statistical power in my sample to estimate the full set of interaction terms $\beta_{\tau,\nu}$. I cluster standard errors at the county of birth-by-county of residence level and I weight the regression using PSID survey weights. In the results that follow, I plot the main effect of Food Stamps α_τ by the age at Food Stamps introduction τ. I also plot the main effect of Food Stamps plus the interaction coefficient $\alpha_\tau + \beta_\tau$ by the age at Food Stamps introduction τ.

21. I study the age range $[-6, 11]$ because, for the ages in this range, some individuals had access to both, one, or neither program depending on their county and year of birth. Outside of this range, there is insufficient variation in program exposure across individuals to identify the interaction effects between Food Stamps and Head Start.

For table estimates, I estimate the following difference-in-differences equation:

$$
\begin{aligned}
y_{icbt} = \gamma_c + \delta_b + \theta ShareHS_{cb}^{3,6} + \alpha_1 ShareFS_{cb}^{-1,3} + \alpha_2 ShareFS_{cb}^{3,6} + \alpha_3 ShareFS_{cb}^{6,11} \\
+ \beta_1 ShareHS_{cb}^{3,6} \times ShareFS_{cb}^{-1,3} \\
+ \beta_2 ShareHS_{cb}^{3,6} \times ShareFS_{cb}^{3,6} \\
+ \beta_3 ShareHS_{cb}^{3,6} \times ShareFS_{cb}^{6,11} \\
+ \eta X_{icbt} + \epsilon_{icbt}
\end{aligned}
\tag{2.4}
$$

where $ShareHS_{cb}^{3,6}$ measures the share of years between ages three and six for which an individual born in county c in year b was exposed to Head Start, and $ShareFS_{cb}^{\tau_0,\tau_1}$ measures the share of years between ages τ_0 and τ_1 for which an individual born in county c in year b was exposed to Food Stamps. In the results that follow, I report estimates for the main program effects θ, α_1, α_2, and α_3. I also report estimates for the interaction effects β_1, β_2, and β_3.

Identification

The event study and difference-in-differences models identify the interaction effect between Food Stamps and Head Start by comparing long-run outcomes for individuals who were exposed to both, only one, or neither program during childhood. The key identifying assumption for these models is that, in the absence of Food Stamps or Head Start, outcomes would have evolved similarly across birth cohorts between individuals exposed to both programs during childhood and individuals exposed to only one or neither program.

I take a few steps to verify that the identifying assumption is satisfied in my setting. First, the individual program event study estimates reported in Figures 2.6, 2.7 and 2.8 show that there is no evidence of statistically significant, non-parallel pre-trends in the outcome variable for cohorts whose exposure to Food Stamps begins after age eighteen (or, whose exposure to Head Start begins after age six). Second, the identifying assumption is violated if the timing of program roll-out for Food Stamps and Head Start is non-randomly assigned across counties. I address this threat to identification by including a set of time-varying county of birth characteristics in each of my empirical models. Third, the identifying assumption is violated if the timing of program roll-out for Food Stamps is perfectly correlated with the timing of program roll-out for Head Start. In Table 2.2, I showed that the raw and regression-adjusted correlations between the year of introduction for Food Stamps and the year of introduction for Head Start are less than one in absolute value.

2.4 Results: Individual Program Effects

In this section, I estimate the individual program effects of Food Stamps and Head Start. The purpose of this section is not to introduce a new set of estimates to the literature. Two previous papers report a similar set of estimates, produced using similar samples from the Panel Study of Income Dynamics and the same program roll-out-based empirical design. Hoynes, Schanzenbach and Almond (2016) estimate the long-run effect of childhood exposure to Food Stamps, while Jackson and Johnson (2019) estimate the long-run effect of childhood exposure to Head Start. Instead, I produce these estimates in order to evaluate whether my results are consistent with the results from the prior literature. By directly comparing my estimates to the estimates in Hoynes, Schanzenbach and Almond (2016) and Jackson and Johnson (2019), I establish the credibility of the interaction effect results reported in the next section.

2.4.1 Estimates for the Effect of Childhood Exposure to Food Stamps

Figure 2.4 presents the event study estimates from empirical model (2.1) for the effect of childhood exposure to Food Stamps on the metabolic syndrome index. For the discussion that follows, note that higher values for the metabolic syndrome index indicate worse health in adulthood and that cumulative years of Food Stamps exposure during childhood are decreasing in the age at the time of Food Stamps introduction. The event study estimates in Figure 2.4 show that children first exposed to Food Stamps beginning at age nine or younger experience a reduction in the metabolic syndrome index in adulthood, relative to children first exposed to Food Stamps beginning at age ten or eleven. In other words, more cumulative years of Food Stamps exposure during childhood (moving from right to left in the figure) leads to fewer metabolic syndromes in adulthood, and the effect is largest among children who are fully exposed to Food Stamps during the in-utero period and childhood (i.e., for whom the age at Food Stamps introduction equals -1 or younger)[22]. The shaded region in Figure 2.4 represents the 90% confidence intervals for the event study estimates. The 90% confidence intervals indicate that most of the event study estimates are not statistically significant at a 10% significance level.

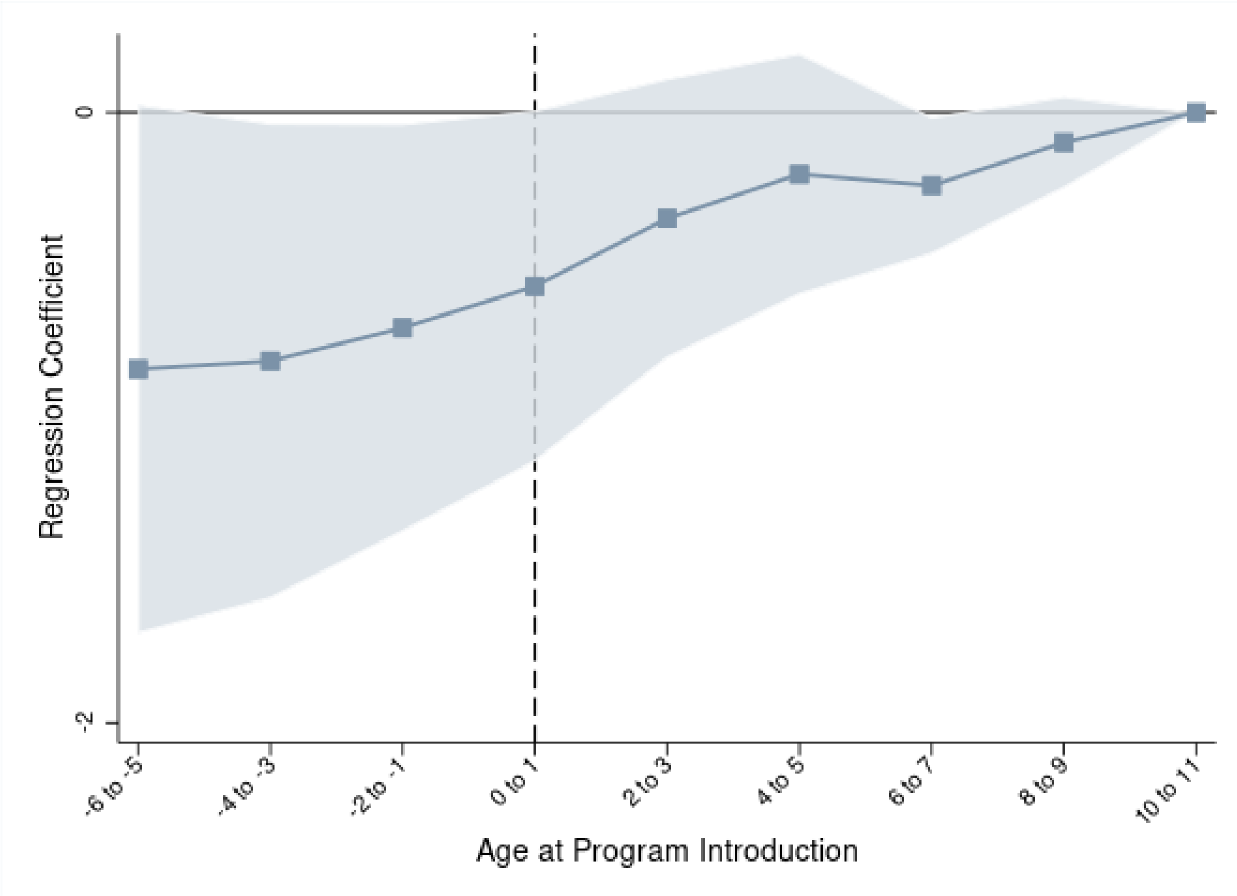
Table 2.3 presents the difference-in-differences estimates from empirical model (2.2) for the effect of exposure to Food Stamps between the in-utero period and age eleven on the metabolic syndrome index and each of its components. The estimate in Column 1 indicates that full exposure to Food Stamps between the in-utero period and age eleven leads to a statistically insignificant 0.11 standard deviation reduction in the metabolic syndrome index. This estimate is consistent in sign with the event study estimates reported in Figure 2.4, though the effect size is smaller than the effect size implied by the event study estimates in Figure 2.4[23]. The estimate in Column 6 shows that the effect of exposure to Food Stamps between the in-utero period and age eleven on the metabolic syndrome index is driven, in part, by a statistically insignificant 24 percentage point (67 percent, relative to the sample

22. This result is consistent with previous work that has shown that the effect of childhood exposure to Food Stamps on long-run outcomes is largely driven by exposure to Food Stamps during early childhood (in-utero to age five). See Hoynes, Schanzenbach and Almond (2016) and Bailey et. al. (2019).

23. The difference-in-differences estimate for the effect of Food Stamps exposure between the in-utero period and age eleven on the metabolic syndrome index (0.11) falls within the 90% confidence interval for each of the event study estimates corresponding to Food Stamps exposure beginning at age 9 or younger.

mean) decrease in the incidence of obesity in adulthood.

Figure 2.4: Effect of Food Stamps on Metabolic Syndrome Index



Notes: Figure plots event study estimates for the effect of Food Stamps exposure on the metabolic syndrome index by the age at the time of Food Stamps introduction. Figure plots event study estimates with 90% confidence intervals. Cohorts to the left of the dashed line are fully exposed to Food Stamps between ages 0 and 11. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level.

Table 2.3: Effect of Food Stamps on Metabolic Syndrome Index and its Components

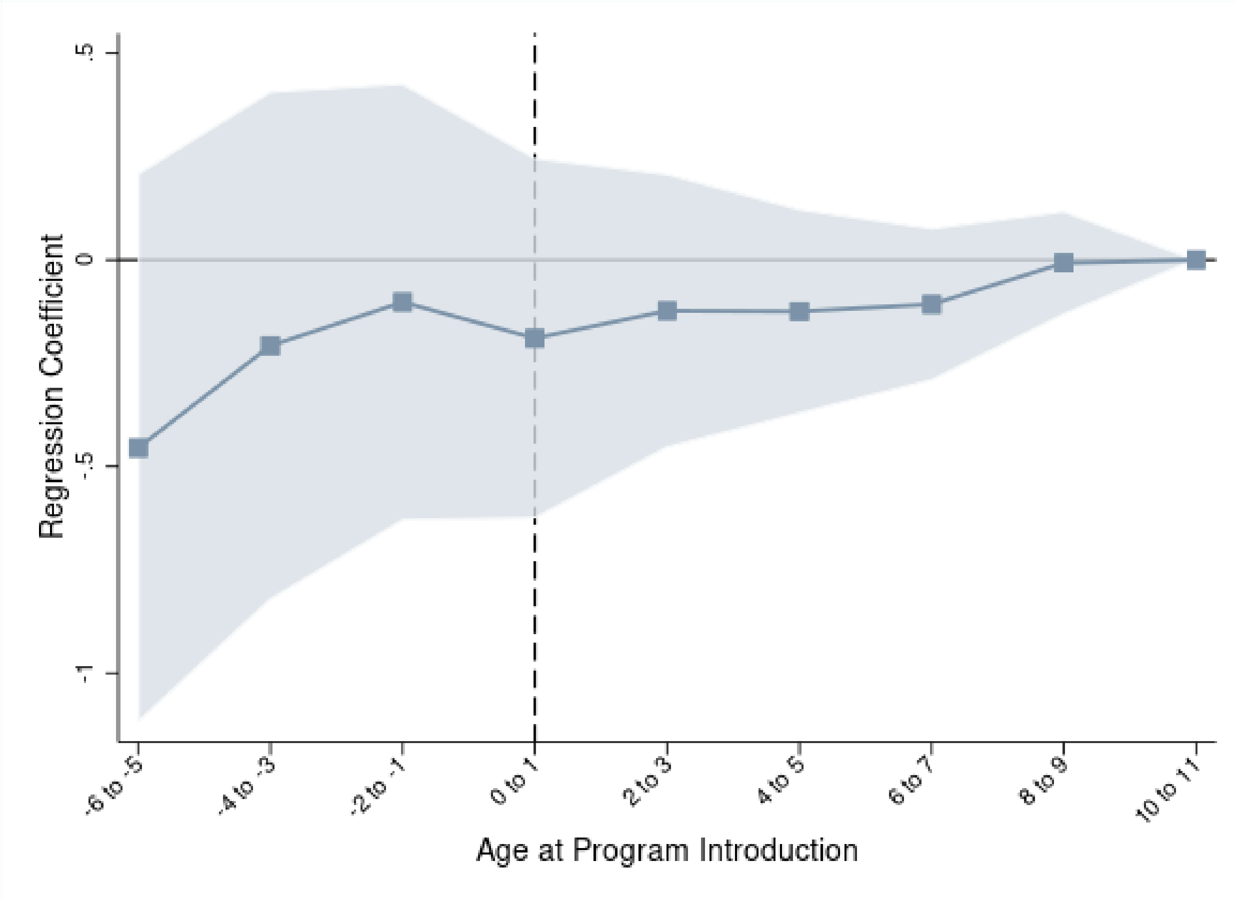
	Metabolic Syndrome Index	Components of Index				
		High Blood Pressure	Diabetes	Heart Attack	Heart Disease	Obesity
Food Stamps IU-11	-0.107	-0.00168	-0.000933	-0.00584	0.00677	-0.241
	(0.146)	(0.116)	(0.0754)	(0.0266)	(0.0438)	(0.148)
Mean of Y	0.0784	0.271	0.0846	0.0178	0.0410	0.358
Number of Person-Years	13,476	13,826	13,828	13,828	13,824	15,139
Number of Persons	2,080	2,090	2,090	2,090	2,089	2,874

Notes: Table reports difference-in-differences estimates for the effect of Food Stamps exposure between the in-utero period and age five on the metabolic syndrome index and each of its components. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level. Standard errors are in parentheses. Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Figure 2.5 presents the event study estimates from empirical model (2.1) for the effect of childhood exposure to Food Stamps on the economic self-sufficiency index. The event study estimates indicate that children first exposed to Food Stamps beginning prior to age eleven experience a slight decrease in the economic self-sufficiency index in adulthood, relative to children first exposed to Food Stamps beginning at age ten or eleven. However, casting doubt on the conclusion that childhood exposure to Food Stamps worsens economic self-sufficiency in adulthood is the fact that the 90% confidence intervals for the event study estimates (represented by the shaded region in Figure 2.5) do not rule out a positive impact of childhood Food Stamps exposure on economic self-sufficiency.

Table 2.4 presents the difference-in-differences estimates from empirical model (2.2) for the effect of exposure to Food Stamps between the in-utero period and age eleven on the economic self-sufficiency index. In contrast with the event study estimates in Figure 2.5, the difference-in-differences estimates provide strong evidence that early childhood exposure to Food Stamps improves economic self-sufficiency in adulthood. The estimate in Column 1 shows that full exposure to Food Stamps between the in-utero period and age eleven results in a statistically significant (at a 5% significance level) 0.48 standard deviation increase in the economic self-sufficiency index. The estimates in Columns 2 through 7 show that early childhood exposure to Food Stamps leads to an improvement in each of the components of the economic self-sufficiency index, including a statistically significant (at a 5% significance level) 25.4 percentage point (29 percent, relative to the sample mean) increase in the likelihood of graduating high school and a statistically insignificant 0.6 year (5 percent, relative to the sample mean) increase in years of completed schooling.

Figure 2.5: Effect of Food Stamps on Economic Self-Sufficiency Index



Notes: Figure plots event study estimates for the effect of Food Stamps exposure on the economic self-sufficiency index by the age at the time of Food Stamps introduction. Figure plots event study estimates with 90% confidence intervals. Cohorts to the left of the dashed line are fully exposed to Food Stamps between ages 0 and 11. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level.

Table 2.4: Effect of Food Stamps on Economic Self-Sufficiency Index and its Components

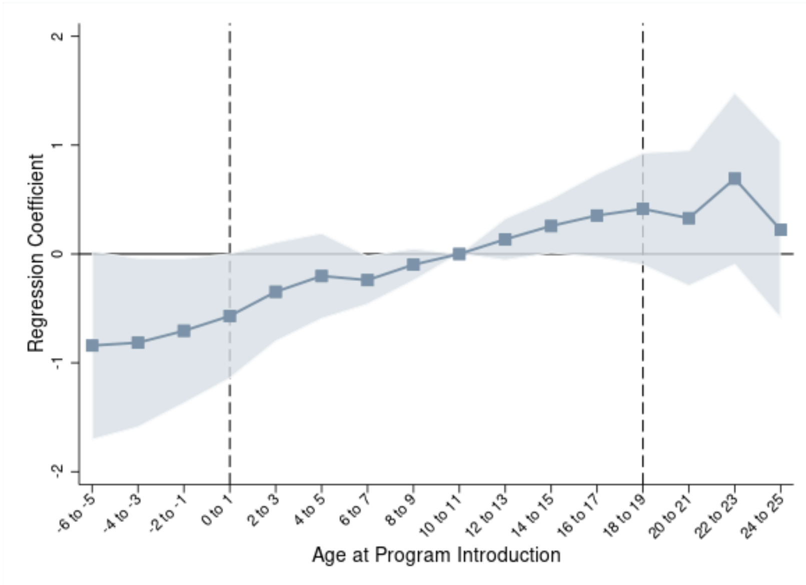
	Economic Self-Sufficiency Index	Components of Index					
		Employed	Personal Earnings	Family Earnings	Above Poverty Line	High School Plus	Years of Education
Food Stamps IU-11	0.482**	0.167*	4,135.3	7,376.7	0.165**	0.254**	0.630
	(0.200)	(0.100)	(6,244.8)	(8,124.6)	(0.0733)	(0.0655)	(0.421)
Mean of Y	-0.245	0.761	22,018.3	46,449.1	0.824	0.867	12.98
Number of Person-Years	47,024	48,332	47,663	56,857	56,927	-	-
Number of Persons	3,642	3,655	3,643	4,128	4,128	4,004	4,004

Notes: Table reports difference-in-differences estimates for the effect of Food Stamps exposure between the in-utero period and age eleven on the economic self-sufficiency index and each of its components. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level. Standard errors are in parentheses. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

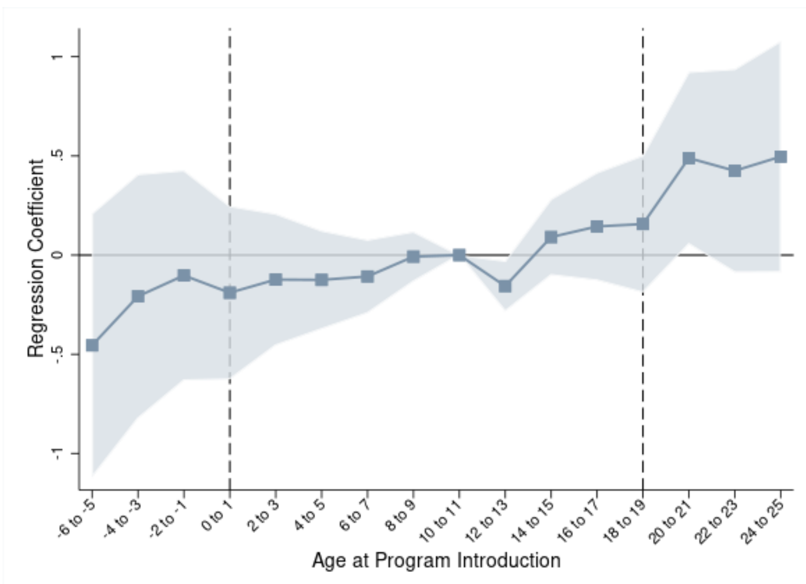
In Figure 2.6, I plot the event study estimates for the effect of childhood exposure to Food Stamps on the metabolic syndrome index (Panel (a)) and the economic self-sufficiency index (Panel (b)) over an expanded range of ages, relative to Figures 2.3 and 2.4. Reassuringly for my identification strategy, the event study estimates in Figure 2.6 show no evidence that exposure to Food Stamps beginning after age 18 has a statistically significant impact on the economic self-sufficiency index or the metabolic syndrome index in adulthood. In other words, neither panel shows evidence of statistically significant, non-parallel pre-trends in the outcome variable between the treatment and control groups (i.e., between individuals with versus without access to Food Stamps beginning at the indicated age)[24].

24. This is not necessarily the ideal test for parallel pre-trends, since individuals whose exposure to Food Stamps begins after age 18 may still be partially treated by their exposure to Food Stamps in adulthood. However, it is reasonable to the extent that the effect of Food Stamps exposure on long-run outcomes is concentrated in childhood.

Figure 2.6: Event Study Estimates for Food Stamps over an Expanded Age Range



(a) Metabolic Syndrome Index



(b) Economic Self-Sufficiency Index

Notes: Figure plots event study estimates for the effect of Food Stamps exposure on the indicated outcome variable by the age at the time of Food Stamps introduction. Figure includes an expanded range of ages, relative to Figures 2.4 and 2.5. Panel (a) plots event study estimates and 90% confidence intervals for the metabolic syndrome index. Panel (b) plots event study estimates and 90% confidence intervals for the economic self-sufficiency index. See notes for Figures 2.4 and 2.5 for more details.

