Design for "The causal effects of sustained unconditional cash transfers: Experimental evidence from two U.S. states" – Health outcomes

Sarah Miller[*] Alex Bartik[†] David E. Broockman[‡] Elizabeth Rhodes[§]
Eva Vivalt[¶]

PRELIMINARY

Abstract

The regular provision of unconditional cash transfers to individuals is a tactic to fight poverty that has attracted significant interest from researchers and policymakers. Despite this interest, many fundamental questions about the effects of receiving sustained unconditional cash transfers remain. Open Research Lab, a nonprofit research organization, aims to help address this absence of data by conducting the U.S.'s first large-scale randomized trial of a guaranteed income. This document describes the design and analysis plan for the study. In the experiment, 1,000 participants will receive $1,000 per month for 3 years. A control group of 2,000 individuals who receive $50 per month will serve as the comparison group. The study offers an opportunity to inform both the debate over unconditional cash assistance and other questions about the effects of income that typically elude causal identification. This document focuses on the design of the study and the health outcomes.

[*]Assistant Professor, University of Michigan Ross School of Business mille@umich.edu
[†]Assistant Professor, University of Illinois
[‡]Associate Professor, University of California, Berkeley
[§]Research Director, Open Research Lab
[¶]Assistant Professor, University of Toronto

Contents

1 Introduction

Since the late 1960s, income inequality in the United States has risen dramatically and the share of income going to the bottom half of the income distribution has fallen by over a third (Piketty, Saez and Zucman 2019). Intergenerational mobility has fallen, wage growth has stagnated for all but the most skilled, and the official poverty rate remains essentially unchanged despite decades of robust economic growth (Chetty and Hendren 2018*a*;*b*; Congressional Research Service 2019; U.S. Department of Health and Human Services 2016). Individuals and communities are struggling as opportunities are increasingly concentrated in urban areas and among the highly skilled. These trends have increased political and social divisions (e.g., Dorn et al. 2016), and the ability of existing social programs to stem them is limited.

Research shows that the current social safety net leaves many Americans cycling in and out of poverty and/or categorically ineligible for aid (Shaefer and Edin 2013; Danziger 2010; Ben-Shalom, Moffitt and Scholz 2012). The patchwork of programs is complex, costly to administer, and difficult to navigate. Take-up rates are often low, particularly among those most in need (Bhargava and Manoli 2015; Finkelstein and Notowidigdo 2019). Due to the high marginal tax rates and eligibility "cliffs" introduced at moderate income levels, families who do find work often face a

difficult trade-off between earnings and the benefits they rely on for survival.

In response to these challenges, policymakers at state and local levels around the country have become increasingly interested in exploring unconditional cash transfers as a solution. Research points to negative economic, social, and psychological feedback loops that keep individuals without a steady income "trapped" in poverty. Sustained unconditional cash transfers seek to break these feedback loops. Interest in unconditional cash assistance has recently skyrocketed, but the debate often relies on conjecture, stereotypes, and studies that are out-of-date, have important methodological shortcomings, or were conducted in very different contexts. This lack of data and experience impedes rigorous policy analyses and data-driven political debate.

To help guide academic, policy, and political debates, we plan to conduct an experiment that will provide new evidence about the effects of sustained unconditional cash transfers in the U.S. We are collaborating with two non-profit organizations that will implement a cash assistance program. Our partners will recruit approximately 3,000 individuals across two U.S. states and randomly assign 1,000 in total to receive $1,000 per month for 3 years. We will conduct extensive quantitative measurement of outcomes related to individuals' economic, social, and physiological self-sufficiency and well-being, as well as gather data on how individuals use their time and money and how their receipt of monthly cash transfers impacts their children and those in their households. We are partnering with state and local government agencies and private entities to measure many outcomes with administrative data. A single study cannot answer all questions about the effects of a guaranteed income, but we view this experiment as the strong foundation for a broader research agenda moving forward.