The difference-in-differences estimates reported in Tables 2.3 and 2.4 are qualitatively similar to the estimates obtained by Hoynes, Schanzenbach and Almond (2016; henceforth HSA) in their study of the long-run effects of childhood exposure to Food Stamps. Using the same empirical strategy and a similar sample from the Panel Study of Income Dynamics, HSA find that exposure to Food Stamps between the in-utero period and age five decreases the metabolic syndrome index and increases an economic self-sufficiency index in adulthood. Because we define our samples differently, construct our outcome variable indices using different component variables, and include different control variables in our empirical models, the estimates that I report in Tables 2.3 and 2.4 are not identical to the estimates reported by HSA[25]. I perform a simple replication exercise to verify that the differences in our estimates are fully explained by the differences in our sample definitions, outcome variable definitions, and empirical specifications. For this exercise, I apply the sample definition used by HSA to the data set that I constructed for my analysis from the Panel Study of Income Dynamics. I construct indices of metabolic health and economic self-sufficiency using the same component variables as HSA, and I estimate the empirical model used in their paper. Table 2.5 describes the results of this replication exercise. Each column in the table reports an estimate for the effect of exposure to Food Stamps between the in-utero period and age five on the indicated outcome variable. Columns 1 and 2 report the estimates obtained by HSA. Columns 3 and 4 report the estimates that I obtain from the replication exercise. Comparing Column 1 to Column 3 and Column 2 to Column 4 reveals that my replication exercise results are nearly identical to the results obtained by HSA. This lends credibility to the estimated interaction effects between Food Stamps and Head Start that I report in the next section of the paper.

25. The sample used by Hoynes, Schanzenbach and Almond (2016) includes individuals born between 1956 and 1981 who appear as heads or spouses in the Panel Study of Income Dynamics (PSID). Their sample includes all PSID survey years between 1968 and 2009.

Table 2.5: Replication Exercise for Hoynes, Schanzenbach and Almond (2016)

	HSA (2016)		Replication of HSA (2016) in My Sample	
	Metabolic Syndrome Index	Economic Self-Sufficiency Index	Metabolic Syndrome Index	Economic Self-Sufficiency Index
Food Stamps Availability, IU-5	-0.294**	0.182	-0.262**	0.198*
	(0.107)	(0.124)	(0.103)	(0.116)
Mean of Y	0.01	-0.25	0.01	-0.25
Person-Year Observations	8,246	20,115	8,267	20,084
R^2	0.26	0.38	0.26	0.38

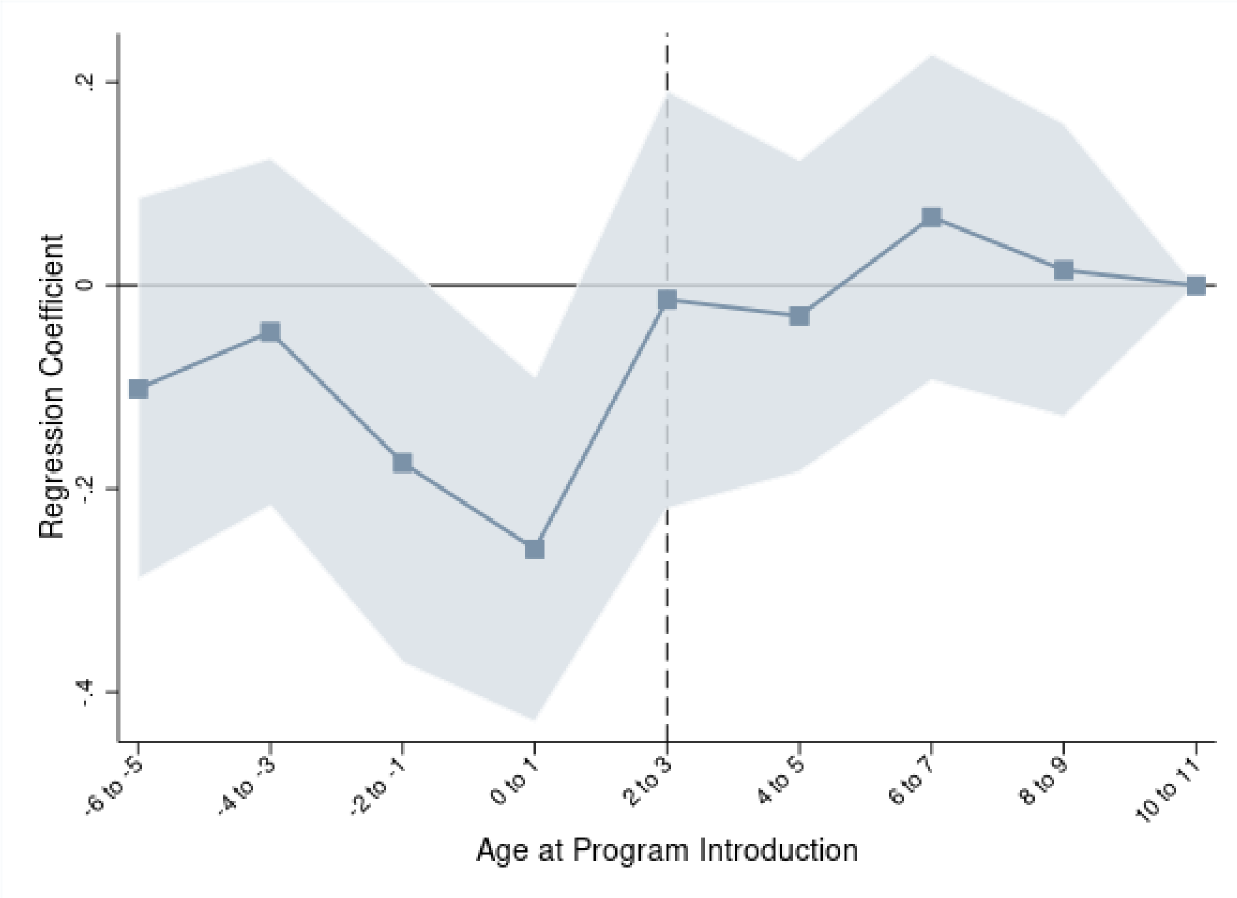
Notes: Table reports difference-in-differences estimates for the effect of Food Stamps exposure between the in-utero period and age five on the indicated outcome variable. Columns 1 and 2 report estimates obtained from Tables 2 and 4 in Hoynes, Schanzenbach and Almond (2016). Columns 3 and 4 report estimates obtained from the replication exercise described in the text. The metabolic health index equals the average of the z-scores for whether an individual experienced each of the following conditions in adulthood: high blood pressure, diabetes, heart attack, heart disease and obesity. The economic self-sufficiency index equals the average of the z-scores for personal income, household income and whether an individual: graduated high school, resided in a household whose income was above the federal poverty line, did not receive Food Stamps and did not receive TANF in adulthood. The empirical model includes controls for sex, race, family background, state fixed effects interacted with a linear year of birth trend and a set of 1960 county characteristics interacted with a linear year of birth trend. The sample includes PSID survey respondents born between 1956 and 1981 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered by county of birth. See Hoynes, Schanzenbach and Almond (2016) for more details. Standard errors are in parentheses. Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

2.4.2 Estimates for the Effect of Childhood Exposure to Head Start

Figure 2.7 plots the event study estimates from empirical model (2.1) for the effect of childhood exposure to Head Start on the metabolic syndrome index. For the remainder of the discussion, note that Head Start attendance occurs between ages three and six during the time period that I study. The event study estimates indicate that children who are first exposed to Head Start beginning at age six or younger experience fewer metabolic syndromes in adulthood, compared to children exposed to Head Start beginning at age ten or eleven. The event study estimates also show that the incidence of metabolic syndromes is decreasing in the amount of Head Start exposure between ages three and six (moving from right to left in the figure), with the largest decreases concentrated among children who are fully exposed to Head Start between ages three and six (i.e., who are three years old or younger at the time of initial Head Start exposure). The shaded region in Figure 2.7 depicts the 90% confidence intervals for the event study estimates. The 90% confidence intervals indicate that, while the estimated effect of Head Start exposure beginning at age zero or one is statistically significant at a 10% significance level, the remaining event study estimates are statistically insignificant.

Table 2.6 reports the difference-in-differences estimates from empirical model (2.2) for the effect of exposure to Head Start between ages three and six on the metabolic syndrome index and each of its components. The estimate in Column 1 shows that full exposure to Head Start between ages three and six leads to a statistically insignificant 0.13 standard deviation reduction in the metabolic syndrome index. This result is consistent in sign and magnitude with the event study estimates reported in Figure 2.7. Further, the estimates in Columns 2, 5, and 6 indicate that exposure to Head Start between ages three and six reduces the incidence of some of the component metabolic syndromes in adulthood, including a statistically significant (at a 5% significance level) 15.5 percentage point (43 percent, relative to the sample mean) decrease in the incidence of obesity and a statistically significant (at a 10% significance level) 11 percentage point (39 percent, relative to the sample mean) decrease in the incidence of high blood pressure.

Figure 2.7: Effect of Head Start on Metabolic Syndrome Index



Notes: Figure plots event study estimates for the effect of Head Start exposure on the metabolic syndrome index by the age at the time of Head Start introduction. Figure plots event study estimates with 90% confidence intervals. Cohorts to the left of the dashed line are fully exposed to Head Start between ages three and six. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level.

Table 2.6: Effect of Head Start on Metabolic Syndrome Index and its Components

	Metabolic Syndrome Index	Components of Index: High Blood Pressure	Diabetes	Heart Attack	Heart Disease	Obesity
Head Start 3-6	-0.128	-0.106*	0.00625	0.00314	-0.00237	-0.155**
	(0.0907)	(0.0599)	(0.0370)	(0.0135)	(0.0232)	(0.0652)
Mean of Y	0.0784	0.271	0.0846	0.0178	0.0410	0.358
Number of Person-Years	13,476	13,826	13,828	13,828	13,824	15,139
Number of Persons	2,080	2,090	2,090	2,090	2,089	2,874

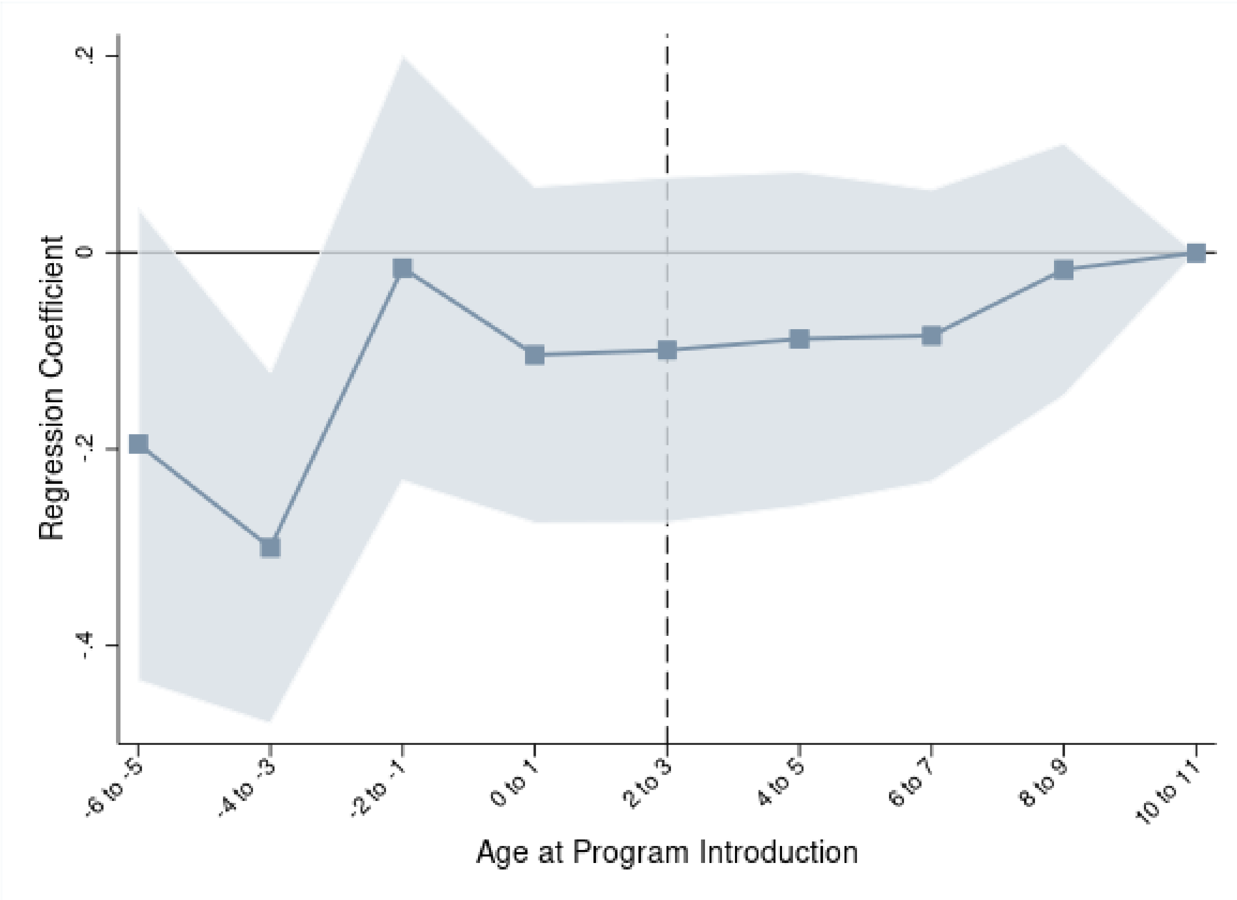
Notes: Table reports difference-in-differences estimates for the effect of Head Start exposure between ages three and six on the metabolic syndrome index and each of its components. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level. Standard errors are in parentheses. Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Figure 2.7 plots the event study estimates from empirical model (2.1) for the effect of childhood exposure to Head Start on the economic self-sufficiency index. The event study estimates indicate that children exposed to Head Start beginning at age six or younger experience a reduction in economic self-sufficiency in adulthood, relative to children exposed to Head Start beginning at age ten or eleven. The 90% confidence intervals for the event study estimates (represented by the shaded region in the figure) show that each of the event study estimates is statistically insignificant at a 10% significance level, with the exception of the event study estimate corresponding to ages -4 to -3.

Table 2.7 provides the difference-in-differences estimates for the effect of exposure to Food Stamps between ages three and six on the economic self-sufficiency index and each of its components. The difference-in-differences estimates reported in Table 2.7 are consistent with the event study estimates reported in Figure 2.7. The estimate in Column 1 shows that full exposure to Head Start between ages three and six leads to a statistically significant (at a 5% significance level) 0.15 standard deviation decrease in the economic self-sufficiency index. The estimates in Columns 2 through 7 show that exposure to Head Start between ages three and six reduces each of the components of the economic self-sufficiency index.

Notice that, in Figures 2.6 and 2.7, the event study estimates for the effect of exposure to Head Start beginning after age six on the metabolic syndrome index and the economic self-sufficiency index are close to zero: there is no evidence that exposure to Head Start beginning after the age of Head Start entry affects either outcome variable. This finding provides empirical evidence in support of my identification strategy, which requires parallel trends in the outcome variable between the treatment and control groups (i.e., between individuals with versus without access to Food Stamps beginning at the indicated age).

Figure 2.8: Effect of Head Start on Economic Self-Sufficiency Index



Notes: Figure plots event study estimates for the effect of Head Start exposure on the economic self-sufficiency index by the age at the time of Head Start introduction. Figure plots event study estimates with 90% confidence intervals. Cohorts to the left of the dashed line are fully exposed to Head Start between ages three and six. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level.

Table 2.7: Effect of Head Start on Economic Self-Sufficiency Index and its Components

	Economic Self-Sufficiency Index	Components of Index					
		Employed	Personal Earnings	Family Earnings	Above Poverty Line	High School Plus	Years of Education
Head Start 3-6	-0.145**	-0.0801**	-137.8	-2,036.2	-0.0597*	-0.0440	-0.343
	(0.0729)	(0.0396)	(2,358.7)	(3,854.4)	(0.0304)	(0.0384)	(0.212)
Mean of Y	-0.245	0.761	22,018.3	46,449.1	0.824	0.867	12.98
Number of Person-Years	47,024	48,332	47,663	56,857	56,927	-	-
Number of Persons	3,642	3,655	3,643	4,128	4,128	4,004	4,004

Notes: Table reports difference-in-differences estimates for the effect of Head Start exposure between ages three and six on the economic self-sufficiency index and each of its components. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level. Standard errors are in parentheses. Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

I compare the difference-in-differences estimates that I obtain in Table 2.7 to a set of related estimates reported by Jackson and Johnson (2019). Jackson and Johnson study the impact of interactions between Head Start and K-12 school spending on educational attainment and earnings. Their empirical strategy exploits (1) variation in age four Head Start spending induced by the county-level roll-out of Head Start and (2) variation in K-12 school spending induced by school finance reforms. Their sample consists of children in the Panel Study of Income Dynamics who were born between 1950 and 1976 and whose average parental earnings between ages 12 and 17 fell in the bottom quartile of the earnings distribution. Their main empirical specification includes a school district of birth fixed effect. Jackson and Johnson estimate that, for children in their sample who were exposed to average K-12 spending levels, exposure to a Head Start center with average per-pupil spending at age four is associated with a statistically significant 17.3 percentage point (24.4 percent, relative to the mean) increase in the likelihood of graduating high school and a statistically significant 0.954 (7.8 percent, relative to the mean) increase in years of completed education. These estimates are reported in Columns 1 and 2 of Table 2.8. In contrast, I find that preschool-age exposure to Head Start leads to a statistically insignificant 4.4 percentage point (5.1 percent, relative to the mean) decrease in the likelihood of graduating high school and a statistically insignificant 0.343 (2.6 percent, relative to the mean) decrease in years of completed education. These estimates are reported in Columns 3 and 4 of Table 2.8.

What can explain the discrepancy between the results reported by Jackson and Johnson and the results from my paper, as summarized in Table 2.8? Here, I describe a few possible explanations. First, the sample used by Jackson and Johnson differs from the sample used in this paper. Jackson and Johnson produce their estimates using a sample of children whose average parental income between ages 12 and 17 fell in the bottom quartile of the earnings distribution. I produce my estimates using a sample of children whose parents did not graduate from high school. Second, for the estimates reported in Table 2.8, the treatment effect for Head Start estimated by Jackson and Johnson differs from the treatment effect for Head Start that is estimated in this paper. Jackson and Johnson estimate the effect of exposure to a Head Start center with average per-pupil spending at age four on educational attainment, for children who were exposed to average K-12 spending levels. I estimate the effect of exposure to a Head Start center between ages three and six on educational attainment. Third, the empirical model used by Jackson and Johnson to produce their estimates is not the same as the empirical model used in this paper. For example, the model that Jackson and Johnson use to produce their main set of results includes a school district of birth fixed effect, whereas my model includes a county of birth fixed effect.

To identify the source of the discrepancy between the results reported by Jackson and Johnson and the results in this paper, I perform a replication exercise. In general, I cannot replicate Jackson and Johnson's main results because these results are produced using an empirical model that includes a school district of birth fixed effect. The school district of birth fixed effect is constructed from restricted-use data to which I do not have access. However, I should be able to replicate one set of results in their paper (Figure 2 of Jackson and Johnson (2019)) that do not depend on the school district of birth fixed effect. I attempt to replicate this set of results using data sets and code that were provided in the replication files for Jackson and Johnson's paper. This replication exercise proceeds in three steps. First, I construct the analysis sample. All variables in the analysis sample are either directly taken from data sets that Jackson and Johnson provide or are produced using code that they provide, with the exception of a county of birth variable that I construct according to the method described in their paper. Second, I impose the sample restriction used by Jackson and Johnson. No code is provided by Jackson and Johnson for this step; instead, I use the method described in their paper and restrict the sample to children whose average parental income between ages 12 and 17 fell in the bottom quartile of the distribution. Third, I produce a set of event study estimates using code that Jackson and Johnson provide with their paper.

I report the findings from the replication exercise in Figure 2.9. Figure 2.9 plots the event study estimates for the effect of Head Start exposure on educational attainment, separately for high and low Head Start spending counties. The left side of each panel shows the plot provided in Jackson and Johnson (2019), and the right side of each panel shows the plot that I produced in the replication exercise. Comparing the left and right plots in each panel shows that I am unable to reproduce the results provided in Jackson and Johnson (2019). Further work is needed to understand how to reconcile our results.[26].