The experiment also offers the opportunity to speak to policy debates about unconditional cash assistance programs. Most directly, the study will provide evidence that will inform debates about the design of public benefits, including whether benefits should be provided as cash or in-kind, whether they should be provided monthly or annually, and whether transfer programs should be extended to groups that they do not traditionally target (such as young adults without children).

More broadly, the study will allow us to better understand the relationship between income, work, and well-being generally, and it can provide new evidence on the mechanisms underlying rich-poor gaps in policy-relevant outcomes such as education, health, and time use. For example, unearned income may relax liquidity constraints and facilitate investments in health, human capital, or geographic mobility that may provide long-run returns to households. Unearned income may also change individual bargaining power with employers, landlords, family members, romantic partners, and others. Additionally, unearned income may reduce the cognitive burdens that may be created by scarce resources (Mani et al. 2013), causing individuals to make different decisions. We discuss a broad array of additional channels through which unearned income may influence outcomes in subsequent sections.

2 Existing Research

Much of the existing literature on unconditional cash transfers in developed countries focuses on estimating effects on labor supply. Traditional economic theory predicts that unconditional cash transfers should cause individuals to work less (e.g., Becker 1965), while also consuming more of most goods. By providing nonwage income, cash transfers make household incomes less dependent on labor market earnings; this "income effect" allows households to consume more leisure. Based on this insight, much of the literature on unconditional cash transfers and welfare programs more broadly has focused on quantifying and understanding the determinants of income effects (Chan and Moffitt 2018).

Less work has been done measuring how unconditional cash transfers influence household consumption, which is the other impact of unconditional cash transfers predicted by traditional economic theory. Moreover, richer models suggest that unconditional transfers could have more nuanced effects than those predicted by traditional theory due to liquidity constraints, behavioral mechanisms, social interactions and spillovers, and other factors. More recent research has started

to provide evidence on these broader effects of unconditional cash transfers.

2.1 Early experiments on unconditional cash transfers

To examine the effects of a negative income tax (NIT) on the labor supply of recipients, the U.S. government conducted four randomized experiments between 1968 and 1980, while the Canadian government sponsored one. A number of studies have aggregated the findings on reduced labor supply among participants across the four U.S. experiments, and these estimates range between a 5% and 7.9% reduction in the number of hours worked annually per individual for men; a 17% to 21.1% reduction for married women with children; and a 7% to 13.2% reduction for single women with children (Burtless 1986; Keeley 1981; Robins 1985).

The goal of the experiments was to examine the effect of a guaranteed income on labor supply, but supplemental analyses revealed positive effects on birth weight, homeownership, health, children's academic achievement, the number of adults pursuing continuing education, and other indicators of well-being (see, e.g., Hanushek et al. 1986; Widerquist et al. 2005; Murnane, Maynard and Ohls 1981; Weiss, Hall and Dong 1980; Rea 1977; Kehrer and Wolin 1979; Keeley 1980*b*;*a*; Baumol 1974; Maynard 1977; Elesh and Lefcowitz 1977; Maynard and Murnane 1979; Kaluzny 1979; O'Connor and Madden 1979). Similarly, a reexamination of Canada's guaranteed annual income experiment in the 1970s using health administration data shows a significant decrease in hospitalizations—particularly due to accident, injury, and mental health concerns—and an overall reduction in health service utilization among guaranteed income recipients relative to controls (Forget 2011; 2013). These overall improvements in health may lead to significant savings in health system expenditures.

Despite their path-breaking design, these experiments were plagued by nonrandom selection, errors in randomization protocols, differential attrition, nonparticipation, and systematic income misreporting, calling their results into question (Hausman and Wise 1979; Greenberg and Halsey 1983). Even without these empirical issues, the experiments were begun a half-century ago in

a different economic and political context, so the results may not generalize to the present day. Moreover, the 1970s studies also did not track a number of outcomes that more recent research suggests may play key mediating roles in the effects of unconditional cash transfers. The proposed study will employ research tools unavailable during the NIT experiments to generate a more holistic picture of the effects of the supplemental income on individuals. Tracking expenditures and financial data and leveraging a mobile application and web-based surveys to gather data on time use enable us to investigate how the cash transfers are spent and whether individuals are able to make investments that promote long-term economic self-sufficiency and build savings to help weather shocks and reduce vulnerability.