26. There are two likely explanations for my inability to replicate Jackson and Johnson's results using the data sets and code that they provide. The first possibility is that the sample restriction imposed by Jackson and Johnson in practice differs from the sample restriction that I impose based on the description in the text of their paper. The second possibility is that the county of birth variable used by Jackson and Johnson in practice differs from the county of birth variable that I construct based on the description in the text of their paper. In either case, I could fail to replicate their results if estimating their model with the sample restriction or county of birth variable that they use in practice produces different results than estimating their model with the sample restriction or county of birth variable that I produce based on the description in the text of their paper.
Neither possibility threatens the validity of my results. If the first possibility were true, this would indicate that the long-run impact of Head Start is heterogeneous across subsamples

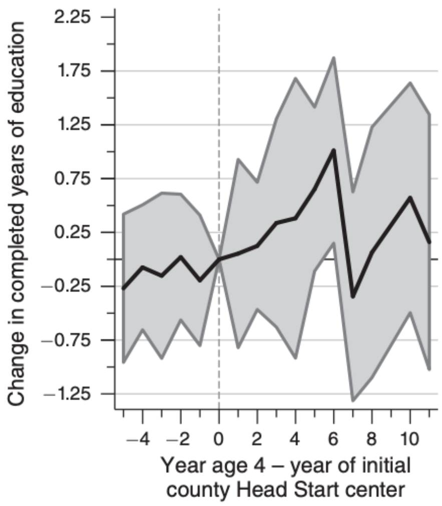
Table 2.8: Comparing Results from Jackson and Johnson (2019) and Carter (2019)

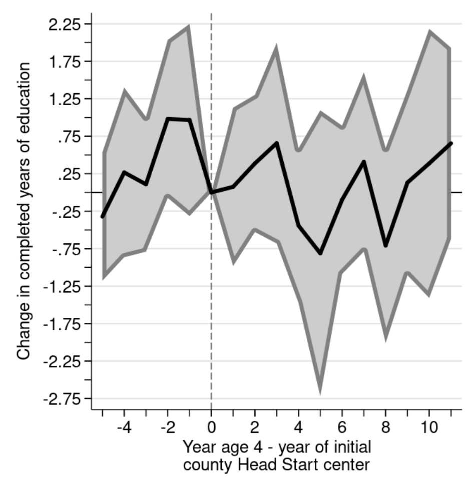
	Jackson and Johnson (2019)		Carter (2019)	
	High School Plus	Years of Education	High School Plus	Years of Education
Head Start with Average Head Start Spending, 4	0.1730* (0.1038)	0.9540* (0.5129)		
Head Start, 3-6			-0.0440	-0.343
			(0.0384)	(0.212)
Mean of Y	0.71	12.29	0.867	12.98
Person Observations	5,419	5,419	4,004	4,004

Notes: Table reports estimates for the effect of Head Start exposure on educational attainment. Columns 1 and 2 report estimates obtained from Table 2, Columns 2 and 4 in Jackson and Johnson (2019). Columns 3 and 4 report estimates obtained from Table 2.7 in this paper. Jackson and Johnson (2019) estimate the effect of having access to a Head Start center with the average county-level per-pupil Head Start spending level (versus having no access to a Head Start center) at age four on the likelihood of completing high school and years of education, for children who were exposed to average K-12 spending levels and whose average parental earnings between ages 12 and 17 fell in the bottom quartile of the earnings distribution. This paper estimates the effect of Head Start exposure between ages three and six on the likelihood of completing high school and years of completed education, for children whose parents did not graduate from high school. Jackson and Johnson's empirical specification includes: school district fixed effects, race-specific year of birth fixed effects, race by census division year of birth trends, a set of 1960 county characteristics interacted with a linear year of birth trend, race by timing of school desegregation fixed effects, and controls for family background and demographic characteristics (including parental education). Carter's (2019) empirical specification is described in Table 2.7. Standard errors are in parentheses. Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

within the Panel Study of Income Dynamics. For my analysis, I select a natural subsample - individuals born between 1950 and 1976 (either just before or during the roll-outs of Food Stamps and Head Start) whose parents did not graduate from high school. If the second possibility were true, this would indicate that the long-run impact of Head Start depends on the method used to construct the county of birth variable. The method that I use to construct the county of birth variable is ideal because (1) it addresses concerns about endogenous mobility in response to program introduction by assigning program exposure based on the county in which each survey respondent is born or, if the respondent is born prior to the first survey wave, to the earliest county of residence that we observe in the data, and (2) it exactly matches the methodology used by Hoynes, Schanzenbach and Almond (2016) in their study of the long-run impacts of childhood access to Food Stamps.

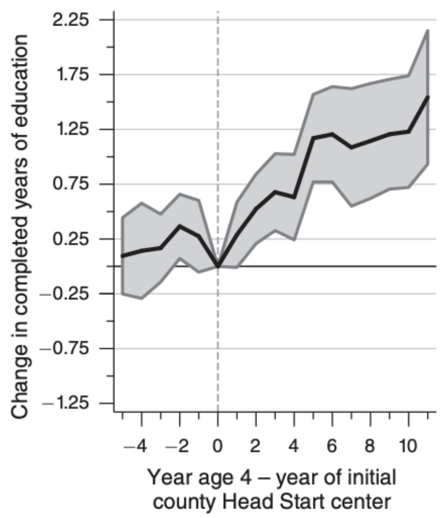
Figure 2.9: Replication Exercise for Jackson and Johnson (2019)

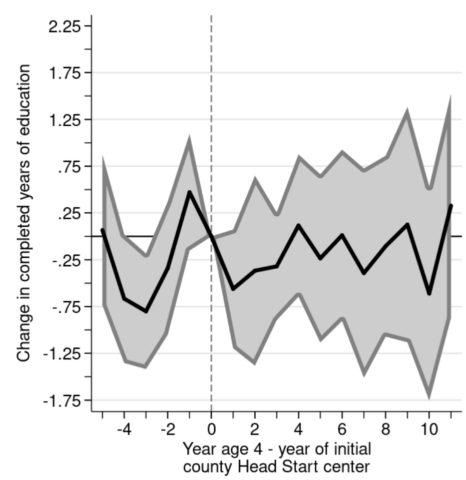




(a) JJ (2019): Low Spending Counties

(b) Replication Exercise: Low Spending Counties





(c) JJ (2019): High Spending Counties

(d) Replication Exercise: High Spending Counties

Notes: Figure reports event study estimates for the effect of Head Start exposure on educational attainment, separately for counties with low Head Start spending (Panels (a) and (b)) and high Head Start spending (Panels (c) and (d)). Panels (a) and (c) are copied directly from Figure 2 in Jackson and Johnson (2019). Panels (b) and (d) are produced from the replication exercise described in the text. Shaded regions in each plot represent 90% confidence intervals. Counties are defined as low or high Head Start spending based on 1980 per-pupil Head Start spending levels. See Jackson and Johnson (2019) for more details.

2.5 Results: Interaction Effects

Based on the estimates presented in the previous section, I conclude that (1) childhood exposure to Food Stamps reduces the incidence of metabolic syndromes and increases economic self-sufficiency in adulthood, and (2) preschool-age exposure to Head Start reduces the incidence of metabolic syndromes and worsens economic self-sufficiency in adulthood. This section asks whether the long-run effect of childhood exposure to Food Stamps depends on the availability of Head Start at preschool age. Specifically, I ask whether the effect of exposure to Food Stamps beginning at age τ on metabolic health and economic self-sufficiency in adulthood is a function of the availability of Head Start between ages three and six, for each age $\tau \in [-6, 11]$. To the best of my knowledge, these estimates have not been previously reported in the literature.

Before I present my estimates, I briefly review the possible values that the interaction effect between Food Stamps and Head Start can take and the implications of each value for public policy. In theory, the interaction effect between exposure to Food Stamps beginning at a particular age τ and Head Start between ages three and six could be positive, negative or zero. A positive value indicates that children with access to a Head Start center between ages three and six experience higher returns to Food Stamps exposure beginning at age τ, compared to children without access to a Head Start center between ages three and six[27]. A negative value indicates the opposite: that children with access to a Head Start center between ages three and six experience lower returns to Food Stamps exposure beginning at age τ, compared to children without access to a Head Start center between ages three and six[28]. For policymakers, complementarity between Food Stamps and Head Start (i.e., a positive value for the interaction effect) implies that a Food Stamps investment made at

27. Complementarity between Food Stamps and Head Start could arise if, for example, the nutritional investments that Food Stamps provides improve cognition, and improved cognition subsequently translates to higher returns from the educational investments that Head Start provides.

28. Substitutability between Food Stamps and Head Start could arise if, for example, Food Stamps and Head Start provide a similar set of nutritional investments to children.

age τ may be more effective if it is paired with a Head Start investment. On the other hand, substitutability between Food Stamps and Head Start (i.e., a negative value for the interaction effect) implies that a Food Stamps investment made at age τ may be more effective if it is targeted to children who miss out on the Head Start investment. A third possibility is that the interaction effect between Food Stamps and Head Start equals zero, in which case policymakers do not need to take into account heterogeneity in the return to Food Stamps based on access to Head Start when determining optimal policy.

2.5.1 Estimates for the Effect of Childhood Exposure to Food Stamps by the Availability of Head Start

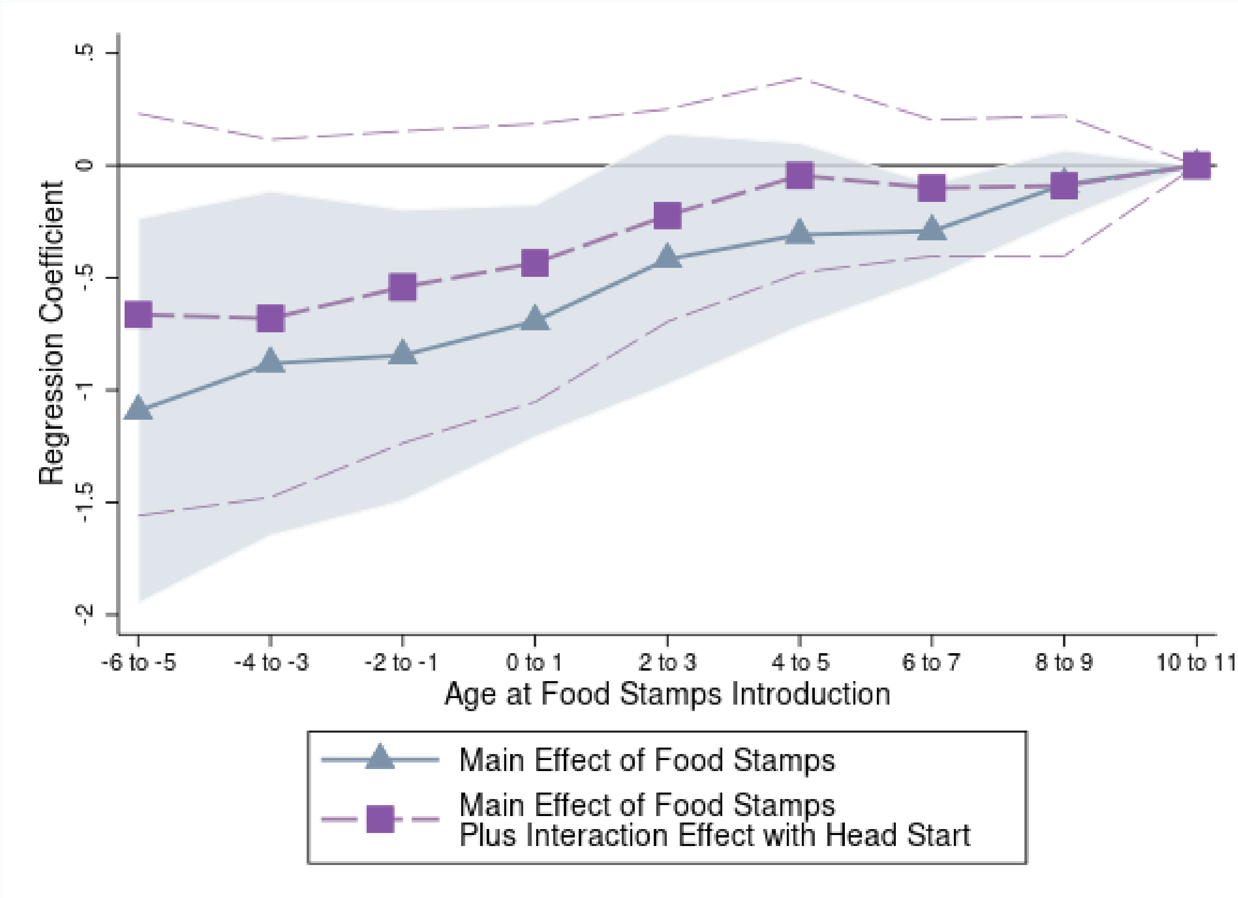
Figures 2.10 and 2.12 present the event study estimates from empirical model (2.3) for the two outcome variables of interest, the metabolic syndrome index and the economic self-sufficiency index. Figures 2.10 and 2.12 each contain two plots. The first plot displays the main effect of Food Stamps exposure beginning at age τ (α_τ in empirical model (2.3)) by the age at program introduction τ. The estimates in this plot measure the effect of Food Stamps exposure for children without access to a Head Start center between ages three and six. The second plot displays the sum of the main effect of Food Stamps exposure beginning at age τ and the interaction effect between Food Stamps exposure beginning at age τ and Head Start exposure between ages three and six ($\alpha_\tau + \beta_\tau$ in empirical model (2.3)) by the age at program introduction τ. The estimates in this plot measure the effect of Food Stamps exposure for children with full access to a Head Start center between ages three and six. For each age τ, the difference between the event study estimates in the first plot and the second plot (equal to β_τ in empirical model (2.3)) measures the interaction effect, or the additional return to Food Stamps exposure beginning at age τ (which may be negative) associated with having full access to a Head Start center between ages three and six.

I begin the discussion by reviewing the event study estimates from model (2.3) for the main effects of Food Stamps and Head Start, plus the interaction effects between the two programs, on the metabolic syndrome index. I present two sets of results. In the first set of results, I present the event study estimates for the main effects of Food Stamps and Head Start, respectively, on the metabolic syndrome index. The event study estimates in Figure 2.10 indicate that, regardless of whether Head Start is available between ages three and six, more cumulative years of exposure to Food Stamps during childhood (moving from right to left in the plot) leads to fewer metabolic syndromes in adulthood. Similarly, the event study estimates in Figure 2.11 indicate that more cumulative years of exposure to Head Start between ages three and six leads to fewer metabolic syndromes in adulthood. The estimates in Figures 2.10 and 2.11 are consistent with the findings presented in the previous section and are, for the most part, statistically insignificant at a 10% significance level (where the shaded or dashed regions in each figure represent the 90% confidence intervals for the event study estimates).

In the second set of results, I present the event study estimates for the interaction effects between Food Stamps and Head Start. I show that my estimates do not provide a clear indication that the effect of childhood exposure to Food Stamps on the metabolic syndrome index depends (either positively or negatively) on the availability of Head Start between ages three and six. Figure 2.10 plots the event study estimates for the effect of Food Stamps exposure by Head Start availability. Figure 2.10 shows that, for a subset of the ages $\tau \in [-6, 11]$, the effect of exposure to Food Stamps beginning at age τ on the metabolic syndrome index is similar for children with and without access to Head Start between ages three and six. Column 1 in Table 2.9 confirms this observation. Column 1 calculates the vertical distance between the two plots shown in Figure 2.10, which equals the estimated interaction effect between Food Stamps and Head Start at the indicated age τ. Column 1 shows that, for ages $\tau \in \{[8, 9]\}$, the estimated interaction effect is less than one-tenth of a standard deviation in magnitude. Further, none of the event study estimates reported in Figure 2.10 and Table 2.9 are statistically significant at a 10% significance level. As a result, I cannot rule out that the interaction effects equal zero at a 10% significance level.

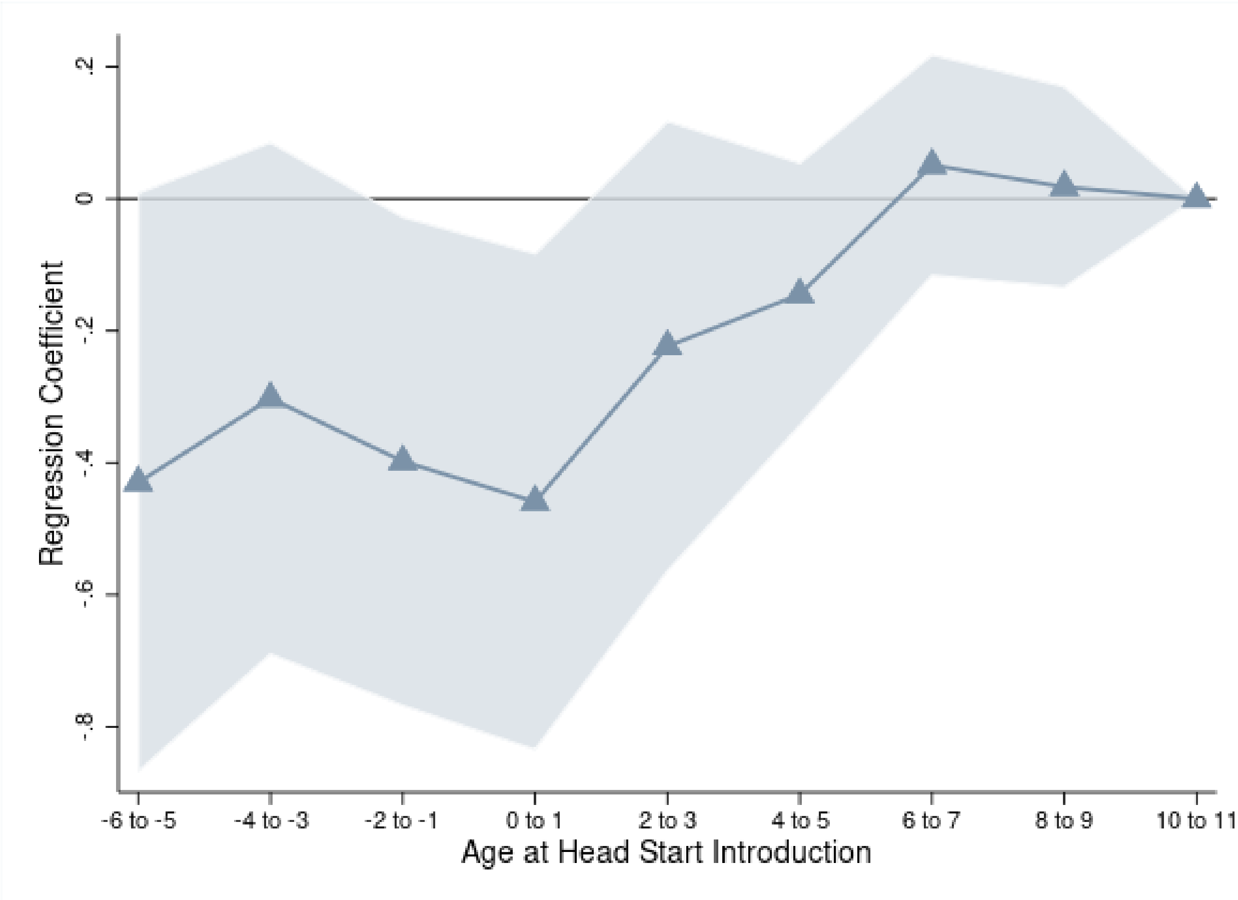
Table 2.10 provides the difference-in-differences estimates from model (2.4) for the effect of the following terms on the metabolic syndrome index and each of its components: Food Stamps exposure between the in-utero period and age two (before Head Start attendance), Food Stamps exposure between ages three and six (during Head Start attendance), Food Stamps exposure between ages seven and 11 (after Head Start attendance), Head Start exposure between ages three and six (during Head Start attendance), and the interaction between the Head Start term and each of the previously-listed Food Stamps terms. The estimated interaction effects in Column 1 range in magnitude from less than one-tenth of a standard deviation (0.07 for Food Stamps exposure between ages three and six) to between one-tenth and one-quarter of a standard deviation (-0.02 for Food Stamps exposure between the in-utero period and age two; 0.23 for Food Stamps exposure between ages seven and 11). In all cases, the estimated interaction effects are not statistically different from zero at a 10% significance level. Additionally, each of the estimated interaction effects differs in sign across the components of the metabolic syndrome index, so that there is no consistent evidence of either a positive or negative interaction effect between Food Stamps and Head Start.

Figure 2.10: Effect of Food Stamps on Metabolic Syndrome Index by Head Start Availability



Notes: Figure plots event study estimates for (1) the main effect of Food Stamps and (2) the main effect of Food Stamps plus the interaction effect between Food Stamps and Head Start on the metabolic syndrome index by the age at the time of Food Stamps introduction. Figure plots event study estimates with 90% confidence intervals. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level.

Figure 2.11: Effect of Head Start on Metabolic Syndrome Index



Notes: Figure plots event study estimates for the main effect of Head Start by the age at the time of Head Start introduction. Figure plots event study estimates with 90% confidence intervals. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level.

Table 2.9: Event Study Estimates for Interaction Effects between Food Stamps and Head Start

Age Range	Metabolic Syndrome Index	Economic Self-Sufficiency Index
-6 to -5	0.423	-0.180
	(0.313)	(0.360)
-4 to -3	0.201	-0.246
	(0.273)	(0.342)
-2 to -1	0.304	0.140
	(0.244)	(0.313)
0 to 1	0.259	0.094
	(0.220)	(0.296)
2 to 3	0.195	-0.109
	(0.301)	(0.245)
4 to 5	0.264	-0.099
	(0.215)	(0.239)
6 to 7	0.192	-0.043
	(0.172)	(0.272)
8 to 9	-0.006	0.042
	(0.202)	(0.252)
10 to 11	0	0

Notes: Table reports event study estimates for the interaction effect between Food Stamps exposure beginning at the indicated age range and Head Start exposure between ages three and six. In Column 1, the outcome variable of interest is the metabolic syndrome index. In Column 2, the outcome variable of interest is the economic self-sufficiency index. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of 1960 county characteristics interacted with a linear year of birth trend. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights.