2.2 Evidence from the Earned Income Tax Credit (EITC)

The expansion of the Earned Income Tax Credit (EITC) in the early 1990s provided another opportunity to examine the effects of exogenous increases in income. Because it is linked to the amount earned, the EITC also affects beneficiaries' incentives to be employed and the number of hours worked, creating a substitution effect in addition to the income effect discussed above. Empirical research has suggested that the EITC increased labor force participation but had negligible impacts on hours worked (Eissa and Liebman 1996; Meyer and Rosenbaum 2001; Nichols and Rothstein 2016). Eissa and Hoynes (2004) show that while there is a positive increase in the labor supply of married men, the increase is more than offset by the reduction in labor force participation by married women, leading to an overall decrease in the total labor supply of married couples. There is ongoing debate about these estimates, however, as more recent analyses suggest that the observed effects on the extensive margin may be confounded by the simultaneous effects of welfare reform and a strong economy (Kleven 2018; 2020).

Additional research has investigated the effects of the EITC beyond measures of labor supply. By transferring money to lower-income households, the EITC substantially reduces the fraction of households in poverty. These gains are concentrated among families near the poverty level, how-

ever, and the EITC has little impact on those who are very poor (Meyer 2010). One analysis of maternal health before and after the expansion documented improvements in self-reported health and mental health as well as reductions in the counts of risky biomarkers for cardiovascular diseases, metabolic disorders, and inflammation (Evans and Garthwaite 2014). Another EITC study found reductions in low infant birth weight that may be at least partially attributable to notable decreases in smoking during pregnancy and increases in prenatal care. More generally, the authors highlight that there are positive externalities to safety net programs that may lead policymakers to underestimate the benefits (Hoynes, Miller and Simon 2015). Other welfare reforms, such as Connecticut's Jobs First program, bundled multiple reforms together, making it difficult to determine the effects of individual components (Kline and Tartari 2016).

2.3 Natural Experiments

Unlike unconditional cash transfers, programs like the EITC affect beneficiaries' incentives to be employed and the number of hours worked because the amount of the benefit is linked to the amount of earned income. To address this limitation, several studies have examined the labor supply of lottery winners. Lottery studies generally find that the income effects of these transfers are modest. Using earnings data from the tax records of consenting Massachusetts lottery players, Imbens, Rubin and Sacerdote (2001) estimate that individuals with winnings up to $100,000 reduce their earnings from labor by about 11 percent of the exogenous increase in income provided by their prize. The effect is larger for individuals between 55 and 65, and the marginal propensity to earn actually increases for those with the lowest pre-lottery earnings, although the effect is not statistically significant.

In a study of Swedish lottery winners, Cesarini et al. (2016) also find negative effects on labor supply, though much smaller in magnitude than earlier studies. The authors report that pretax earnings decrease by approximately 1.1 percent of the payout amount per year, mainly due to a reduction in wages from working fewer hours. It is also important to note that, for lottery winners

with a large lump sum or large monthly payments, negative effects on labor supply could also be attributed to higher marginal tax rates on wages. Furthermore, the lottery studies generally either had small samples (Imbens, Rubin and Sacerdote 2001) or took place in policy contexts very different from the U.S. (Cesarini et al. 2016).

Other recent quasi-experimental evidence of responses to exogenous increases in income comes from examinations of the Alaska Permanent Fund and casino disbursements to Native American families in the U.S. The Alaska Permanent Fund provides an annual unconditional cash transfer to every resident of the state. In 2019, this transfer amounted to $1,606. Feinberg and Kuehn's (2018) analysis using data from the American Community Survey shows a negative effect of dividend receipt on hours worked. In contrast, Jones and Marinescu (2018) employ synthetic controls using data from the Current Population Survey and find no effect on the extensive margin and a small positive effect on the intensive margin. Available data was insufficient to determine if the latter is a result of people shifting from full to part time work or more people entering the labor force part time. A study of the effects of casino disbursements to Native American families found that a $4,000 annual increase in income per adult had no effect on parental labor force participation (Akee et al. 2010).

In addition to the effects on labor supply, some of the recent quasi-experimental papers have examined broader outcomes. Research on casino disbursements to Native American families finds that an average increase in annual household income of $1,750 is associated with statistically significant reductions in obesity, hypertension, and diabetes (Wolfe et al. 2012). Casino windfall cash disbursements have also been linked to higher achievement and educational attainment, reduced incidence of risk behaviors in adolescence, improvements in children's mental health, and better parent-child relationships (Akee et al. 2010; 2018; Costello et al. 2003). The Swedish lottery study found that winners consumed fewer mental health medications after winning, particularly those targeting anxiety (Cesarini et al. 2016). Though they did not report statistically significant changes in health service utilization and other indicators of health, the generalizability of the results to the

U.S. context is questionable given the presence of universal health coverage and a generous social safety net.

2.4 Unconditional Cash Transfers in Developing Countries

There is also an important literature on cash transfers in a developing country context. Most of this work focuses on conditional cash transfers and children's outcomes (reviewed, for example, in Fiszbein et al. 2009). However, some studies leverage unconditional cash transfers and consider employment outcomes. Banerjee et al. (2017) review seven government-run cash transfer programs, plus Haushofer and Shapiro's evaluation of a Give Directly program in Kenya (2016), and find no systematic effect on labor supply on either the intensive or extensive margin.

One of the largest and most widely available of these recent cash transfer programs was the 2011 policy enacted in Iran that distributes the equivalent of 28% of the median per capita household income to over 70 million individuals. Despite the size of these transfers, no impacts were found on labor force participation (Salehi-Isfahani and Mostafavi-Dehzooei 2018). Individuals under thirty worked slightly less, though the effect was not statistically significant, and there were very small positive effects on labor supply for some groups (e.g., women and men in industrial and service sectors). These results may not generalize to the U.S., given the significant contextual differences.[1]

Other studies have focused on the impacts of cash transfers targeted at business owners or workers in particular industries (de Mel, McKenzie and Woodruff 2008; Blattman, Fiala and Martinez 2014; Fafchamps and Quinn 2017; McKenzie 2015). Schady and Rosero's (2007) analysis of data from an Ecuadorian unconditional cash transfer program reveals no impact on the labor supply of recipients. In a study of three-generation households in South Africa, Bertrand, Mullainathan, and Miller 2003 find a sharp decline in both the extensive and intensive margin in working-age individuals' labor supply when an older individual in the household receives a pension.

[1] There is also a large literature on conditional cash transfers in developing countries we do not review here.

2.5 Recent Experiments

More recently, there have been a growing number of conditional and unconditional cash transfer pilots in high-income countries. In the U.S., there have been two recent experiments with conditional cash transfers (CCTs) in New York City and Memphis, Tennessee, but results were mixed. The transfers reduced poverty and led to modest improvements in other areas that varied across sites, but researchers did not observe expected gains in academic achievement, employment, and health (Miller et al. 2016; Riccio and Miller 2016). However, a disproportionate amount of the cash rewards went to more advantaged families; in households that earned more rewards, parents had higher education levels and were more likely to be employed and married. There are a number of possible explanations for the lack of impact, including challenges with implementation, the complexity of the incentives, the process of documenting participation, and the small amount of money relative to the cost of living.

Finland recently piloted a basic income scheme targeted to those experiencing long-term unemployment. Two thousand unemployed individuals were randomly selected to receive 560 euros per month unconditionally for two years in lieu of traditional unemployment benefits. Final results are due in 2020, but no significant impacts were found on labor market participation in preliminary analyses (Kangas et al. 2019). It is important to note, however, that the control group was asymmetrically affected by changes to the unemployment system implemented in the middle of the experiment that require unemployment benefit recipients to prove they are looking for a job in order to continue receiving financial assistance. Though survey response rates were low, survey data indicated that basic income recipients experienced less stress, fewer symptoms of depression, and better cognitive functioning than the control group. Positive effects were also found on financial well-being, trust, and confidence in their future possibilities (Kangas et al. 2020).