Table 2.10: Effect of Food Stamps on Metabolic Syndrome Index by Head Start Availability

		Components of Index				
	Metabolic Syndrome Index	High Blood Pressure	Diabetes	Heart Attack	Heart Disease	Obesity
Head Start 3-6	-0.391	-0.519**	0.0139	-0.00775	-0.0317	-0.312
	(0.302)	(0.183)	(0.191)	(0.0284)	(0.0505)	(0.264)
Food Stamps IU-2	-0.213*	-0.221**	-0.0763	-0.00715	0.0170	-0.107
	(0.128)	(0.0932)	(0.0832)	(0.0171)	(0.0328)	(0.124)
Food Stamps 3-6	0.0259	0.135*	0.0511	-0.0252	-0.0244	-0.0686
	(0.151)	(0.0692)	(0.0979)	(0.0211)	(0.0327)	(0.121)
Food Stamps 7-11	-0.123	-0.211**	-0.0705	0.0247	0.0257	-0.132
	(0.0989)	(0.0769)	(0.0568)	(0.0235)	(0.0370)	(0.0875)
Head Start 3-6 X Food Stamps IU-2	-0.0178	0.104	-0.00839	-0.0156	-0.0325	-0.0820
	(0.122)	(0.0782)	(0.0708)	(0.0209)	(0.0368)	(0.137)
Head Start 3-6 X Food Stamps 3-6	0.0734	-0.0421	0.0658	0.0344	0.0221	0.0619
	(0.184)	(0.105)	(0.109)	(0.0339)	(0.0465)	(0.154)
Head Start 3-6 X Food Stamps 7-11	0.227	0.412**	-0.0573	-0.00750	0.0287	0.158
	(0.284)	(0.176)	(0.190)	(0.0340)	(0.0521)	(0.240)
Mean of Y	0.0784	0.271	0.0846	0.0178	0.0410	0.358
Number of Person-Years	13,476	13,826	13,828	13,828	13,824	15,139
Number of Persons	2,080	2,090	2,090	2,090	2,089	2,874

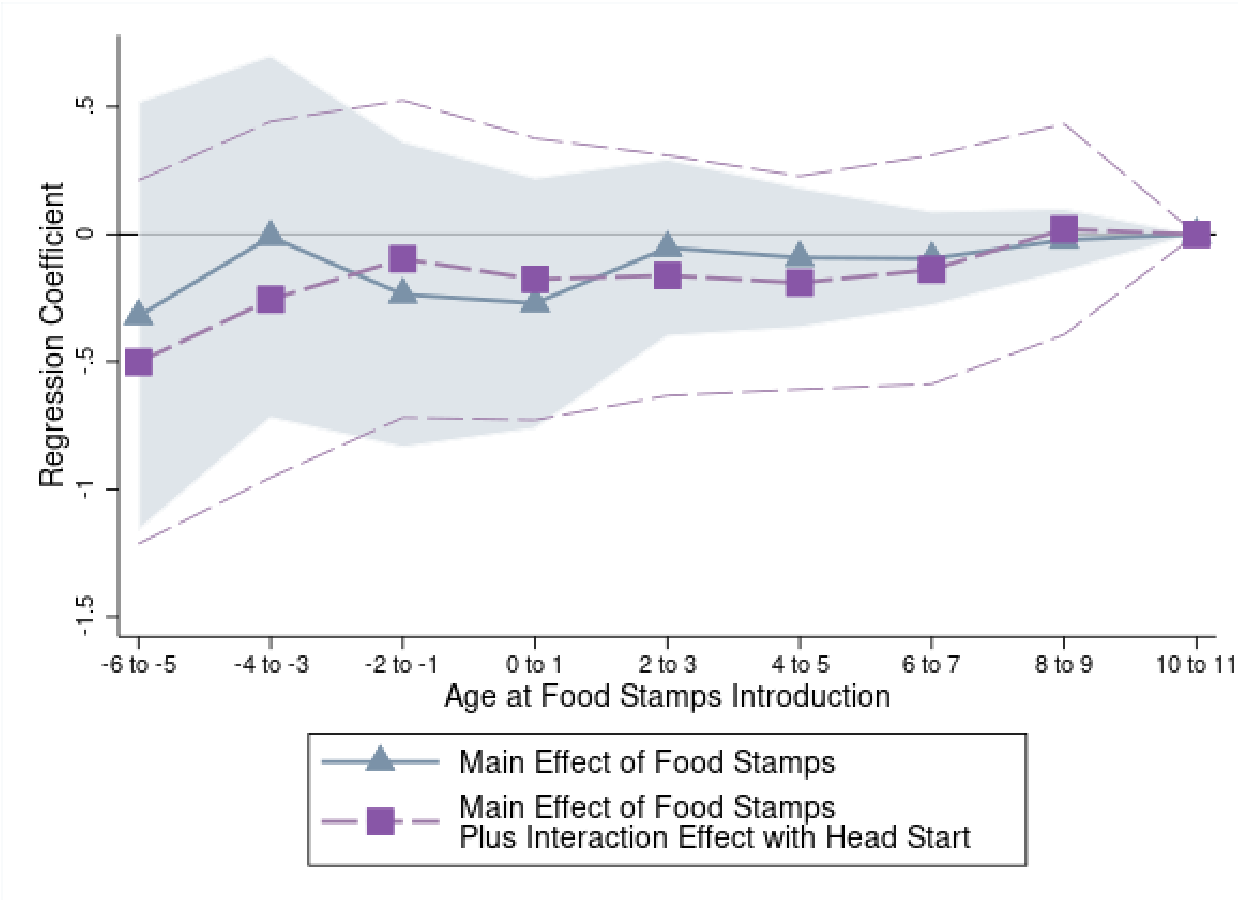
Notes: Table reports difference-in-differences estimates for the main effects and the interaction effects between Food Stamps exposure and Head Start exposure. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level. Standard errors are in parentheses. Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Next, I review the event study estimates from model (2.3) for the main effects of Food Stamps and Head Start, plus the interaction effects between the two programs, on the economic self-sufficiency index. Again, I present two sets of results. In my first set of results, I describe the main effects of Food Stamps and Head Start, respectively, on the economic self-sufficiency index. The event study estimates in Figure 2.12 indicate that exposure to Food Stamps beginning prior to age ten has a small, negative impact on economic self-sufficiency in adulthood; however, these estimates are not statistically significant and I cannot rule out a positive impact of Food Stamps exposure prior to age ten on the economic self-sufficiency index at a 10% significance level. The event study estimates in Figure 2.13 indicate that exposure to Head Start between ages three and six leads to a reduction in economic self-sufficiency in adulthood. None of the estimates in in Figure 2.13 are statistically significant at a 10% significance level. All of the estimates in Figures 2.12 and 2.13 are consistent with the findings presented in the previous section.

In the second set of results, I review the interaction effects between Food Stamps and Head Start on the economic self-sufficiency index. I show that my event study estimates do not provide a clear indication that the effect of childhood Food Stamps exposure on the economic self-sufficiency index depends on the availability of Head Start. Figure 2.12 plots the event study estimates for the effect of exposure to Food Stamps on the economic self-sufficiency index by Head Start availability. Children without access to Head Start experience similar returns to Food Stamps compared to children with access to Head Start, for most ages $\tau \in [-6, 11]$. Column 2 in Table 2.9 shows that this is the case. Column 2 calculates the vertical distance between the two plots shown in Figure 2.12, which equals the estimated interaction effect between Food Stamps and Head Start at the indicated age τ. Column 2 shows that, for ages τ in $[0, 9]$, the estimated interaction effects are (approximately) less than one-tenth of a standard deviation in magnitude. In addition, all of the event study estimates reported in Figure 2.12 and Table 2.9 are statistically insignificant at a 10% significance level. As a result, I cannot rule out that the interaction effects equal zero at a 10% significance level.

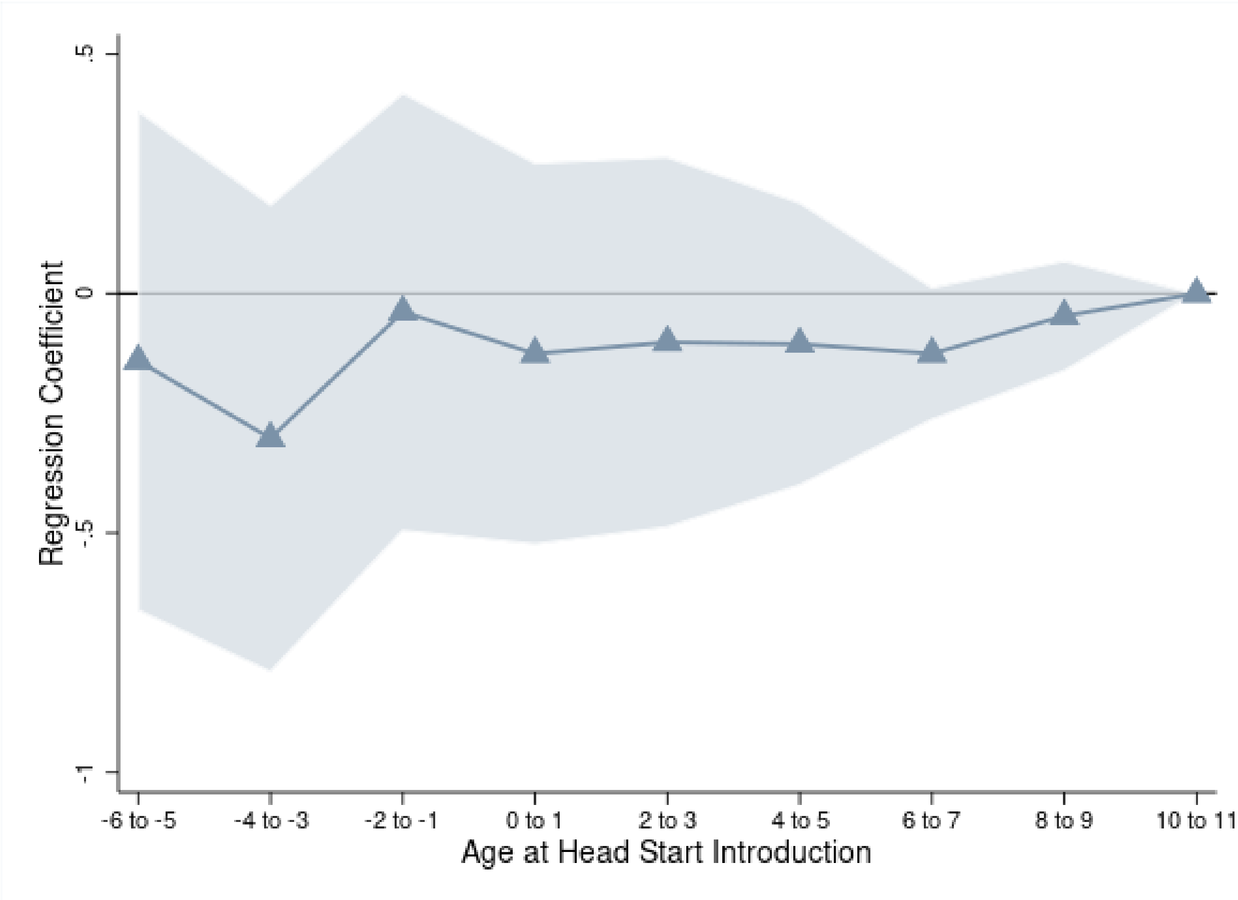
Table 2.11 provides the difference-in-differences estimates from model (2.4) for the effect of the following terms on the economic self-sufficiency index and each of its components: Food Stamps exposure between the in-utero period and age two (before Head Start attendance), Food Stamps exposure between ages three and six (during Head Start attendance), Food Stamps exposure between ages seven and 11 (after Head Start attendance), Head Start exposure between ages three and six (during Head Start attendance), and the interaction between the Head Start term and each of the previously-listed Food Stamps terms. The estimated interaction effects are less than one-tenth of a standard deviation in magnitude for Food Stamps exposure between ages in-utero and two (0.01 standard deviations), and for Food Stamps exposure between ages three and six (0.06 standard deviations). In all cases, the estimated interaction effects are not statistically different from zero at a 10% significance level, and the signs of the interaction effects differ across the components of the economic self-sufficiency index.

Figure 2.12: Effect of Food Stamps on Economic Self-Sufficiency Index by Head Start Availability



Notes: Figure plots event study estimates for (1) the main effect of Food Stamps and (2) the main effect of Food Stamps plus the interaction effect between Food Stamps and Head Start on the economic self-sufficiency index by the age at the time of Food Stamps introduction. Figure plots event study estimates with 90% confidence intervals. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level.

Figure 2.13: Effect of Head Start on Economic Self-Sufficiency Index



Notes: Figure plots event study estimates for the main effect of Head Start by the age at the time of Food Stamps introduction. Figure plots event study estimates with 90% confidence intervals. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level.

Table 2.11: Effect of Food Stamps on Economic Self-Sufficiency Index by Head Start Availability

	Economic Self-Sufficiency Index	Components of Index					
		Employed	Personal Earnings	Family Earnings	Above Poverty Line	High School Plus	Years of Education
Head Start 3-6	-0.000126	0.0823	4,506.2	–9,984.7	0.0373	0.0692	-0.852
	(0.254)	(0.0954)	(5,279.3)	(9,641.1)	(0.0836)	(0.123)	(0.783)
Food Stamps IU-2	0.240	-0.00844	7,696.3	10,959.4	0.0147	0.0275	0.747**
	(0.149)	(0.0745)	(6,501.1)	(7,771.0)	(0.0634)	(0.0586)	(0.356)
Food Stamps 3-6	0.0497	0.0804	-3,1987.8	-9,367.5**	0.0500	0.103**	-0.0540
	(0.101)	(0.0545)	(2,907.1)	(4,735.2)	(0.0417)	(0.0415)	(0.263)
Food Stamps 7-11	0.306**	0.125**	5,008.4**	8,763.7**	0.115**	0.128**	0.200
	(0.0884)	(0.0454)	(2,511.1)	(3,585.0)	(0.0338)	(0.0445)	(0.265)
Head Start 3-6 X Food Stamps IU-2	0.0113	0.0879	-2,019.0	-5,746.7	0.0198	0.0731	-0.486
	(0.141)	(0.0624)	(5,376.7)	(7,701.8)	(0.0674)	(0.0644)	(0.419)
Head Start 3-6 X Food Stamps 3-6	0.0639	-0.0817	-2,341.5	10,178.2	0.0161	-0.0203	-0.136
	(0.145)	(0.0842)	(3,670.9)	(6,580.2)	(0.0639)	(0.0708)	(0.453)
Head Start 3-6 X Food Stamps 7-11	-0.221	-0.145	-6,107.0	2,730.0	-0.124	-0.140	0.591
	(0.296)	(0.112)	(5,918.9)	(9,101.8)	(0.0976)	(0.138)	(0.920)
Mean of Y	-0.245	0.761	22,018.3	46,449.1	0.824	0.867	12.98
Number of Person-Years	47,024	48,332	47,663	56,857	56,927	-	-
Number of Persons	3,642	3,655	3,643	4,128	4,128	4,004	4,004

Notes: Table reports difference-in-differences estimates for the main effects and the interaction effects between Food Stamps exposure and Head Start exposure. The empirical model includes county of birth fixed effects, year of birth fixed effects, survey year fixed effects, sex, race and a set of time-varying county of birth characteristics. The sample includes PSID survey respondents born between 1950 and 1976 whose parents did not complete high school. Regressions are weighted using PSID survey weights. Standard errors are clustered at the county of birth-by-county of residence level. Standard errors are in parentheses. Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

2.6 Conclusion

Participation in multiple safety net programs during the course of childhood is common, yet little is known about how safety net programs accessed at different points during childhood interact to affect long-run outcomes. In this paper, I study two key components of the social safety net in the United States, Food Stamps and Head Start. I ask whether the effect of childhood Food Stamps exposure on labor market and health outcomes in adulthood varies depending on the availability of Head Start at preschool age. To answer this question, I estimate event study and difference-in-differences models that exploit variation in childhood exposure to Food Stamps and Head Start induced by the county-level roll-outs of the two programs during the 1960s and 1970s. I find that, among children whose parents did not complete high school, childhood access to Food Stamps reduces the incidence of metabolic syndromes and increases economic self-sufficiency in adulthood, while preschool-age access to Head Start reduces the incidence of metabolic syndromes and has a negative impact on economic self-sufficiency in adulthood. Using the same sample, I find no clear evidence that the return to childhood access to Food Stamps depends on the availability of Head Start at preschool age, for either the metabolic syndrome index or the economic self-sufficiency index. However, the estimated interaction effects that I obtain are statistically insignificant at a 10% level, so that I cannot rule out non-zero interaction effects between the two programs with any reasonable degree of certainty.

My results raise the question of whether I do not detect clear evidence of interaction effects between Food Stamps and Head Start because (1) Food Stamps and Head Start are complements or substitutes in the production of human capital, but I lack the statistical power to estimate non-zero interaction effects with precision, or (2) Food Stamps and Head Start are neither complements nor substitutes in the production of human capital. To address the first possibility, future work should aim to estimate the interaction effects between Food Stamps and Head Start more precisely by combining my empirical strategy with a larger administrative data set that links information on individuals' outcomes in adulthood to their county and year of birth. To address the second possibility, future work should incorporate additional data on parental investment responses to Food Stamps and Head Start in order to isolate the direct impact of Head Start-induced increases in human capital on the return to Food Stamps.

CHAPTER 3

REDUCING INEQUALITY THROUGH DYNAMIC COMPLEMENTARITY: EVIDENCE FROM HEAD START AND PUBLIC SCHOOL SPENDING: COMMENT

3.1 Introduction

Several recent papers in the economics literature evaluate the long-run effects of the Head Start program on children's outcomes by pairing quasi-experimental variation in preschool-age Head Start availability with panel data. One example is Jackson and Johnson (2019)[1]. Jackson and Johnson estimate event study and difference-in-differences models that compare long-run outcomes for poor children who were quasi-randomly exposed to different levels of age-four Head Start spending. They produce their estimates using a nationally representative sample from the Panel Study of Income Dynamics. Their identification strategy exploits exogenous variation in age-four Head Start spending induced by the county-level roll-out of Head Start during the 1960s and 1970s. Based on their estimates, they conclude that higher levels of age-four Head Start spending improve poor children's educational and labor market outcomes in adulthood.

In this paper, I show that I am unable to reproduce Jackson and Johnson's estimates for the effect of age-four Head Start spending on poor children's outcomes in adulthood. In my main set of results, I perform a replication exercise where I attempt to reproduce a set of event study estimates from Jackson and Johnson. I produce the replication sample using the data sets and code files that Jackson and Johnson include with their published paper, combined with a childhood county of residence variable that I construct. I show that the event study estimates from the replication exercise do not match the estimates published by Jackson and Johnson. Jackson and Johnson's published estimates indicate that poor children from high

1. Other examples include Ludwig and Miller (2007), Thompson (2017), and Bailey et. al. (2018).

Head Start spending counties experience a statistically significant increase in educational attainment following the introduction of Head Start. In contrast, the estimates from the replication exercise indicate that poor children experience no statistically significant change in educational attainment following the introduction of Head Start, in either high or low Head Start spending counties.

In a second set of results, I perform an empirical exercise where I use the replication sample to estimate a pair of difference-in-differences models that are similar (though not identical) to a pair of difference-in-differences models estimated by Jackson and Johnson. I show that the estimates from the empirical exercise are inconsistent with the estimates published by Jackson and Johnson, despite the fact that both sets of estimates are identified by a similar source of variation in Head Start spending.

3.1.1 Organization

This paper proceeds as follows. In Section 2, I describe Jackson and Johnson's data and empirical methods. In Section 3, I perform a replication exercise. I estimate an event study model that is identical to an event study model estimated by Jackson and Johnson. I compare the results of the replication exercise to Jackson and Johnson's published results. In Section 4, I perform an empirical exercise. I estimate a pair of difference-in-differences models that are similar (though not identical) to a pair of models estimated by Jackson and Johnson. I compare the results of the empirical exercise to Jackson and Johnson's published results. I also compare Jackson and Johnson's published results to similar estimates from the literature. In Section 5, I conclude.

3.2 Description of Jackson and Johnson (2019)

In this section, I describe the data set and empirical methods used by Jackson and Johnson.

3.2.1 Data

Jackson and Johnson construct their analysis sample using data from the Panel Study of Income Dynamics (PSID). The PSID is a nationally representative, longitudinal survey that began in 1968 with a sample of 5,000 households. Since the survey's inception, members of the original households and their descendants have been interviewed annually or bi-annually. The PSID collects information on a wide range of outcomes for survey respondents, including earnings, educational attainment, and incarceration status. Jackson and Johnson restrict their analysis sample to individuals born between 1950 and 1976, and include all observations from survey years 1968 through 2015 for which the individual is between the ages of 20 and 50. Jackson and Johnson divide their analysis sample into a poor child subsample and a non-poor child subsample based on childhood family earnings, where the poor child subsample consists of individuals whose average family earnings between ages 12 and 17 fell in the bottom quartile of the family earnings distribution. They produce their main set of empirical results using the poor child subsample.

Jackson and Johnson collect county-level data on annual Head Start expenditures between 1965 and 1980. Using this data, they determine Head Start availability and average Head Start spending per poor four-year-old (measured in $2000) at age four in the childhood county of residence for each individual in their analysis sample. Jackson and Johnson collect additional local area-level data, including school district-level K-12 public school spending; a database of state-level, court-ordered school finance reforms; county-level demographic characteristics from the Census; and county-level social safety net program expenditures.

Geographic Variables

The PSID collects three geographic variables that can be used to identify an individual's childhood county of residence: (1) the county in which the survey respondent resides in each survey year, (2) retrospective self-reports of the county in which the survey respondent was born, and (3) retrospective self-reports of the county in which the survey respondent grew up. For each individual in the analysis sample, Jackson and Johnson construct a childhood county of residence variable equal to the earliest-observed county of residence between ages 0 and 17. If the individual is not observed during childhood, Jackson and Johnson assign childhood county of residence using retrospective self-reports of the county in which the individual grew up[2].

In their paper, Jackson and Johnson obtain access to an additional, restricted-use geographic variable that identifies the Census block in which each survey respondent resides in each survey year. They use this variable, in combination with historical data on school district boundaries and GIS software, to identify the childhood school district of residence for each individual in their analysis sample. Unlike Jackson and Johnson, I do not have access to the restricted-use Census block variable. As a result, I am unable to (and do not attempt to) replicate any results in their paper that depend on the childhood school district of residence.

3.2.2 Empirical Methods

Jackson and Johnson estimate a set of event study and difference-in-differences models in their paper. In the remainder of this section, I describe these models in detail.

Event Study Models

Jackson and Johnson produce event study estimates for the effect of age-four Head Start exposure on two outcome variables of interest: county-level Head Start spending per poor four-year-old at age four and educational attainment in adulthood. In the discussion that follows, Jackson and Johnson define high [low] Head Start spending counties as counties with

2. I determine that this is the approach that Jackson and Johnson use to construct the childhood county of residence variable based on the written description in their paper, online appendix, and code files. See Appendix B.2 for sources.

top [bottom] quartile Head Start spending per poor four-year-old in 1980[3].

To measure the effect of age-four Head Start exposure on county-level Head Start spending per poor four-year-old at age four, Jackson and Johnson estimate the following model on the poor child subsample. They estimate this model separately for children from high versus low Head Start spending counties:

$$Y_{icb} = \sum_{\alpha} \beta_{\alpha}^{ES} HS_{cb}^{\alpha} + \theta_c + \epsilon_{icb} \tag{3.1}$$

Y_{icb} is county-level Head Start spending per poor four-year-old at age four, for an individual i who is from childhood county c and born in year b. HS_{cb}^{α} is an indicator variable equal to one if the cohort that is from childhood county c and born in year b turns age four α years after (or before, if α is negative) the introduction of Head Start in county c. θ_c is a childhood county fixed effect. Jackson and Johnson cluster standard errors at the childhood county level and weight the regression using PSID survey weights.