3 Sample Definition and Sampling Procedures

3.1 Population

3.1.1 Eligibility Criteria

We define the population of interest as all individuals with Social Security Numbers between the ages of 21 and 40, inclusive, whose self-reported total household income in the calendar year prior to enrollment did not exceed 300% of the federal poverty level (FPL). In addition, we will exclude individuals that receive Supplemental Security Income (SSI) or Social Security Disability Income (SSDI), live in public housing or have a Section 8 voucher (also called Housing Choice Voucher) or other housing subsidy, and live in households in which another member receives SSI. Receiving an income supplement could jeopardize individuals' eligibility for housing assistance and SSI, and getting back on these benefits is very difficult and may take years. Losing this assistance could cause permanent harm, so these individuals will be excluded from the study.

3.1.2 Geography

The study will be conducted in regions in two states. Within each state, we chose a mixture of urban counties with large city centers, urban counties with medium-sized city centers, suburban counties, and rural counties.[2] We selected 1-5 counties of each type in each state that are demographically representative of counties of that type in the region. Nationally, roughly 19% of households that meet the eligibility criteria for the cash assistance program live in rural areas, 35% live in suburban

[2]Counties are divided into rural, suburban, small urban, medium urban, and large urban based on the share of households living in rural census tracts, the population density, whether the county is the largest in its metropolitan or micropolitan area, and population. Rural counties are those that have at least 50% of the population living in rural census tracts or population densities of less than 100 per square mile. Suburban counties are those that are not rural counties, but are not the largest city in their metropolitan or micropolitan area and have populations of less than two million. Small urban counties are those non-rural counties that are the largest in their micropolitan area but have urban cores of smaller than 40,000 people. Medium urban counties are those that are the largest in their metropolitan area, but have population densities of less than 1000 per square mile and populations of less than one million. Large urban counties are those that are the largest in their metropolitan area and have populations of at least one million or densities of greater than 1000 per square mile.

areas, less than 1% live in small urban areas, 17% live in medium-sized urban areas, and 28% live in large urban areas. Small urban counties make up a small share of the overall eligible population (less than 1%), so we excluded them from the sample. We aimed to recruit a sample that roughly matched these population shares, but we oversampled large urban areas to reduce recruitment and survey costs. This approach resulted in a sample of program participants composed of 13% individuals living in rural counties, 18% living in suburban counties, 16% living in medium urban counties, and 53% living in large urban counties.

3.1.3 Demographic Characteristics

In additional to the geographically stratified sampling described above, we used stratified random sampling to ensure that low-income individuals are over-represented in the sample of program participants and the share of males and females is approximately proportionate to their shares of the eligible population (which is roughly 62% female). Table 1 reports basic summary statistics of both eligible mailer respondents and enrolled program participants and compares both groups to the population mean characteristics computed using the American Community Survey for eligible households living in study counties. We report estimates of the eligible population both unweighted and reweighted to reflect the FPL group and county type stratification variables that were used.

On most dimensions, the characteristics of the sample closely match the eligible population in study counties. Our sample is slightly poorer, less likely to be Hispanic, and more likely to be female than eligible households as a whole. The biggest differences between our sample and the full eligible population are that our sample is more likely to report having a college degree and to be a renter than the eligible population.