To measure the effect of age-four Head Start exposure on educational attainment in adulthood among children from high [low] Head Start spending counties, Jackson and Johnson estimate the following model on the poor child subsample:

$$Y_{icb} = \sum_{\alpha} \eta_{\alpha}^{ES} OtherSpend_c \times HS_{cb}^{\alpha} + \sum_{\alpha} \beta_{\alpha}^{ES} HS_{cb}^{\alpha} + \gamma C_{icb} + \theta_c + \epsilon_{icb} \tag{3.2}$$

$OtherSpend_c$ is an indicator variable equal to one if county-level Head Start spending per poor four-year-old is not in the top [bottom] quartile in 1980. C_{icb} is a vector of individual-level and childhood county-level control variables that includes: race by Census division-specific year of birth trends, a set of 1960 county characteristics (education, percent urban, log population, percent black, percent black interacted with race) interacted with a linear year of birth trend, a set of demographic characteristics (race, gender, parental education,

3. Top quartile Head Start spending corresponds to Head Start spending per poor four-year-old greater than \$1800 (in \$2000). Bottom quartile Head Start spending corresponds to Head Start spending per poor four-year-old below \$1500 (in \$2000).

parental marital status, childhood poverty status, birth weight, age at most recent survey), and a set of dummy variables that indicate missing safety net program (Medicaid, AFDC, Food Stamps, unemployment insurance) expenditures data[4]. All other terms share the same definition as the corresponding terms in model (3.1). Jackson and Johnson cluster standard errors at the childhood county level and weight the regression using PSID survey weights.

Models (3.1) and (3.2) are not directly specified in the text of Jackson and Johnson's paper. Instead, I specify these models based on the code that is provided in the replication files that they include with their published paper. Appendix B.1 provides the relevant excerpts of code from the replication files.

Difference-in-Differences Models

Jackson and Johnson produce difference-in-differences estimates for the effect of age-four Head Start spending on the following outcome variables in adulthood: high school graduation, years of completed education, log wages in adulthood, the incidence of poverty in adulthood, and whether the individual has ever been incarcerated. They do so by estimating a pair of models that include terms for age-four Head Start spending, K-12 public school spending, and the interaction between Head Start spending and K-12 public school spending.

Jackson and Johnson begin by estimating the following model, which uses exposure to court-ordered school finance reforms as an instrument for K-12 public school spending. They estimate this model on the poor child subsample:

$$Y_{icb} = \beta_{HS} HS_{cb}^{age4} + \beta_{K-12} \widehat{ppe}_{idb}^{5-17} + \beta_{int} \widehat{INT}_{idb} + \gamma C_{icb} + \theta_d + \tau_b + \epsilon_{idb} \tag{3.3}$$

4. The control variables in vector C_{icb} match the control variables that Jackson and Johnson include in model (3.2) according to the code files that they provide with their published paper. They do not match the control variables that Jackson and Johnson state that they include in model (3.2) in the text of their paper. Jackson and Johnson state that they include the following control variables: race by Census division-specific year of birth trends, a set of 1960 county characteristics (poverty rate, percent black, education, percent urban, population size) interacted with a linear year of birth trend, and a set of demographic characteristics (parental income and education, mother's marital status at birth, birth weight, gender) [see table notes for Table 2 in Jackson and Johnson (2019), p. 323].

Y_{icb} is the outcome variable for an individual i who is from childhood county c and born in year b. HS_{cb}^{age4} measures county-level Head Start spending per poor four-year-old at age four, for the cohort from childhood county c and born in year b. ppe_{idb}^{5-17} is the natural log of school district-level, per-pupil K-12 public school spending from ages 5 to 17, for the cohort from childhood school district d and born in year b. INT_{idb} is the interaction between HS_{cb}^{age4} and ppe_{idb}^{5-17}. $\widehat{dose}_d$ is the predicted change in K-12 public school spending in school district d that results from the state's court-ordered school finance reform. $SFRExp_{idb}$ is the number of years that individual i is exposed to the state's court-ordered school finance reform. $\widehat{ppe}_{idb}^{5-17}$ and $\widehat{INT}_{idb}$ are the fitted values from first stage regressions for ppe_{idb}^{5-17} and INT_{idb}, which include as instruments all possible interactions between HS_{icb}^{age4}, $\widehat{dose}_d$, and $SFRExp_{idb}$ (with the exception of $HS_{icb}^{age4} \times \widehat{dose}_d$). C_{icb} is a vector of individual-level and childhood county-level control variables that includes: race by year of birth fixed effects; race by Census division-specific year of birth trends; a set of county-level controls for race by timing of hospital desegregation, race by years of exposure to school desegregation, and years of exposure to community health centers; a state-level control for years of exposure to kindergarten; a set of 1960 county characteristics (education, percent urban, log population, percent black, percent black interacted with race) interacted with a linear year of birth trend; a set of demographic characteristics (race, gender, parental education, parental marital status, childhood poverty status, birth weight, age at most recent survey); a set of controls for county-level safety net program (Medicaid, AFDC, Food Stamps, unemployment insurance) expenditures; and, for wage and poverty regressions only, a cubic in age-by-gender and a control for pregnancy in the last five years. θ_d is a childhood school district fixed effect and τ_b is a year of birth fixed effect. Jackson and Johnson cluster standard errors at the childhood state level and weight the regression using PSID survey weights.

Next, Jackson and Johnson estimate the following model, which uses exposure to a Head Start center during the program's roll-out as an instrument for Head Start spending and exposure to court-ordered school finance reforms as an instrument for K-12 public school spending. They estimate this model on the poor child subsample:

$$Y_{icb} = \beta_{HS}\widehat{HS_{cb}^{age4}} + \beta_{K-12}\widehat{ppe}_{idb}^{5-17} + \beta_{int}\widehat{INT}_{idb} + \gamma C_{icb} + \theta_d + \tau_b + \epsilon_{idb} \tag{3.4}$$

$Exposed\_HS_{cb}^{age4}$ is an indicator variable equal to one if Head Start is available at age four, for the cohort from childhood county c and born in year b. $\widehat{HS_{cb}^{age4}}$, $\widehat{ppe}_{idb}^{5-17}$ and $\widehat{INT}_{idb}$ are the fitted values from first stage regressions for HS_{cb}^{age4}, ppe_{idb}^{5-17} and INT_{idb}, which include as instruments all possible interactions between $Exposed\_HS_{cb}^{age4}$, $\widehat{dose}_d$, and $SFRExp_{idb}$ (with the exception of $Exposed\_HS_{cb}^{age4} \times \widehat{dose}_d$). All other terms in the model share the same definition as the corresponding terms in model (3.3). Jackson and Johnson cluster standard errors at the childhood state level and weight the regression using PSID survey weights.

3.3 Replication Results

In this section, I replicate a set of event study estimates from Jackson and Johnson. I describe the results of my replication exercise, and I compare my replication estimates to the estimates published by Jackson and Johnson.

3.3.1 Summary Statistics

I begin by briefly describing the methodology that I use to replicate the analysis sample from Jackson and Johnson. Greater detail on this methodology is provided in Appendix B.2. Jackson and Johnson include with their published paper most of the data sets and code files that they use to produce their analysis sample[5]. However, they do not provide, nor do they include a complete set of code files to produce, the childhood county of residence

5. Jackson and Johnson's data sets and code files can be downloaded here: https://www.aeaweb.org/articles?id=10.1257/pol.20180510

variable[6]. As a result, I replicate the analysis sample from Jackson and Johnson in two steps. First, I construct the childhood county of residence variable according to the procedure that Jackson and Johnson describe in their published paper, online appendix, and code files. Second, I produce the analysis sample using the data sets and code files that Jackson and Johnson include with their published paper, in combination with the childhood county of residence variable that I constructed in the first step[7].

Table 3.1 presents the summary statistics from the replication sample (Columns 2 and 4), alongside the summary statistics published by Jackson and Johnson (Columns 1 and 3). The summary statistics are reported separately for the poor child subsample and the non-poor child subsample. Table 3.1 shows that the replication sample is not identical to the analysis sample that Jackson and Johnson use to produce their published summary statistics. Several discrepancies between the two samples should be noted. First, the poor child subsample is more female and less white in the replication sample compared to Jackson and Johnson's analysis sample: 50 percent of the poor child subsample is female in the replication sample, compared to 43 percent in Jackson and Johnson's analysis sample; 57 percent of the poor child subsample is white in the replication sample, compared to 66 percent in Jackson and Johnson's analysis sample. Second, the share of the sample that is born into a two-parent family appears to be much lower in the replication sample compared to Jackson and Johnson's analysis sample: 54 percent of the poor child subsample is born into a two-parent family in the replication sample, compared to 74 percent in Jackson and

6. Jackson and Johnson cannot directly provide any geographic variables used in their analysis because these variables can only be accessed after establishing a restricted-use data contract with the PSID. The fact that the childhood county of residence variable cannot be constructed using the code files that Jackson and Johnson provide is problematic because the childhood county of residence variable is key for defining the analysis sample (survey respondents for whom childhood county of residence cannot be assigned are dropped from the analysis sample).

7. I make several modifications to the code files that Jackson and Johnson include with their published paper. I do so because some of the data sets and variables that Jackson and Johnson reference in their code files are neither provided by Jackson and Johnson nor produced by the code files that they do provide. I describe each of these modifications in Appendix B.2.

Johnson's analysis sample; 82 percent of the non-poor child subsample is born into a two-parent family in the replication sample, compared to 95 percent in Jackson and Johnson's analysis sample. However, this discrepancy arises due to a coding error made by Jackson and Johnson. They mistakenly calculate the share of the sample that is born into a two-parent family using only the subsample of respondents for whom the unrelated low birth weight variable is non-missing. Third, and perhaps most strikingly, the replication sample consists of 10,425 individuals, while Jackson and Johnson's analysis sample consists of 15,232 individuals.

Before proceeding, it is important to note that there is some ambiguity regarding the correct weight variables to use in the replication exercise (both for producing the summary statistics and for producing the regression results). Jackson and Johnson do not provide, nor do they include a code file to produce, the weight variables that they use to produce the results in their empirical analysis. For the remainder of this paper, I assume that the weight variable is equal to either the PSID survey weight from the most recent survey wave (for person-level analyses), or the PSID survey weight in a given survey year (for person by survey year-level analyses).

Table 3.1: Summary Statistics

	Poor Child		Non-Poor Child	
	Published	Replication	Published	Replication
Adult outcomes				
High school graduate	0.71	0.68	0.89	0.89
Years of education	12.29	12.24	13.61	13.82
ln(wages), at age 30	2.24	2.17	2.56	2.53
Adult family income, at age 30	$35,372	$32,428	$52,448	$53,054
In poverty, at age 30	0.18	0.21	0.05	0.06
Ever incarcerated	0.08	0.08	0.04	0.04
Age (range: 20-50)	30.3	32.5	31.0	33.1
Year born (range: 1950-1976)	1962	1962	1962	1962
Female	0.43	0.5	0.44	0.49
White	0.66	0.57	0.93	0.9
Childhood school variables				
Any Head Start center in county, age 4	0.33	0.29	0.34	0.31
Post-rollout: Head Start spending per poor 4-year-old, age 4	$4,204	$4,335	$4,072	$4,306
Child attended Head Start	0.19	0.18	0.02	0.04
Child attended any preschool program	0.31	0.28	0.23	0.26
Childhood family variables				
Income (average, ages 12-17)	$22,520	$20,043	$65,130	$65,370
Income-to-needs ratio (average, ages 12-17)	1.31	1.14	3.62	3.6
Mother's years of education	10.61	10.23	12.24	12.49
Father's years of education	10.04	9.19	12.36	12.73
Born into two-parent family	0.74	0.54	0.95	0.82
Low birth weight (<5.5 pounds)	0.07	0.08	0.07	0.06
Observations	6,373	4,480	8,859	5,945

Notes: Columns 1 and 3 report summary statistics obtained by Jackson and Johnson (2019) (see Table 1, p. 322 of Jackson and Johnson (2019)). Columns 2 and 4 report summary statistics obtained from the replication sample. Summary statistics are reported for the the poor child subsample and the non-poor child subsample. Poor children are defined as children for whom average family earnings between the ages of 12 and 17 fell in the bottom quartile of the family earnings distribution. Summary statistics are produced using PSID sample weights.

3.3.2 Event Study Estimates

I replicate a set of event study estimates reported by Jackson and Johnson. I do so by estimating models (3.1) and (3.2) using the poor child subsample of the replication sample. These event study estimates measure the effect of Head Start exposure on county-level Head Start spending per poor four-year-old at age four and educational attainment in adulthood, respectively. In the discussion that follows, I compare my replication estimates to the estimates published by Jackson and Johnson. In Figures 3.1, 3.2, and 3.3, Panel (a) plots the event study estimates published by Jackson and Johnson, while Panel (b) plots the event study estimates that I obtain from the replication exercise.

Figure 3.1 plots the event study estimates for the effect of Head Start exposure on county-level Head Start spending per poor four-year-old at age four (measured in $2000), separately for individuals from counties with high versus low levels of Head Start spending. Recall, Jackson and Johnson define high Head Start spending counties as counties with top quartile Head Start spending in 1980, and they define low Head Start spending counties as counties with bottom quartile Head Start spending in 1980. The replication estimates in Panel (b) follow a similar pattern as, but are not identical to, the published estimates in Panel (a): in both plots, county-level Head Start spending per poor four-year-old at age four increases following Head Start introduction, and the increase in Head Start spending is largest for individuals from high Head Start spending counties.

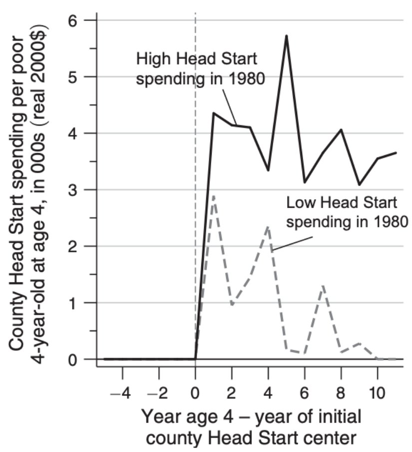
Figure 3.2 plots the event study estimates for the effect of Head Start exposure on educational attainment in adulthood, for individuals from low Head Start spending counties. The replication estimates in Panel (b) are not identical to the published estimates in Panel (a); however, both plots show no evidence that the introduction of Head Start has a statistically significant effect on educational attainment in adulthood for individuals from low Head Start spending counties.

Figure 3.3 plots the event study estimates for the effect of Head Start exposure on educational attainment in adulthood, for individuals from high Head Start spending counties.

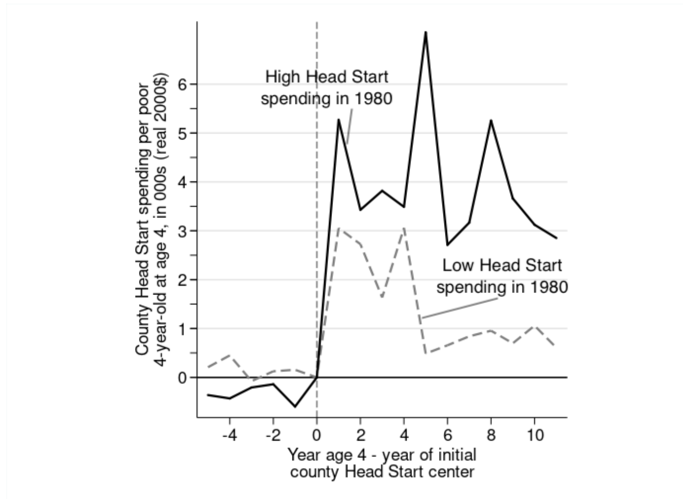
While Panel (a) indicates that the introduction of Head Start leads to a statistically significant increase in educational attainment in adulthood, Panel (b) shows no evidence of a statistically significant change in educational attainment in adulthood following the introduction of Head Start. Further, while the point estimates in Panel (a) depict a striking upward trend in educational attainment following the introduction of Head Start, the point estimates in Panel (b) show no increase or decrease in educational attainment following the introduction of Head Start. Additionally, the point estimates in Panel (b) tend to be estimated with less precision than the corresponding point estimates in Panel (b).

Based on the findings in Figures 3.1, 3.2, and 3.3, I conclude that my replication estimates are not identical to Jackson and Johnson's published estimates. In particular, my replication estimates for the effect of Head Start exposure on educational attainment among individuals from high Head Start spending counties differ substantially from Jackson and Johnson's published estimates. Unlike Jackson and Johnson's published estimates, my replication estimates provide no indication that the introduction of Head Start has a statistically significant impact on poor children's educational attainment.

Figure 3.1: Event Study Estimates for Head Start Spending by Level of Head Start Spending in 1980



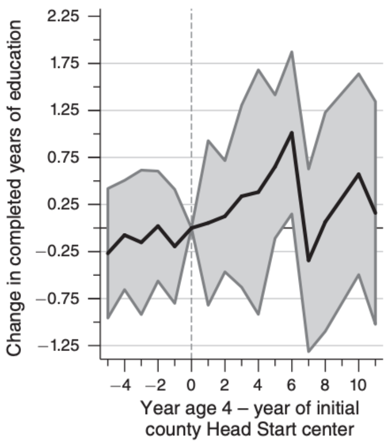
(a) Published Estimates



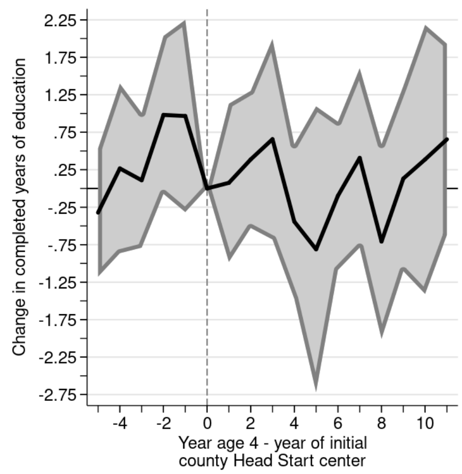
(b) Replication Estimates

Notes: Figure plots event study estimates for county-level Head Start spending per poor four-year-old at age four (measured in $2000, in thousands of dollars) by the level of Head Start spending in 1980. Panel (a) plots estimates obtained by Jackson and Johnson (2019) (see Figure 2, Panel (a) on p. 323 of Jackson and Johnson (2019)). Panel (b) plots estimates from the replication exercise. Event study estimates are produced by estimating model 3.1 on the poor child subsample (children for whom average family earnings between ages 12 and 17 fell in the bottom quartile of the family earnings distribution), separately for children from high versus low Head Start spending counties. Event study regressions are weighted using PSID survey weights.

Figure 3.2: Event Study Estimates for Educational Attainment in Low Head Start Spending Counties



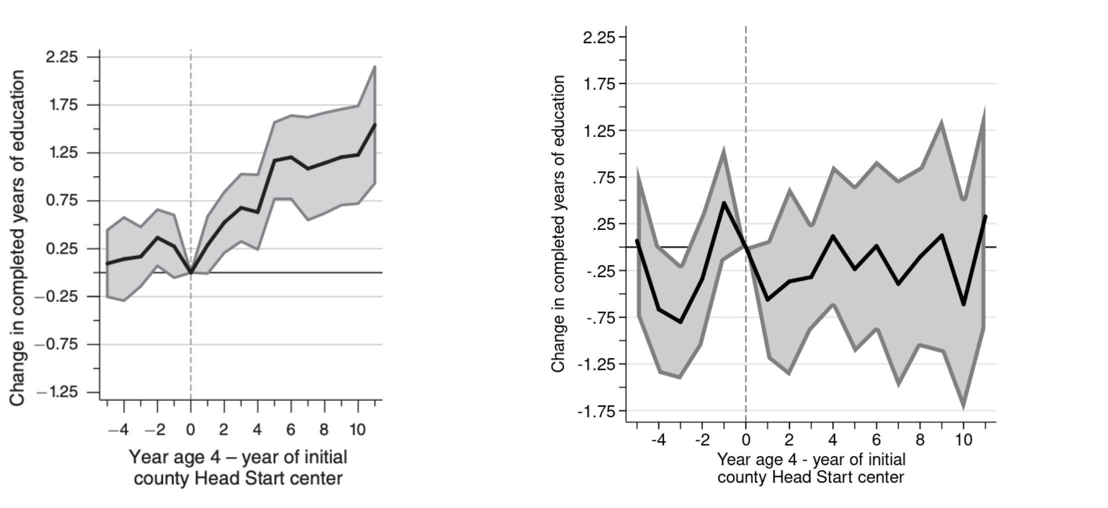
(a) Published Estimates



(b) Replication Estimates

Notes: Figure plots event study estimates for educational attainment among individuals born in counties with bottom quartile Head Start spending in 1980. Panel (a) plots estimates obtained by Jackson and Johnson (2019) (see Figure 2, Panel (b) on p. 323 of Jackson and Johnson (2019)). Panel (b) plots estimates from the replication exercise. Event study estimates are produced by estimating model 3.2 on the poor child subsample (children for whom average family earnings between ages 12 and 17 fell in the bottom quartile of the family earnings distribution). Event study regressions are weighted using PSID survey weights. Standard errors are clustered at the childhood county level.

Figure 3.3: Event Study Estimates for Educational Attainment in High Head Start Spending Counties



(a) Published Estimates (b) Replication Estimates

Notes: Figure plots event study estimates for educational attainment among individuals born in counties with top quartile Head Start spending in 1980. Panel (a) plots estimates obtained by Jackson and Johnson (2019) (see Figure 2, Panel (c) on p. 323 of Jackson and Johnson (2019)). Panel (b) plots estimates from the replication exercise. Event study estimates are produced by estimating model 3.2 on the poor child subsample (children for whom average family earnings between ages 12 and 17 fell in the bottom quartile of the family earnings distribution). Event study regressions are weighted using PSID survey weights. Standard errors are clustered at the childhood county level.

3.4 Other Results

In this section, I use the replication sample to estimate a pair of difference-in-differences models that are similar (though not identical) to a pair of difference-in-differences models estimated by Jackson and Johnson. This is not a replication exercise, and therefore I do not expect my estimates to match the estimates published by Jackson and Johnson exactly. I compare my estimates to the estimates published by Jackson and Johnson. To provide context, I also compare the estimates published by Jackson and Johnson to similar estimates reported in the existing literature.

3.4.1 Difference-in-Differences Estimates

Jackson and Johnson use models (3.3) and (3.4) to produce difference-in-differences estimates for the effect of age-four Head Start spending on a set of outcomes in adulthood. I cannot replicate the estimates from models (3.3) and (3.4) because these estimates depend on the childhood school district of residence, which I do not observe in my data set[8]. However, I can estimate another pair of difference-in-differences models that use a similar source of variation in Head Start spending to identify the effect of age-four Head Start spending on outcomes in adulthood, but do not depend on the childhood school district of residence. I construct this pair of models by aggregating each school district-level variable in models (3.3) and (3.4) to the county level[9]. While I do not expect the estimates that I obtain from this empirical exercise to be identical to Jackson and Johnson's published estimates, it would be surprising if our estimates were not similar since our models exploit similar sources of variation in Head Start spending.

I estimate the following difference-in-differences model on the poor child subsample of the replication sample. I compare the estimates from this model to the published estimates from model (3.3):

$$Y_{icb} = \beta_{HS} HS_{cb}^{age4} + \beta_{K-12} \widehat{\overline{ppe}}_{icb}^{5-17} + \beta_{int} \widehat{\overline{INT}}_{icb} + \gamma C_{icb} + \theta_c + \tau_b + \epsilon_{idb} \qquad (3.5)$$

$\overline{ppe}_{icb}$ is the natural log of county-level, per-pupil K-12 public school spending from ages 5-17, for the cohort from childhood county c and born in year b. $\overline{INT}_{icb}$ is the interaction between HS_{cb}^{age4} and $\overline{ppe}_{icb}$. $\widehat{\overline{dose}}_c$ is the average predicted change in K-12 public school spending that results from the state's court-ordered school finance reform, among school districts located in county c. The terms $\widehat{\overline{ppe}}_{icb}^{5-17}$ and $\widehat{\overline{INT}}_{icb}$ are the fitted values from first stage regressions for $\overline{ppe}_{icb}$ and $\overline{INT}_{icb}$, which include as instruments all possible interactions

8. See the discussion in Section 3.2.1 for more details.

9. In other words, I take the average value for each variable across all school districts that are located within a given county.

between HS_{cb}^{age4}, $\widehat{dose_c}$, and $SFRExp_{idb}$ (with the exception of $HS_{cb}^{age4} \times \widehat{dose_c}$). I cluster standard errors at the childhood state level and weight the regression using PSID survey weights.

Similarly, I estimate the following difference-in-differences model on the poor child subsample of the replication sample. I compare the estimates from this model to the published estimates from model (3.4):

$$Y_{icb} = \beta_{HS}\widehat{HS_{cb}^{age4}} + \beta_{K-12}\widehat{\overline{ppe}}_{icb}^{5-17} + \beta_{int}\widehat{\overline{INT}}_{icb} + \gamma C_{icb} + \theta_c + \tau_b + \epsilon_{idb} \tag{3.6}$$

The terms $\widehat{HS_{cb}^{age4}}$, $\widehat{\overline{ppe}}_{icb}^{5-17}$ and $\widehat{\overline{INT}}_{icb}$ are the fitted values from first stage regressions for HS_{cb}^{age4}, $\overline{ppe}_{icb}$ and $\overline{INT}_{icb}$, which include as instruments all possible interactions between $Exposed\_HS_{cb}^{age4}$, $\widehat{dose_c}$, and $SFRExp_{idb}$ (with the exception of $Exposed\_HS_{cb}^{age4} \times \widehat{dose_c}$). I cluster standard errors at the childhood state level and weight the regression using PSID survey weights.

Table 3.2 compares my estimates from model (3.5) to Jackson and Johnson's published estimates from model (3.3), for each of the indicated outcome variables in adulthood. My estimates (shown in the even-numbered columns) differ substantially from Jackson and Johnson's published estimates (shown in the odd-numbered columns). For the discussion that follows, I assume that average Head Start spending equals $4,230 (measured in $2000; calculated by Jackson and Johnson using county-level Head Start spending per poor four-year-old) and average K-12 public school spending equals $4,953 (measured in $2000; calculated by Jackson and Johnson using school district-level K-12 public school spending per pupil). My estimates imply that introducing a Head Start center with average Head Start spending in a county with average K-12 public school spending leads to worse outcomes for poor children in adulthood, including: a 36.46 percentage point decrease in the likelihood of graduating high school, a 1.11 year decrease in years of completed education, a 30.46 percent decrease in wages in adulthood, an 11.25 percentage point increase in the incidence of poverty in adulthood, and a 1.11 percentage point increase in the likelihood of ever being incarcerated. In contrast, Jackson and Johnson's estimates imply that introducing a Head Start center with average Head Start spending in a school district with average K-12 public school spending leads to improved outcomes for poor children in adulthood, including: a 10.59 percentage point increase in the likelihood of graduating high school, a 0.31 year increase in years of completed education, a 9.87 percent increase in wages in adulthood, a 7.65 percentage point decrease in the incidence of poverty in adulthood, and a a 2.54 percentage point decrease in the likelihood of ever being incarcerated.

Table 3.3 compares my estimates from model (3.6) to Jackson and Johnson's published estimates from model (3.4), for each of the indicated outcome variables in adulthood. There are several inconsistencies (in either sign or magnitude) between my estimates (shown in even-numbered columns) and Jackson and Johnson's estimates (shown in odd-numbered columns). These inconsistencies are most apparent for the following outcome variables: the likelihood of graduating high school, the incidence of poverty in adulthood, and the likelihood of ever being incarcerated. My estimates imply that introducing a Head Start center with average Head Start spending in a county with average K-12 public school spending leads to a 49.49 percentage point decrease in the likelihood of graduating high school, a 1.29 year increase in years of completed education, a 27.07 percent increase in wages in adulthood, a 99.83 percentage point decrease in the incidence of poverty in adulthood, and a 47.8 percentage point increase in the likelihood of ever being incarcerated. On the other hand, Jackson and Johnson's estimates imply that introducing a Head Start center with average Head Start spending in a school district with average K-12 public school spending leads to a 17.3 percentage point increase in the likelihood of graduating high school, a 0.95 year increase in years of completed education, a 15.29 percent increase in wages in adulthood, a 10.9 percentage point decrease in the incidence of poverty in adulthood, and an 8.56 percentage point decrease in the likelihood of ever being incarcerated. Further, my estimates are, in general, measured with less precision than Jackson and Johnson's published estimates.

Based on the results in Tables 3.2 and 3.3, I conclude that my estimates from models (3.5) and (3.6) are not consistent with Jackson and Johnson's published estimates from models (3.3) and (3.4). This is true, despite the fact that I produce my estimates using the replication sample, and I exploit a similar source of variation in Head Start spending to identify the effect of Head Start on long-run outcomes. An important caveat is that the empirical exercise that I performed in this section is not a replication exercise. Therefore, it is possible that the discrepancy between my estimates and Jackson and Johnson's estimates can be explained by the fact that I produce my estimates using empirical models that do not rely on school district-level variation (and that therefore are not identical to the empirical models that Jackson and Johnson use to produce their estimates).

3.4.2 Comparison to Literature

Figure 3.4 presents Jackson and Johnson's published estimates from models (3.3) and (3.4), alongside estimates from six related studies of the long-run effects of the Head Start program. Two of these studies, Garces et. al. (2002) and Deming (2009), estimate the effect of Head Start on long-run outcomes using sibling fixed effects models; the remaining studies, Ludwig and Miller (2007), Thompson (2017), and Bailey et. al. (2018), exploit quasi-experimental variation in the availability of Head Start during early childhood. Each bar in Figure 3.4 represents an estimate from the literature for the average treatment effect on the treated (ATET) of the Head Start program on the indicated outcome variable. For any paper where the ATET is not directly reported by the authors, I calculate the ATET by scaling the estimated intent to treat effect by the estimated first stage effect[10].

Panels (a) and (c) show that Jackson and Johnson's published estimates for the effect of Head Start on high school graduation and log wages, respectively, fall within the range of estimates reported in the prior literature. Panel (b) presents estimates for the effect

10. Bailey et. al. (2018) use a similar methodology to construct Figure 6 (p. 34) in their paper.

Table 3.2: Regression Results: Published Estimates from Model (3.3) Compared to My Estimates from Model (3.5)

	High School Grad		Years of Completed Education		ln(Wage), Age 20-50		Annual Incidence of Poverty, Age 20-50		Ever Incarcerated	
	Published (1)	My Estimate (2)	Published (3)	My Estimate (4)	Published (5)	My Estimate (6)	Published (7)	My Estimate (8)	Published (9)	My Estimate (10)
Head Start Spending, Age 4	0.02503***	-0.0862**	0.07721***	-0.263	0.02334***	-0.0720**	-0.01808***	0.0266*	-0.006002*	0.00262
	(0.006942)	(0.0361)	(0.01992)	(0.189)	(0.004503)	(0.0282)	(0.005302)	(0.0147)	(0.003494)	(0.0130)
SFR-Instrumented ln(PPE), Age 5-17	1.1016***	3.607**	4.0399**	27.11**	2.0561***	2.059	-0.7923***	-0.364	-0.8080**	-0.219
	(0.3268)	(1.625)	(1.6751)	(11.40)	(0.4348)	(1.810)	(0.2969)	(0.810)	(0.3397)	(0.966)
Head Start Spending, Age 4 × ln(PPE), Age 5-17	0.1012*	0.363**	0.6460***	0.945	0.1698**	0.251**	-0.1079**	-0.101*	0.05169*	-0.00969
	(0.05454)	(0.127)	(0.2354)	(0.585)	(0.06985)	(0.101)	(0.04267)	(0.0544)	(0.02777)	(0.0448)
Person Observations	5,419	1,834	5,419	1,833	5,613	1,807	6,373	1,855	4,536	1,875
Person-Year Observations	-	-	-	-	55,706	21,568	88,124	29,072	-	-

Notes: Table presents regression results for the effect of age-four Head Start spending, K-12 public school spending, and the interaction between age-four Head Start spending and K-12 public school spending on outcomes in adulthood, produced using DID-2SLS models. Columns (1), (3), (5), (7), and (9) report published estimates that Jackson and Johnson obtain from model (3.3). Columns (2), (4), (6), (8), and (10) report the estimates that I obtain from model (3.5). All estimates are produced using the poor child subsample. All regressors in the table are centered at their approximate means ($4,230 for Head Start spending; 1.6 for K-12 public school spending).

Table 3.3: Regression Results: Published Estimates from Model (3.4) Compared to My Estimates from Model (3.6)

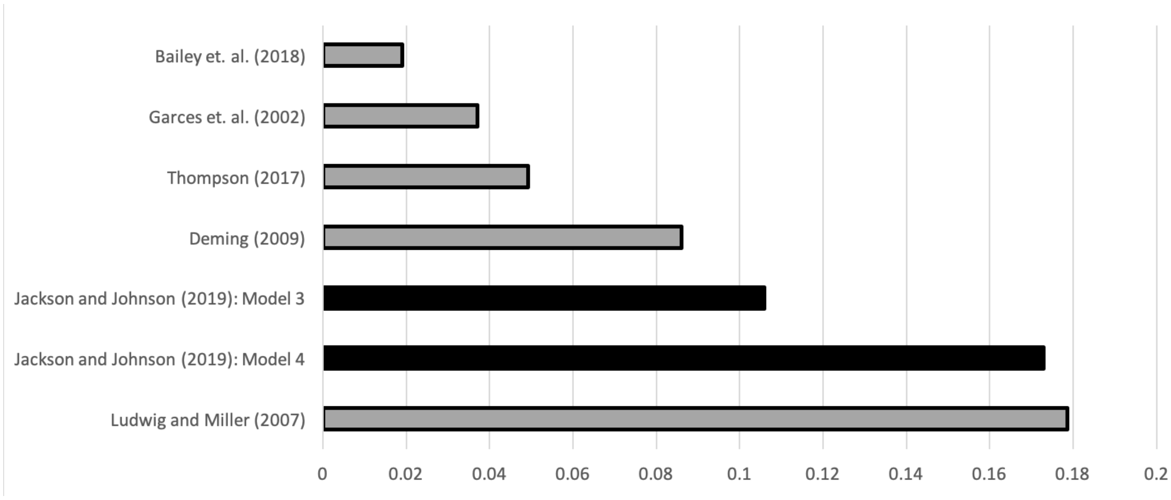
	High School Grad		Years of Completed Education		ln(Wage), Age 20-50		Annual Incidence of Poverty, Age 20-50		Ever Incarcerated	
	Published (1)	My Estimate (2)	Published (3)	My Estimate (4)	Published (5)	My Estimate (6)	Published (7)	My Estimate (8)	Published (9)	My Estimate (10)
Head Start Spending, Age 4	0.04089*	−0.117	0.2255*	0.304	0.03615*	0.0624	-0.02576*	−0.236	-0.02024*	0.113
	(0.02453)	(0.145)	(0.1212)	(0.708)	(0.01956)	(0.134)	(0.01385)	(0.169)	(0.01082)	(0.120)
SFR-Instrumented ln(PPE), Age 5-17	1.4163***	1.492	4.0218**	16.58**	1.2596***	-0.565	-0.7971***	0.434	-1.1822***	0.103
	(0.3390)	(1.330)	(1.7856)	(5.506)	(0.2690)	(1.130)	(0.2903)	(1.314)	(0.4550)	(0.893)
Head Start Spending, Age 4 × ln(PPE), Age 5-17	0.2273***	0.396**	0.8345*	0.840	0.2561***	0.165	-0.1852***	0.173	-0.1808*	-0.217**
	(0.06518)	(0.143)	(0.4824)	(0.801)	(0.07191)	(0.133)	(0.05038)	(0.201)	(0.1076)	(0.109)
Person Observations	5,419	1,834	5,419	1,833	5,613	1,807	6,373	1,855	4,536	1,875
Person-Year Observations	-	-	-	-	55,706	21,568	88,124	29,072	-	-

Notes: Table presents regression results for the effect of age-four Head Start spending, K-12 public school spending, and the interaction between age-four Head Start spending and K-12 public school spending on outcomes in adulthood, produced using 2SLS-IV models. Columns (1), (3), (5), (7), and (9) report published estimates that Jackson and Johnson obtain from model (3.4). Columns (2), (4), (6), (8), and (10) report the estimates that I obtain from model (3.6). All estimates are produced using the poor child subsample. All regressors in the table are centered at their approximate means ($4,230 for Head Start spending; 1.6 for K-12 public school spending).

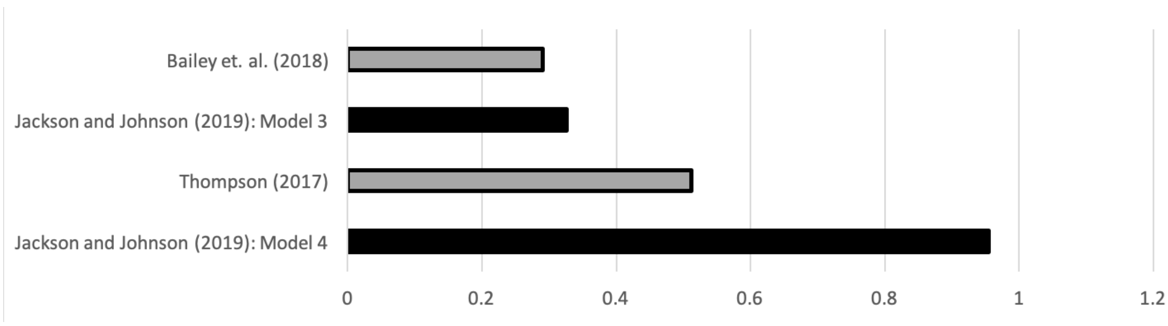
of Head Start on years of completed education. Panel (b) shows that, while Jackson and Johnson's published estimate from model (3.3) falls within the range of estimates reported in the prior literature, their published estimate from model (3.4) falls outside and above this range. Hence, for the most part, Jackson and Johnson's published estimates for the effect of Head Start on high school graduation, years of completed education, and log wages are consistent with the estimates reported in the prior literature[11]. This is true, despite the fact that I cannot reproduce their published estimates using the replication sample.

11. I do not include ever incarcerated or annual incidence of poverty in Figure 3.4 because I was unable to identify more than one paper that produced estimates for the effect of Head Start on these two outcome variables.

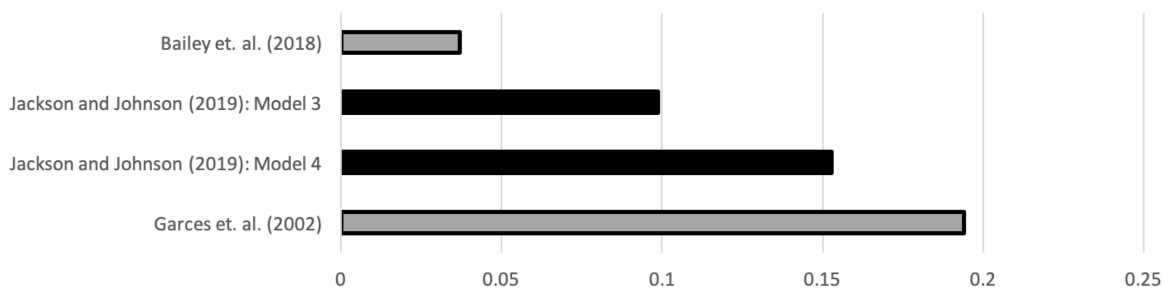
Figure 3.4: Effect of Head Start on Outcomes in Adulthood



(a) High School Graduation



(b) Years of Completed Education



(c) Log Wages

Notes: Figure plots estimates for the average treatment effect on the treated of Head Start on high school graduation, years of completed education, and log wages. Thompson (2017) uses a sample of children whose parents have a high school education or less. Jackson and Johnson (2019) use a sample of children whose average family earnings between ages 12 and 17 fell in the bottom quartile of the family earnings distribution. Log wages are calculated between ages 23 and 25 in Garces et. al. (2002), between ages 25 and 54 in Bailey et. al. (2018), and between ages 20 and 50 in Jackson and Johnson (2019).

3.5 Conclusion

Jackson and Johnson (2019) exploit quasi-random variation in the timing of Head Start introduction across counties to estimate the causal effect of age-four Head Start spending on poor children's long-run outcomes. Their estimates indicate that higher levels of age-four Head Start spending improve poor children's educational and labor market outcomes in adulthood. In particular, they present a set of striking event study estimates which show that, following the introduction of Head Start, educational attainment increased among children from high Head Start spending counties.

In this paper, I construct a replication sample from the replication files that Jackson and Johnson provide with their published paper. I show that, using this replication sample, I am unable to reproduce Jackson and Johnson's published estimates for the effect of age-four Head Start spending on children's long-run outcomes. I do so in two ways. First, I perform a replication exercise where I show that I am unable to reproduce Jackson and Johnson's published summary statistics or event study estimates. Second, I perform an empirical exercise where I estimate a pair of difference-in-differences models that are similar to Jackson and Johnson's difference-in-differences models. I show that the estimates from this empirical exercise are not consistent with Jackson and Johnson's published difference-in-differences estimates.

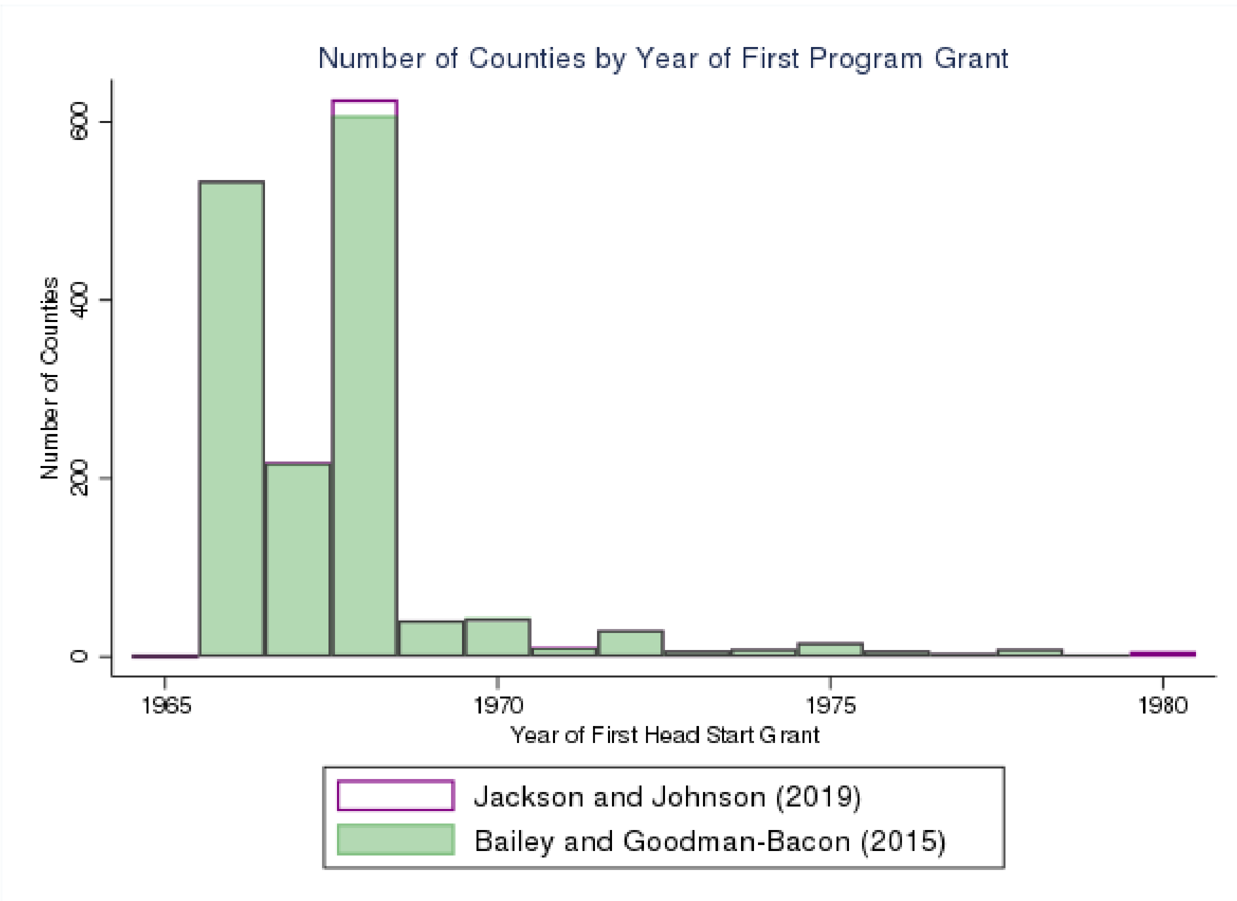
Jackson and Johnson's published estimates consistently indicate that age-four Head Start spending improves children's outcomes in the long run. I show that this same consistent pattern of results does not arise when I attempt to reproduce their published estimates using the replication sample. Further correspondence with Jackson and Johnson is needed to understand (1) how the analysis sample that Jackson and Johnson use to produce their published estimates differs from the replication sample that I construct using the replication files provided with their paper, and (2) why these sample differences translate to differences in the estimated long-run impact of the Head Start program.