Table 1: Study Sample Characteristics Compared to Eligible Population

	Eligible Populaton Comparison(ACS)				Study Sample		
	Full US Population			Study Counties	Eligible Mailer Respondents		Enrolled Active Survey Group
	Unweighted	Reweighted to Match Enrolled Sample FPL Distribution	Reweighted to Match Enrolled Sample FPL and County	Reweighted to Match Enrolled Sample FPL and County Type	Unweighted	Reweighted to Match Enrolled Sample FPL and County Type Distribution	Unweighted
	(1)	(2)	(3)	(3)	(4)	(5)	(6)
Panel A. Key active group stratification variables							
Income < 100% of FPL	0.25	0.37	0.37	0.37	0.38	0.37	0.37
Income 100-200% of FPL	0.36	0.39	0.39	0.40	0.31	0.40	0.40
Income 200% + of FPL	0.38	0.23	0.23	0.23	0.30	0.23	0.23
Rural County	0.26	0.22	0.13	0.13	0.15	0.13	0.13
Suburban County	0.32	0.34	0.18	0.18	0.18	0.18	0.18
Medium-Sized Urban County	0.16	0.18	0.16	0.16	0.18	0.16	0.16
Large Urban County	0.24	0.26	0.53	0.53	0.49	0.53	0.53
Panel B. Demographic Characteristics							
Any Children	0.59	0.60	0.59	0.62	0.62	0.59	0.58
HH Size	3.4	3.3	3.2	3.3	3.1	3.2	3.0
Age < 30	0.52	0.53	0.54	0.54	0.41	0.54	0.54
White (non-hispanic)	0.59	0.53	0.46	0.40	0.45	0.46	0.46
Black (non-hispanic)	0.17	0.22	0.25	0.30	0.30	0.27	0.30
Hispanic	0.17	0.18	0.21	0.25	0.19	0.23	0.22
Female	0.57	0.59	0.59	0.61	0.69	0.69	0.67
HH Income	36,204	29,822	29,549	30,158	28,715	28,297	28,800
College Degree or more	0.17	0.15	0.16	0.15	0.28	0.29	0.26
Renter	0.56	0.66	0.69	0.67	0.79	0.84	0.85
	919,395	904,792	904,792	35,086	14,708	14,708	3,000

Notes: This table compares the study sample to estimates of the characteristics of the study in the US as a whole. Eligible individuals are thoe ages 21-40 with household incomes of less than 300% of the federal poverty line. Columns (1) - (4) report estimates of the characteristics of eligible households using the American Community Survey (ACS) 2013-2017 pooled sample. Column (1) presents the unweighted means for eligible individuals, Column (2) reweights this sample to match the enrolled sample distribution of income groups as a share of the FPL (which was a stratification target when assigning individuals to the active survey group), Column (3) reweights the ACS sample to match both the income group distribution and the county-type distribution in the enrolled active survey group sample, and Column (4) presents estimates of characteristics of eligible individuals in sttudy counties, reweighted to match the enrolled sample FPL group and county type distribution. Columns (5)-(7) report characteristics of the study sample. Columns (5) and (6) report characteristics of eligible respondents to the mailer and online advertisement reruitment methods. Column (5) is unweighted, while Column (6) is reweighted o match the enrolled sample FPL and county type distribution. Column (7) reports the unweighted mean of the ultimate enrolled actve survey group (i.e. the 3000 individuals assigned to the active group who answered the baseline survey).

3.2 Sampling Frames

3.2.1 Address-based Sampling

The majority of the sample—approximately 87%—was recruited through mailers. We selected addresses in eligible Census tracts from Target Smart (`targetsmart.com`). This vendor appends commercial data on name, income, race, and other available information to addresses from a variety of state and commercial sources. We understand that the accuracy of these commercial data varies widely, but using the data for targeting significantly improved the efficiency and cost of recruitment in pilots of the mailing strategy. About 69% of mailers were targeted to individuals

who appear income and age eligible on the basis of these commercial data. We refer to these as the "targeted mailers".

To ensure that we did not systematically exclude from the sample individuals who are income and age eligible but did not appear as eligible in the commercial data (for example, because they moved or lost a job recently, they have missing or incomplete information in the commercial data, or they do not appear in any of the commercial data), the remaining 31% of the mailers were sent to addresses that were chosen randomly without regard to information from the Target Smart data. We refer to these as the "untargeted mailers." Where data on names was available, we randomly selected one name per household to whom to address the letter.[3] We appended "or Current Resident" to the end of each name.