APPENDIX A

APPENDIX MATERIAL FOR CHAPTER 1

A.1 Results Appendix

Figure A.1: Histograms for Year of Head Start Introduction



Notes: Figure compares histogram for year of Head Start introduction from Jackson and Johnson (2019) to histogram for year of Head Start introduction from Bailey and Goodman-Bacon (2015).

A.2 Data Appendix

In this appendix, I describe in detail the method that I use to construct the analysis sample in Carter (2020).

1. Prepare a data set containing county-level characteristics.

   The county-level characteristics that I include in this data set are year of Food Stamps introduction from the AER replication file for Hoynes, Schanzenbach, and Almond (2016); year of Head Start introduction from the AEJ replication file for Jackson and Johnson (2019); and safety net program expenditures from the Bureau of Economic Analysis Regional Information System and the AER replication file for Hoynes, Schanzenbach, and Almond (2016).

2. Prepare a data set containing a panel of observations for each survey respondent in the Panel Study of Income Dynamics.

   (a) Import demographic, labor market, and health variables from the 2017 individual file and the 1968-2017 family files for the Panel Study of Income Dynamics.

   (b) Construct year of birth and county of birth variables.

       I construct a year of birth variable, equal to the most recent retrospective report of the year of birth. I construct a county of birth variable, equal to the county of residence in the year of birth for individuals born on or after 1968 or equal to the county of residence in 1968 for individuals born prior to 1968. I fill in any remaining missing values for the county of birth variable using the county of residence for the survey respondent's household in the 13 years prior to the year of birth.

   (c) Merge county-level characteristics into the data set using the year of birth and county of birth variables. Identify the year of Food Stamps introduction and the year of Head Start introduction in the county of birth. Identify safety net program expenditures in the year of birth and county of birth.

3. Select sample of interest.

   The sample of interest consists of individuals who were born between 1950 and 1976, whose parents completed fewer than 12 years of schooling, and for whom the county of birth variable is non-missing (i.e., who were observed in the Panel Study of Income Dynamics during childhood). For each individual in the sample, I keep all observations from survey years 1968 to 2017 where the individual is between the ages of 24 and 54.

4. Construct new variables for the empirical analysis.

   (a) I construct new variables to measure the age at the time of program introduction for Food Stamps and Head Start, the share of years between ages three and six for which the individual was exposed to Head Start, and the share of years between ages in-utero and 11 for which the individual was exposed to Food Stamps.

   (b) I construct a metabolic syndrome index to summarize health outcomes, and I construct an economic self-sufficiency index to summarize labor market outcomes. The metabolic syndrome index equals the average of the z-scores for whether the individual experienced each of the following conditions in adulthood: high blood pressure, diabetes, heart attack, heart disease, and obesity. The economic self-sufficiency index equals the average of the z-scores for personal income, household income, years of education, and whether the individual: was employed, graduated high school, and resided in a household whose income was above the federal poverty line in adulthood. I calculate the sample means and standard deviations that I use to construct the z-scores using the subsample of individuals who were born between 1950 and 1960.

APPENDIX B

APPENDIX MATERIAL FOR CHAPTER 2

B.1 Empirical Methods Appendix

In this appendix, I list the excerpts of code from Jackson and Johnson's replication files that motivate the empirical specifications that I write down in Section 3.2.

1. The following excerpts of code from EventStudyFig2.do in Jackson and Johnson's replication files motivate the empirical specification for model (3.1):

Figure B.1: Code from Jackson and Johnson for Model (3.1)

```
/* TO CREATE HEAD START EVENT STUDY FIGURE 2 (column 1) */

char yearage4_initialyrHeadStart[omit] -1
xi: areg rc_headstartexpage4 i.yearage4_initialyrHeadStart  [
   pweight=wgt]   if agemriw_rc>=25&rcfed_hsperpoorkid1980
   >1800&rcfed_hsperpoorkid1980<., absorb(cntygrp) cluster(
   cntygrp)

char yearage4_initialyrHeadStart[omit] -1
xi: areg rc_headstartexpage4 i.yearage4_initialyrHeadStart  [
   pweight=wgt] if agemriw_rc>=25&rcfed_hsperpoorkid1980<1500&
   rcfed_hsperpoorkid1980<., absorb(cntygrp) cluster(cntygrp)
```

2. The following excerpts of code from EventStudyFig2.do in Jackson and Johnson's replication files motivate the empirical specification for model (3.2):

Figure B.2: Code from Jackson and Johnson for Model (3.2), Control Variables

```
local controls1  "white hisp other  female  agemriw_rc
  rcchildpov1 rcchildpov2 rcchildpov3
  mom_mreducyrs_createfamind dad_mreducyrs_createfamind
  missmomeduc missdadeduc  marriedatbirth lbwt  misslbwt
  notorigsmplkids  yob  FE_division1xyr - FE_division4xyr
  FE_division6xyr - FE_division9xyr   FE_division1w -
  FE_division4w FE_division6w - FE_division9w FE_division1h -
  FE_division4h FE_division6h - FE_division9h FE_division1o -
  FE_division4o FE_division6o - FE_division9o
  pct_black_1960xyr  wpct_black_1960xyr hpct_black_1960xyr
  opct_black_1960xyr  pct_urban_1960xyr lnpop60xyr
  CensusGovt1962_v36xyr    miss_pct_black_1960
  miss_pct_urban_1960 miss_lnpop60   miss_CensusGovt1962_v36
  miss_transpcmcaidage5_17 miss_transpcafdcage5_17
  miss_transpcfspage5_17 miss_transpcuiage5_17"
```

Figure B.3: Code from Jackson and Johnson for Model (3.2), High Spending Counties

```
areg rc5mreduc  ‘controls1’ _Iyearage4__1 _Iyearage4__2
   _Iyearage4__3 _Iyearage4__4 _Iyearage4__5 _Iyearage4__6
   _Iyearage4__7 _Iyearage4__8 _Iyearage4__9 _Iyearage4__10
   _Iyearage4__11 _Iyearage4__12 _Iyearage4__13 _Iyearage4__14
   _Iyearage4__15 _Iyearage4__16 _Iyearage4__17 _Iyearage4__18
   _Iyearage4__19 _Iyearage4__20 _Iyearage4__21 _Iyearage4__22
   _Iyearage4__23 _Iyearage4__24 _Iyearage4__25 _Iyearage4__26
   _Iyearage4__27 _Iyearage4__28 _Iyearage4__29 _Iyearage4__30
   _Iyearage4__31 _Iyearage4__33 _Iyearage4__34 _Iyearage4__35
   _Iyearage4__36 _Iyearage4__37 _Iyearage4__38 _Iyearage4__39
   _Iyearage4__40 _Iyearage4__41 _Iyearage4__42 _Iyearage4__43
   _Iyearage4__44 _Iyearage4__45 _Iyearage4__46 _Iyearage4__47
       lo_Iyearage4__1 lo_Iyearage4__2 lo_Iyearage4__3
   lo_Iyearage4__4 lo_Iyearage4__5 lo_Iyearage4__6
   lo_Iyearage4__7 lo_Iyearage4__8 lo_Iyearage4__9
   lo_Iyearage4__10 lo_Iyearage4__11 lo_Iyearage4__12
   lo_Iyearage4__13 lo_Iyearage4__14 lo_Iyearage4__15
   lo_Iyearage4__16 lo_Iyearage4__17 lo_Iyearage4__18
   lo_Iyearage4__19 lo_Iyearage4__20 lo_Iyearage4__21
   lo_Iyearage4__22 lo_Iyearage4__23 lo_Iyearage4__24
   lo_Iyearage4__25 lo_Iyearage4__26 lo_Iyearage4__27
   lo_Iyearage4__28 lo_Iyearage4__29 lo_Iyearage4__30
   lo_Iyearage4__31 lo_Iyearage4__33 lo_Iyearage4__34
   lo_Iyearage4__35 lo_Iyearage4__36 lo_Iyearage4__37
   lo_Iyearage4__38 lo_Iyearage4__39 lo_Iyearage4__40
   lo_Iyearage4__41 lo_Iyearage4__42 lo_Iyearage4__43
   lo_Iyearage4__44 lo_Iyearage4__45 lo_Iyearage4__46
   lo_Iyearage4__47  [pweight=wgt] , absorb(cntygrp) cluster(
   cntygrp)
```

Figure B.4: Code from Jackson and Johnson for Model (3.2), Low Spending Counties

```
areg rc5mreduc  ‘controls1’ _Iyearage4__1 _Iyearage4__2
   _Iyearage4__3 _Iyearage4__4 _Iyearage4__5 _Iyearage4__6
   _Iyearage4__7 _Iyearage4__8 _Iyearage4__9 _Iyearage4__10
   _Iyearage4__11 _Iyearage4__12 _Iyearage4__13 _Iyearage4__14
   _Iyearage4__15 _Iyearage4__16 _Iyearage4__17 _Iyearage4__18
   _Iyearage4__19 _Iyearage4__20 _Iyearage4__21 _Iyearage4__22
   _Iyearage4__23 _Iyearage4__24 _Iyearage4__25 _Iyearage4__26
   _Iyearage4__27 _Iyearage4__28 _Iyearage4__29 _Iyearage4__30
   _Iyearage4__31 _Iyearage4__33 _Iyearage4__34 _Iyearage4__35
   _Iyearage4__36 _Iyearage4__37 _Iyearage4__38 _Iyearage4__39
   _Iyearage4__40 _Iyearage4__41 _Iyearage4__42 _Iyearage4__43
   _Iyearage4__44 _Iyearage4__45 _Iyearage4__46 _Iyearage4__47
       hi_Iyearage4__1 hi_Iyearage4__2 hi_Iyearage4__3
   hi_Iyearage4__4 hi_Iyearage4__5 hi_Iyearage4__6
   hi_Iyearage4__7 hi_Iyearage4__8 hi_Iyearage4__9
   hi_Iyearage4__10 hi_Iyearage4__11 hi_Iyearage4__12
   hi_Iyearage4__13 hi_Iyearage4__14 hi_Iyearage4__15
   hi_Iyearage4__16 hi_Iyearage4__17 hi_Iyearage4__18
   hi_Iyearage4__19 hi_Iyearage4__20 hi_Iyearage4__21
   hi_Iyearage4__22 hi_Iyearage4__23 hi_Iyearage4__24
   hi_Iyearage4__25 hi_Iyearage4__26 hi_Iyearage4__27
   hi_Iyearage4__28 hi_Iyearage4__29 hi_Iyearage4__30
   hi_Iyearage4__31 hi_Iyearage4__33 hi_Iyearage4__34
   hi_Iyearage4__35 hi_Iyearage4__36 hi_Iyearage4__37
   hi_Iyearage4__38 hi_Iyearage4__39 hi_Iyearage4__40
   hi_Iyearage4__41 hi_Iyearage4__42 hi_Iyearage4__43
   hi_Iyearage4__44 hi_Iyearage4__45 hi_Iyearage4__46
   hi_Iyearage4__47  [pweight=wgt]  , absorb(cntygrp) cluster(
   cntygrp)
```

B.2 Data Appendix

In this appendix, I describe in detail the method that I use to replicate the analysis sample, summary statistics, and event study estimates from Jackson and Johnson. I use data sets and code files that Jackson and Johnson provide in their replication files, which can be downloaded here: https://www.aeaweb.org/articles?id=10.1257/pol.20180510.

1. Construct childhood county of residence variable.

   Jackson and Johnson do not include a complete set of code files to produce the childhood county of residence variable in their replication files. I construct the childhood county of residence variable using the following procedure: I assign childhood county of residence as the earliest observed county of residence for the individual between the ages of zero and 17. I fill in remaining missing values using the most recent retrospective report of the county where the individual grew up.

   I construct the childhood county of residence variable according to the procedure that is described by Jackson and Johnson in their published paper, online appendix, and code files. The following excerpts describe Jackson and Johnson's approach:

   (a) "We use the confidential restricted-use geocode PSID data that includes census block identifiers that correspond with childhood respondent addresses. We match respondent earliest childhood residential location (typically, 1968) to school districts via the combination of GIS mapping methods and school-to-census tract relationship files." (Jackson and Johnson, Online Appendix, p. 6)

   (b) "commands must be done in 2 stages; 1) first do for inds observed while in childhood; 2) then do for inds not observed in childhood but for whom retrospective reports of childhood residential location of upbringing available; and 3) then append the 2 data sets from steps 1-2 together again"
   (Johnson_PSIDgeocode_multistep_mergeSchoolDatafiles_post.do)

2. Run makeregdata_part2_post.do in order to replicate the analysis sample from Jackson and Johnson. This .do file is included in Jackson and Johnson's replication files.

   For reasons that I explain in detail below, I make the following modifications to the .do file:

   (a) Jackson and Johnson import the data set psidall_kidgeocodevars_GIS-ICPSRschoolmerge_9-4-17.dta. This data set contains restricted-use geographic variables (including the childhood county of residence variable), and is neither directly provided in the replication files nor produced by the code that is provided in the replication files. I replace this data set with the data set that I produced in step (1).

   (b) Jackson and Johnson import the data set SPENDlvl_j_clean62_rjedit_sortcnty_8-3-16.dta. This data set is neither directly provided in the replication files nor produced by the code that is provided in the replication files. I replace this data set with the data set SPENDlvl_j_clean62_rjedit_sortcnty.dta, which is provided in the replication files.

   (c) Jackson and Johnson construct a childhood family income variable, which they use to divide the analysis sample into a poor child subsample and a non-poor child subsample. They construct the childhood family income variable using the following procedure: they assign childhood family income as average childhood family income between ages 12 and 17, and they fill in the remaining missing values using the predicted values from a linear regression. Some of the regressors that Jackson and Johnson include in this linear regression are neither directly provided in the replication files nor produced by the code that is provided in the replication files. I omit these regressors (denoted by nonpoorkid, notorigsmplkids, and nognonpoorkid in the .do file) from the linear regression.

       For reference, I include the excerpt of code where Jackson and Johnson specify the original linear regression:

Figure B.5: Code from Jackson and Johnson (2019) for Childhood Family Income Regression

```
reg avgfamincage12_17  parentincome_svyr88wave i.
   parentincomecat_svyr88wave dadincome_svyr88wave i.
   dadincomecat_svyr88wave momincome_svyr88wave i.
   momincomecat_svyr88wave  miss_parentincome_svyr88wave
   miss_parentincomecat_svyr88wave miss_dadincome_svyr88wave
   miss_dadincomecat_svyr88wave miss_momincome_svyr88wave
   miss_momincomecat_svyr88wave /* rcchildpov1 rcchildpov2
   rcchildpov3 */  mom_mreducyrs_createfamind
   dad_mreducyrs_createfamind missmomeduc missdadeduc /*
   parhddropout parhdattendcollege miss_parhdeduc */
   mom_DuncanSEIScore dad_DuncanSEIScore missmom_DuncanSEIScore
    missdad_DuncanSEIScore    marriedatbirth lbwt  misslbwt  i.
   parents_econstatus miss_parents_econstatus nonpoorkid
   notorigsmplkids nognonpoorkid
predict pred_avgfamincage12_17, xb
```

(d) Jackson and Johnson construct a variable, rcdistrict_lnbetacase, which measures the school district-level predicted change in K-12 public school spending that results from the state's court-ordered school finance reform. I do not observe childhood school district of residence in my data set, so I construct rcdistrict_lnbetacase as the county-level, average (among school districts located in each county) predicted change in K-12 public school spending that results from the state's court-ordered school finance reform.

(e) Jackson and Johnson refer to the following variables, which are neither directly provided in the replication files nor produced by the code that is provided in the replication files: South, yearage17, yearage4_initialyrHeadStart, district_lnbetacaselow, district_lnbetacaseneg,district_lnbetacasetop, and district_lnbetacaseoth. I construct South as an indicator variable equal to one if the individual grew up in one of the following states: Virginia, Tennessee, Arkansas, Louisiana, North Carolina, South Carolina, Mississippi, Alabama, Georgia, Florida, or Texas. I construct yearage17 as year of birth plus 17. I construct yearage4_initialyrHeadStart as year of birth plus four, minus the year of Head Start introduction in the individual's childhood county. I construct district_lnbetacaselow as an indicator variable equal to one if rcdistrict_lnbetacase equals 0. I construct district_lnbetacaseneg as an indicator variable equal to one if rcdistrict_lnbetacase is less than 0. I construct district_lnbetacasetop as an indicator variable equal to one if rcdistrict_lnbetacase is greater than 0. I construct district_lnbetacaseoth as an indicator variable equals to one if rcdistrict_lnbetacase is missing (meaning no school finance reform occurs).

3. Run SummaryDesStat_Tab1.do in order to replicate the sample statistics from Jackson and Johnson. This .do file is included in Jackson and Johnson's replication files.

   For reasons that I explain in detail below, I make the following modifications to the .do file:

   (a) Jackson and Johnson import the data sets Table1-c1.dta, Table1-c2.dta, Table1-c3.dta, Table1-c1c.dta, Table1-c2c.dta, and Table1-c3c.dta. These data sets are neither directly provided in the replication files nor produced by the code that is provided in the replication files. I replace these data sets with the data set that I produced in (2). For files Table1-c2.dta and Table1-c2c.dta, I restrict the sample to individuals whose childhood family income between ages 12 and 17 fell in the bottom quartile of the family income distribution. For files Table1-c3.dta and Table1-c3c.dta, I restrict the sample to individuals whose childhood family income between ages 12 and 17 did not fall in the bottom quartile of the family income distribution. I impose these sample restrictions to be consistent with the description of Table 1 on page 322 of Jackson and Johnson (2019).

(b) Jackson and Johnson refer to the following variables, which are neither directly provided in the replication files nor produced by the code that is provided in the replication files: wgt, any_rc_headstartexpage4, and rcheadstart_95iw. I construct wgt as either (1) the PSID survey weight from the most recent survey wave (i.e., the variable iwgt_mostrecent from Jackson and Johnson's replication files), for person-level analyses, or (2) the PSID survey weight in a given survey year (i.e., the variable iwgt from Jackson and Johnson's replication files), for person-by-survey-year-level analyses. I construct any_rc_headstartexpage4 as an indicator variable equal to 1 if Head Start expenditures at age four are greater than 0 in the individual's childhood county. I set rcheadstart_95iw equal to the variable headstart_95iw from Jackson and Johnson's replication files.

(c) Jackson and Johnson mistakenly produce the summary statistic for marriedatbirth using the subsample of individuals for whom the unrelated low birthweight variable is non-missing. I correct this coding mistake, and I calculate the summary statistic for marriedatbirth using the subsample of individuals for whom marriedatbirth is non-missing.

4. Run EventStudyFig2.do in order to replicate the event study estimates from Jackson and Johnson. This .do file is included in Jackson and Johnson's replication files.

   For reasons that I explain in detail below, I make the following modifications to the .do file:

   (a) Jackson and Johnson import the data sets Figure2-c1.dta and Figure2-c2.dta. These data sets are neither directly provided in the replication files nor produced by the code that is provided in the replication files. I replace these data sets with the data set that I produced in (2). I restrict the sample to individuals whose childhood family income between ages 12 and 17 fell in the bottom quartile of the family income distribution. I impose this sample restriction to be consistent with the description of Figure 2 on page 323 of Jackson and Johnson (2019).

(b) Jackson and Johnson refer to the following variables, which are neither directly provided in the replication files nor produced by the code that is provided in the replication files: wgt, rcyearage4_initialyrHeadStart, lowHeadStartspend, hiHeadStartspend, nothighHeadStartspend, and notlowHeadStartspend, I construct wgt as the PSID survey weight from the most recent survey wave (i.e., the variable iwgt_mostrecent from Jackson and Johnson's replication files). I set rcyearage4_initialyrHeadStart equal to yearage4_initialyrHeadStart, which is constructed in (2).

Jackson and Johnson provided the definitions for lowHeadStartspend, hiHeadStartspend, nothighHeadStartspend, and notlowHeadStartspend through email correspondence. Based on this correspondence, I construct lowHeadStartspend as an indicator variable that equals one if Head Start spending in 1980 (measured by rcfed_hsperpoorkid1980) is less than $1500 and non-missing. I construct hiHeadStartspend as an indicator variable that equals one if Head Start spending in 1980 (measured by rcfed_hsperpoorkid1980) is greater than $1800. I construct notlowHeadStartspend as an indicator variable that equals one if Head Start spending in 1980 (measured by rcfed_hsperpoorkid1980) is greater than or equal to $1500 and non-missing. I construct nothiHeadStartspend as an indicator variable that equals one if Head Start spending in 1980 (measured by rcfed_hsperpoorkid1980) is less than or equal to $1800.

(c) The event study regressions for educational attainment include a control variable (denoted by notorigsmplkids in the .do file) that is neither directly provided in the replication files nor produced by the code that is provided in the replication files. I omit this control variable from these regressions.

(d) Based on email correspondence with Jackson and Johnson, I drop observations for which rcfed_hsperpoorkid1980 is missing from the event study regressions.

B.3 Copy of Memo Sent to Jackson and Johnson

In this appendix, I include a copy of the memo that I sent to Kirabo Jackson and Rucker Johnson describing the results of the replication exercise.

B.3.1 Introduction

Jackson and Johnson (2019) posted materials with the AEJ that contain the code files and data sets used to produce the empirical results in their paper. However, these AEJ materials do not contain everything needed to reproduce the results in Jackson and Johnson (2019). This memo focuses on one missing item. The childhood county of residence variable is not in the AEJ files. Further, the AEJ files do not contain the code that Jackson and Johnson (2019) used to create county of childhood residence.