We sent mailers to Census tracts roughly in proportion to their share of the eligible population within the county type in the region. For example, if a Census tract contains 2% of the eligible households in rural counties in a state, that county was sent roughly the number of mailers required to ensure that the tract represents 2% of the ultimate sample. The number of mailers this procedure required for each tract depended on the share of households in the tract that are eligible for the program, the targeting effectiveness of the commercial data, and the share of respondents we aimed to recruit using targeted versus untargeted mailers. Ultimately, we sent mailers to 1,138,130 unique addresses, making up about 23% of households in the average Census tract in the study.[4]

To identify the optimal mailing strategy and generate variation in selection into the study, we randomized both the number of letters sent to each address (ranging from one to four) and the gift card incentive offered for completing the online screening questionnaire, which ranged from \$0 to \$20. Roughly 2% of mailed households received one letter, 55% received two letters, 26% received three letters, and 17% received four letters. In terms of gift cards amounts, 37% of households received no gift card, 21% received \$5, 17% received \$10, 2% received \$15, and 23%

[3]For the "targeted" mailers and 50% of the "untargeted" mailers, we randomly selected one name per household among those names that appear age eligible in the commercial mailer data.

[4]The exact share varies with response and eligibility rates across different geography types.

received $20.

3.2.2 Alternative Recruitment Methods

In an effort to include in the sample participants selected differently from those who chose to respond to mailers, we employed two alternative methods to recruit the remaining 13% of the sample. First, the partner organizations purchased ads on the Facebook and Instagram platforms that were shown to all age eligible individuals located in program counties. Participants recruited through this method make up about 1 percent of study participants.

Second, the partners placed ads on the Fresh EBT platform. FreshEBT is a free mobile application developed by Propel (`www.joinpropel.com`) that allows Supplemental Nutrition Assistance Program (SNAP, also known as food stamps) recipients to check their balance and manage their benefits. FreshEBT has over 4 million users nationwide, including more than 180,000 active users in the program counties. The partner organization recruited app users in eligible zip codes by placing ads for the study within the app. Participants recruited through this method comprise roughly 12% of study participants.

3.2.3 Mitigating Spillovers Between Participants

We took three primary measures to reduce potential spillovers between study participants (either through direct interactions or through changing housing or labor market conditions). First, we sent mailers in 6 waves, composed of 0.4%, 9.5%, 19%, 25%, 20%, and 26% of the total mailers, spread out over 8 months. We stratified the number of mailers sent across each wave within a Census tract. This meant that, at most, 6% of households in the average tract received a mailer during any given mailer wave.[5]

Second, we capped the number of households we randomized into the program participation

[5]There are a few rural counties where we needed to send mailers to essentially all households within the county during the course of recruitment.

group at 2 for each Census block and 20 for each Census tract. This reduces the probability that participants in the program interact socially.

Third, prior to randomization into treatment and control, we conducted a survey of study participants to ask if they knew anyone else in the study and, if so, who that person was. Individuals who knew another person in the program were randomized in clusters with the other person(s) they knew in the study to avoid spillovers between people with different treatment status. For more details, see Section 4.5 below.

4 Recruitment and Randomization Procedures

4.1 Recruitment to Eligibility Survey

4.1.1 Mailers

The non-profit organizations implementing the cash assistance program first sent the mailers described above, informing individuals they may be eligible to participate in a new program in which participants receive "$50 or more" per month for three years. Following Broockman, Kalla and Sekhon (2017), the mailers directed recipients to a website where they could register their interest in the program and complete a short eligibility screening survey. This screening survey collected demographic data that was used to verify eligibility for the program (e.g., household size and income to determine if respondents' incomes were below the cap, age, participation in public assistance programs). Respondents were also presented with an e-consent form to give the research team permission to access their administrative data. In order to facilitate linkages to administrative data, individuals who consented to share admin data had the option of providing their social security numbers during this process. Consent to share admin data was not a requirement for program participation, and it did not affect the probability of being selected for the program or randomized into the treatment group.

The partner organizations provided a phone number on the letter that people could call with questions or to receive assistance accessing and completing the survey. Ultimately, 38,823 individuals responded to the mailers and completed the eligibility survey, of whom 12,745 were program eligible (33%).