In this memo, I perform a replication exercise. I produce a childhood county of residence variable according to the methodology described in the Jackson and Johnson text and the files they posted with AEJ. Given this variable and the variables provided in their AEJ files, I am not able to reproduce a key figure from Jackson and Johnson (Panel C of Event Study Figure 2). Further, even though I tried several different strategies, I am not able to produce a figure that looks anything like the published figure in Panel C of Event Study Figure 2.

After running many tests, I conjecture that the childhood county of residence variable is the key factor driving the difference between my replication results and Jackson and Johnson?s published results. I show that the total number of distinct counties in the childhood county of residence variable that I construct does not match the total number of counties that Jackson and Johnson report. However, my total number of childhood counties of residence is fairly close to the count implied by the Hoynes, Schanzenbach, and Almond (2016) definition of childhood county of residence. Further, Hoynes, Schanzenbach, and Almond

(2016) follow a similar research design that employs PSID data, and I have been able to replicate their key findings concerning the impacts of access to Food Stamps. (I am glad to provide more details about these replication files.)

Before proceeding, let me highlight an important limitation of the scope of the analysis that I perform in this memo. I have only attempted to replicate two results from Jackson and Johnson: the descriptive statistics for the childhood county of residence variable and Event Study Figure 2. I do not have access to the childhood school district of residence variable that Jackson and Johnson use to produce the remaining results in their paper.

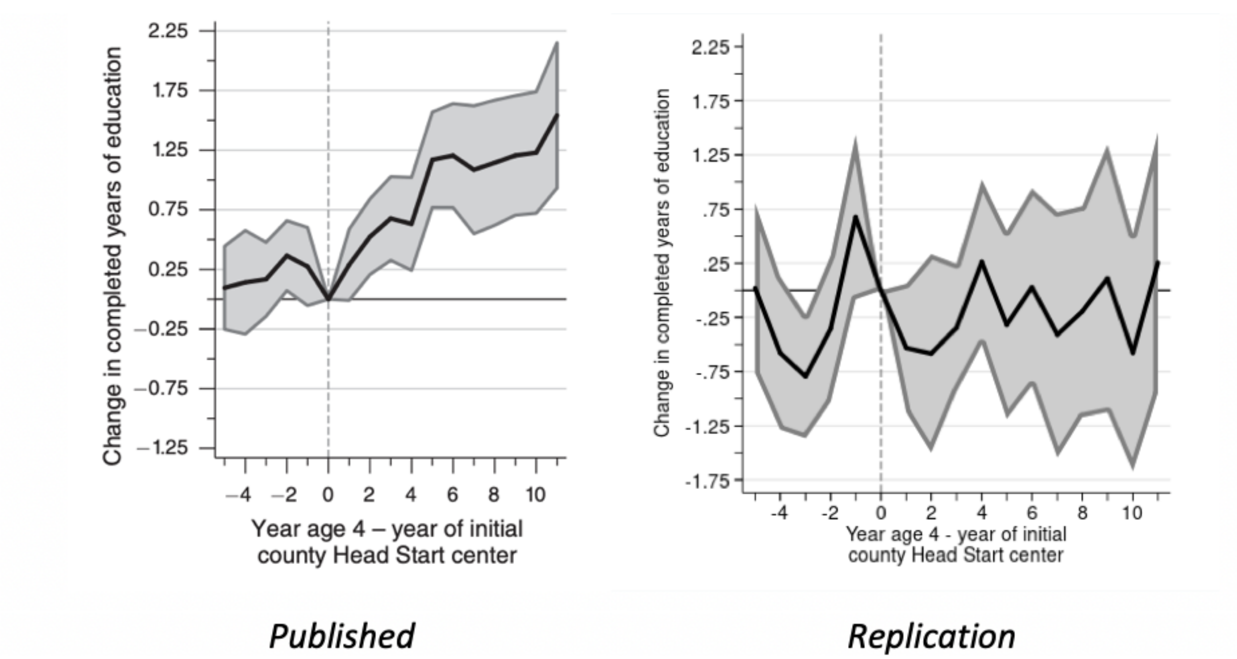
B.3.2 Replication Exercise for Panel 3 of Event Study Figure 2 in Jackson and Johnson (2019)

Here, I describe my attempts to reproduce Panel C of Event Study Figure 2 in Jackson and Johnson (2019). First, I constructed a sample ("Jackson and Johnson replication sample") that contains all the variables required to produce the figure. The key step in this sample construction is the creation of the childhood county of residence variable. This variable is not included in the AEJ files that Jackson and Johnson posted. Further, these AEJ files do not contain code that produces this variable. Therefore, I constructed county of childhood residence with my own code. This code seeks to follow how Jackson and Johnson define this variable in the text of their paper.

My Jackson and Johnson replication sample also includes the following variables that are either directly provided in the AEJ files or are produced using code that is provided in the AEJ files: rc5mreduc, iwgt_mostrecent, white, hisp, other, female, agemriw_rc, rcchildpov1, rcchildpov2, rcchildpov3, yob, fivep_child, mom_mreducyears_createfamind, dad_mreducyears_createfamind, missmomeduc, missdadeduc, marriedatbirth, lbwt, misslbwt, rc_headstartexpage4, miss_headstartexpage4, yearage4_initialyrHeadStart, rcfed_hsperpoorkid1980, pct_black_1960xyr, wpct_black_1960xyr, hpct_black_1960xyr, opct_black_1960xyr, pct_urban_1960xyr, lnpop60xyr, CensusGovt1962_v36, miss_pct_urban1960, miss_lnpop60, miss_CensusGovt1962_V36, miss_transpcmcaidage5_17, miss_transpcafdcage5_17, miss_transpcfspage5_17, and miss_transpcuiage5_17.

Panel C of Event Study Figure 2 plots the event study estimates for the effect of Head Start exposure on educational attainment in adulthood, among children who grew up in high Head Start spending counties. I produced Panel C of Event Study Figure 2 using the Jackson and Johnson replication sample described above and the progam EventStudyFig2.do - which is provided in the AEJ files. Figure B.6 shows that, using the variables that I obtained from the AEJ files combined with the childhood county of residence variable that I constructed, I was not able to reproduce the plot that is published in the paper.

Figure B.6: Panel C of Event Study Figure 2 in Jackson and Johnson (2019)



At this point, I feared that I made an error in processing the AEJ samples and code. To address this concern, I constructed a new sample ("Carter replication sample") that contains the same variables as the Jackson and Johnson replication sample, except that I constructed every variable that I could from scratch in the PSID. Most variables in the Carter replication sample are identical or nearly identical to the corresponding variable in the Jackson and Johnson replication sample. However, differences between the two samples arise because the AEJ files do not provide all of the code that Jackson and Johnson use to make imputations when constructing the variables in the AEJ samples. I identified three sets of variables that differ significantly between the two samples: the educational attainment variable (rc5mreduc), the parental background variables (rcchildpov, mom_mreducyrs_createfamind, dad_mreducyrs_createfamind, missmomeduc, missdadeduc, marriedatbirth, fivep_child), and the race variables (white, black, other, hisp).

To understand how differences in these sets of variables may impact Panel C of Event Study Figure 2, I produced several versions of the figure using different combinations of variables from the Carter replication sample and the Jackson and Johnson replication sample. In particular, I substituted into the Jackson and Johnson replication sample each possible combination of the educational attainment variable, the parental background variables, and the race variables from the Carter replication sample. In total, I created nine additional figures, and each one looks quite similar to the Replication figure above. I am glad to provide these figures if they would be helpful.

At this point, I concluded that the difference between my Jackson and Johnson replication figure and the published plot likely arises from some difference between the childhood county of residence variable that I construct and the childhood county of residence variable that Jackson and Johnson used to produce the published plot. The next section explores this issue further.

B.3.3 Method for Constructing the Childhood County of Residence Variable

Below, I provide all excerpts of text from Jackson and Johnson and all sections of code from my replication file for Jackson and Johnson that relate to the methodology used to construct the childhood county of residence variable.

1. "We linked persons in the PSID using their census blocks during childhood to school spending data, SFR data, and Head Start spending data. We then match the earliest

available childhood residential address to the school district boundaries that prevailed in 1969 to avoid complications arising from endogenously changing district boundaries over time." Jackson and Johnson (2019), p. 320.

2. "We use the confidential restricted-use geocode PSID data that includes census block identifiers that correspond with childhood respondent addresses. We match respondent earliest childhood residential location (typically, 1968) to school districts via the combination of GIS mapping methods and school-to-census tract relationship files." Jackson and Johnson (2019), Online Appendix, p. 6.

3. "commands must be done in 2 stages; 1) first do for inds observed while in childhood; 2) then do for inds not observed in childhood but for whom retrospective reports of child- hood residential location of upbringing available; and 3) then append the 2 data sets from steps 1-2 together again"
Johnson_PSIDgeocode_multistep_mergeSchoolDatafiles_post.do.

Based on the excerpts of text and code provided above, I constructed the childhood county of residence variable according to the following methodology. For individuals in the sample whose county of residence is observed between ages zero and eighteen, I assigned the childhood county of residence as the earliest-observed county of residence between ages zero and eighteen. For all other individuals in the sample, I assigned the childhood county of residence as the most recently-observed retrospective report of the county where the individual grew up. (My results are quite similar if I use the earliest retrospective report.)

Table B.1 shows that the number of childhood counties reported in the published paper does not equal the number of childhood counties in my Jackson and Johnson replication sample. My Jackson and Johnson replication sample contains 513 counties (1,262 counties if we include individuals whose sample weight equals zero), while the published paper reports 1,120 counties. Further, the 513 count is the most relevant number because the regression that produces Panel C of Figure 2 is a weighted regression.

Table B.1: Number of Counties, Published versus Jackson and Johnson Replication Sample

Sample	Number of Counties
Published	1120
Jackson and Johnson Replication Sample	513 (weighted), 1262 (unweighted)

At this point, I wanted to test my code by trying to replicate some other paper that uses similar data. In a 2016 AER paper, Hoynes, Schanzenbach, and Almond (HSA) estimate the effect of Food Stamps exposure during childhood on long-run outcomes using the sample of heads and spouses in the PSID who were born between 1956 and 1981 and who were observed during childhood. They construct a childhood county of residence variable using a methodology that is similar (though not identical) to the methodology that I used to construct the childhood county of residence variable in the Jackson and Johnson replication sample. Table B.2 shows that the number of childhood counties in the Jackson and Johnson replication sample that I created is comparable to the number of childhood counties in the HSA sample. The Jackson and Johnson replication sample contains 513 counties. When I built the replication files that HSA posted at AER, I found 578 childhood residence counties using their definition of childhood residence. Furthermore, when I used these AER replication files to construct the HSA childhood county of residence variable on the Jackson and Johnson replication sample, I obtained 543 counties.

Table B.2: Number of Counties, Hoynes, Schanzenbach, and Almond Sample versus Jackson and Johnson Replication Sample with Hoynes, Schanzenbach, and Almond Childhood County of Residence Variable

Sample	Number of Counties
Hoynes, Schanzenbach, and Almond Sample	578
Jackson and Johnson Replication Sample with Hoynes, Schanzenbach, and Almond Childhood County of Residence Variable	543

These counts are not identical. However, I have also used files that I have created from scratch to replicate the key findings in HSA. Using many of the same lines of code that I used to create my Jackson and Johnson replication file, I have successfully produced results that are quite close to the published results in HSA. This gives me some confidence that I do understand the PSID data and its structure.

In an appendix below, I show that my replication sample produces summary measures of head start spending and availability that are different but not terribly different from the published results in Jackson and Johnson (2019). This pattern further suggests that differences in how these measures of spending are linked to individuals through their county of birth are likely driving the differences between the Replication figure and the published version of Figure 2: Panel C.

B.3.4 Conclusion

I conclude that there are three possibilities.

- HSA and I made a common error in constructing childhood county of residence variables, and that this mistake affects the results in Panel C of Event Study Figure 2.
- There is a programming error in the missing code that Jackson and Johnson used to produce the childhood county of residence variable they used to produce their published AEJ results.
- The code in the AEJ files that is alleged to produce Event Study Figure 2 is not the same as the code that the authors actually used to produce Event Study Figure 2.

B.3.5 Memo Appendix

Appendix Table B.3 shows that, using the Carter replication sample, I am able to replicate the estimate from Hoynes, Schanzenbach, and Almond (2016) for the effect of Food Stamps exposure on the likelihood of completing high school. This result lends credibility to the Carter replication sample.

Table B.3: Effect of Food Stamps Exposure on Completing High School, Replication Exercise

	Hoynes, Schanzenbach, and Almond Sample	Carter Replication Sample
Share Food Stamps Program,	0.184	0.185
In Utero to Age Five	(0.108)	(0.108)
Person-Year Observations	21197	20653

Appendix Table B.4 shows additional descriptive statistics for the childhood county of residence variable, in the published paper compared to the Jackson and Johnson replication sample.

Table B.4: Descriptive Statistics for Childhood County of Residence Variable

Descriptive Statistics	Published	Jackson and Johnson Replication Sample
Share of Individuals with Any Head Start Center in Childhood County at Age 4	0.33	0.30
Post Roll-Out Head Start Spending per Poor 4-Year-Old at Age 4 in Childhood County	$4103	$4320

Notes: Calculated using PSID sample weights. For the replication exercise, I exclude from the sample all individuals for whom average family income between ages 12 and 17 is missing (so that the individual cannot be designated as belonging to either the poor or the non-poor subsample).

Appendix Table B.4 shows that the descriptive statistics for Head Start, which depend on the childhood county of residence variable, are similar (though not identical) between the published paper and the Jackson and Johnson replication sample. The share of individuals with any age-four Head Start exposure is 0.30 in the Jackson and Johnson replication sample, compared to 0.33 in the published paper. The average Head Start per-pupil spending level at age four is $4,318 in the Jackson and Johnson replication sample, compared to $4,103 in the published paper.

REFERENCES

Achyuta Adhvaryu, Anant Nyshadham, Teresa Molina, and Jorge Tamayo. Helping Children Catch Up: Early Life Shocks and the PROGRESA Experiment. Technical report, National Bureau of Economic Research, 2018. Working Paper.

Anna Aizer, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review*, 106(4):935–71, 2016.

Douglas Almond and Janet Currie. Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, 25(3):153–172, 2011.

Douglas Almond, Janet Currie, and Valentina Duque. Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature*, 56(4):1360–1446, 2018.

Douglas Almond, Hilary Hoynes, and Diane Whitmore Schanzenbach. Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes. *The Review of Economics and Statistics*, 93(2):387–403, 2011.

Douglas Almond and Bhashkar Mazumder. Fetal Origins and Parental Responses. *Annual Review of Economics*, 5(1):37–56, 2013.

Martha J. Bailey and Nicolas J. Duquette. How Johnson Fought the War on Poverty: The Economics and Politics of Funding at the Office of Economic Opportunity. *The Journal of Economic History*, 74(2):351–388, 2014.

Martha J. Bailey and Andrew Goodman-Bacon. The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans. *American Economic Review*, 105(3):1067–1104, 2015.

Martha J. Bailey, Hilary Hoynes, Maya Rossin-Slater, and Reed Walker. Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program. Working Paper, 2019.

Martha J. Bailey, Shuqiao Sun, and Brenden Timpe. Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency. Working Paper, 2018.

Andrew Barr and Chloe Gibbs. Breaking the Cycle? Intergenerational Effect of an Anti-Poverty Program in Early Childhood. Working Paper, 2018.

Jacob Bastian and Katherine Michelmore. The Long-Term Impact of the Earned Income Tax Credit on Children's Education and Employment Outcomes. *Journal of Labor Economics*, 36(4):1127–1163, 2018.

Prashant Bharadwaj, Katrine Vellesen Loken, and Christopher Neilson. Early Life Health Interventions and Academic Achievement. *American Economic Review*, 103(5):1862–91, 2013.

Marinne P. Bitler, Hilary W. Hoynes, and Thurston Domina. Experimental Evidence on Distributional Effects of Head Start. Technical report, National Bureau of Economic Research, 2014. Working Paper.

David W. Brown, Amanda E. Kowalski, and Ithai Z. Lurie. Long-Term Impacts of Childhood Medicaid Expansions on Outcomes in Adulthood. *Review of Economic Studies*, Forthcoming.

Pedro Carneiro and Rita Ginja. Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start. *American Economic Journal: Economic Policy*, 6(4), 2014.

Flavio Cunha and James Heckman. The Technology of Skill Formation. *American Economic Review*, 97(2):31–47, 2007.

Janet Currie. Healthy, Wealthy, and Wise? Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature*, 47(1):87–122, 2009.

Janet Currie and Douglas Almond. Human Capital Development Before Age Five. In *Handbook of Labor Economics*, volume 4, pages 1315–1486. 2011.

David Deming. Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1(3):111–34, 2009.

Manasi Deshpande. Does Welfare Inhibit Success? The Long-Term Effects of Removing Low-Income Youth from the Disability Rolls. *American Economic Review*, 106(11):3300–3330, 2016.

Manasi Deshpande. The Effect of Disability Payment on Household Earnings and Income: Evidence from the SSI Children's Program. *The Review of Economics and Statistics*, 98(4):638–654, 2016.

Manasi Deshpande and Yue Li. Who is Screened Out? Application Costs and the Targeting of Disability Programs. *American Economic Journal: Economic Policy*, Forthcoming.

Eliana Garces, Duncan Thomas, and Janet Currie. Longer-Term Effects of Head Start. 92(4):999–2012, 2002.

Alexander M. Gelber and Adam Isen. Children's Schooling and Parents' Behavior: Evidence from the Head Start Impact Study. *Journal of Public Economics*, 101:25–38, 2013.

Chloe Gibbs, Jens Ludwig, and Douglas L. Miller. Does Head Start Do Any Lasting Good? Technical report, National Bureau of Economic Research, 2011. Working Paper.

Andrew Goodman-Bacon. The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes. Technical report, National Bureau of Economic Research, 2016. Working Paper.

Andrew Goodman-Bacon. Public Insurance and Mortality: Evidence from Medicaid Implementation. *Journal of Political Economy*, 126(1):216–262, 2018.

Snaebjorn Gunnsteinsson, Achyuta Adhvaryu, Parul Christian, Alain Labrique, Jonathan Sugimoto, Abu Ahmen Shamim, and Keith P. West Jr. Resilience to Early Life Shocks: Evidence from the Interaction of a Natural Experiment and a Randomized Control Trial. Working Paper, 2016.

Justine Hastings, Ryan Kessler, and Jesse M. Shapiro. The Effect of SNAP on the Composition of Purchased Foods: Evidence and Implications. Working Paper, 2018.

Justine Hastings and Jesse M. Shapiro. How Are SNAP Benefits Spent? Evidence from a Retail Panel. *American Economic Review*, 108(12), 2018.

Tarjei Havnes and Magne Mogstad. No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes. *American Economic Journal: Economic Policy*, 3(2):97–129, 2011.

James Heckman. The Economics, Technology, and Neuroscience of Human Capability Formation. *Proceedings of the National Academy of Sciences*, 104(33):13250–13255, 2007.

James Heckman and Stefano Mosso. The Economics of Human Development and Social Mobility. *Annual Review of Economics*, 6(1):689–733, 2014.

James J. Heckman, Rodrigo Pinto, and Peter A. Savelyev. Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, 103(6):2052–86, 2013.

Hilary Hoynes, Marianne Page, and Ann Huff Stevens. Can Targeted Transfers Improve Birth Outcomes?: Evidence from the Introduction of the WIC Program. *Journal of Public Economics*, 95(7-8):813–827, 2011.

Hilary Hoynes and Diane Whitmore Schanzenbach. Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program. *American Economic Journal: Applied Economics*, 1(4):109–39, 2009.

Hilary Hoynes, Diane Whitmore Schanzenbach, and Douglas Almond. Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review*, 106(4):903–34, 2016.

Adam Isen, Maya Rossin-Slater, and W. Reed Walker. Every Breath You Take - Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970. *Journal of Political Economy*, 125(3), 2017.

Rucker C. Johnson and C. Kirabo Jackson. Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending. *American Economic Journal: Economic Policy*, Forthcoming.

William R. Johnson and Derek Neal. The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy*, 104(5):869–895, 1996.

Patrick Kline and Christopher R. Walters. Evaluating Public Programs with Close Substitutes: The Case of Head Start. *The Quarterly Journal of Economics*, 131(4):1795–1848, 2016.

Hamish Low, Costas Meghir, Luigi Pistaferri, and Alessandra Voena. Marriage, Labor Supply and the Dynamics of the Social Safety Net. Technical report, National Bureau of Economic Research, 2018. Working Paper.

Jens Ludwig and Douglas L. Miller. Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *The Quarterly Journal of Economics*, 122(1):159–208, 2007.

Ofer Malamud, Cristian Pop-Eleches, and Miguel Urquiola. Interactions Between Family and School Environment: Evidence on Dynamic Complementarities? Technical report, National Bureau of Economic Research, 2016. Working Paper.

Sarah Miller and Laura Wherry. The Long-Term Effects of Early Life Medicaid Coverage. *Journal of Human Resources*, 54(3):785–824, 2019.

Office of Management and Budget. Historical Tables, 2018.

Petra Persson and Maya Rossin-Slater. Family Ruptures, Stress, and the Mental Health of the Next Generation. *American Economic Review*, 108(4-5), 2018.

Maya Rossin-Slater. WIC in your Neighborhood: New Evidence on the Impacts of Geographic Access to Clinics. *Journal of Public Economics*, 102:51–69, 2013.

Maya Rossin-Slater and Miriam Wust. What is the Added Value of Preschool for Poor Children? Long-Term and Intergenerational Impacts and Interactions with an Infant Health Intervention. Technical report, National Bureau of Economic Research, 2019. Working Paper.

University of Michigan Survey Research Center, Institute for Social Research. Panel Study of Income Dynamics, Restricted Use Dataset, 2019.

Owen Thompson. Head Start's Long-Run Impact: Evidence from the Program's Introduction. *Journal of Human Resources*, 53(4):1100–1139, 2017.

United States Department of Agriculture. SNAP Data Tables, 2019.

United States Department of Health & Human Services. Office of Head Start, 2019.

Christopher R. Walters. Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start. *American Economic Journal: Applied Economics*, 7(4):76–102, 2015.