4.1.2 Facebook and Instagram

As described above, each implementing partner organization purchased ads that appeared on Instagram and in the Facebook news feeds of users in all eligible counties who are predicted to be age-eligible for the program. The ads ran for 1-3 weeks and had varied levels of concentration, as measured by ad spending, by zip code group in each state; more money was spent on ads in zip code groups with the highest poverty rates.

The ad included a thumbnail picture of a calculator and a notepad with a list of monthly bills and text announcing a new program in which "Participants will receive $50 or more per month." Clicking a button that said "Learn more" directed respondents to a website hosted by each partner organization that included a brief description of the program, contact information for questions, and a link to complete the same online eligibility survey that mailer recipients completed.

4.1.3 FreshEBT

Also as described above, each implementing partner organization posted ads on the FreshEBT app to users in eligible counties. These notices ran for 1-2 weeks and advertised a "new financial assistance program" in which "selected participants receive $50 or more per month." When a user clicked the "Learn More" button, they were directed to a short form that collected their email address, phone number, age, and zip code. Age-eligible respondents who confirmed that they live in an eligible zip code were sent an email that provided instructions to complete the same online eligibility survey administered to individuals recruited through other methods.

4.2 Randomization 1: To In-Person Enrollment or Passive Monitoring

We then randomized individuals to be targeted for in-person enrollment or to remain in an "administrative data only" control group. Though individuals in the latter group will not participate in any research activities, their de-identified administrative data can be used for comparison on outcomes measured using these data.

Once we had a pool of eligible individuals, we blocked participants by demographics (age, gender, and race) and pre-treatment values of high-priority outcomes collected in the eligibility survey. We randomly assigned participants to the **"administrative data control"** or the **"program participation"** sample. To ensure that we met our demographic quotas[6] in the program participation group, we sent a larger number of mailers than required to reach our sample size and then randomly selected the program participation group to satisfy the demographic quotas. This means that participants had different probabilities of assignment to the "administrative data control." We include all eligible screener respondents who are not randomized into the program participation group in the administrative data control group, but we will reweight the administrative data control group to have the same demographic averages as the program participation group.

In total, 9,504 individuals were placed in the "administrative data control" group, of whom 55% consented to share their non-health related administrative data, yielding an admin control group of 5,266.[7] [8]

We plan to compare outcomes measured using administrative data for the administrative data control group to the control group enrolled in the main study (as described in Randomization 2

[6]There are three demographic quotas that we targeted for the sample. Specifically, we designed the randomization to ensure that i) the share of women in the sample resembles the share of women in the eligible population in study counties; ii) the sample is least 20% non-Hispanic White, 20% Black, and 20% Hispanic; and iii) the household income of at least 30% of the sample is 0-100% of the federal poverty level (FPL), the household income of at least 30% is 101-200% of FPL, and the household income of no more than 25% of the sample is 201-300% of FPL.

[7]Individuals in the admin control group are disproportionately in the middle and high income groups (with household incomes of 101%-200% and 201%-300% of the FPL) given the need to assign households with incomes of 0-100% of the FPL to the program participation group with higher probability in order to achieve our sample income group target goals.

[8]A smaller proportion, 51%, agreed to also share health related administrative data.

below). This comparison will reveal whether participation in the study and receipt of the $50 per month transfer had any effects on outcomes.[9]

4.3 In-Person Enrollment

The partner organizations then attempted to enroll individuals who had been randomized into the group targeted for in-person enrollment into the cash assistance program. As part of this enrollment, we administered the baseline survey to program participants who consented to take part in the research. We contracted with the University of Michigan Survey Research Center (SRC), a survey research firm with extensive experience fielding national studies, to manage recruitment and conduct in-person enrollment and baseline surveys. SRC employees aimed to ultimately complete 3,000 enrollments from the larger pool of possible participants. During the first 3 weeks of an attempted enrollment, interviewers made a total of 12 phone calls to primary and secondary phone numbers and sent follow up emails and text messages. The non-profit partner reached out to the individual at least once during week 4 if no contact had been made, and a different interviewer attempted 3 additional phone calls in week 5. If there had been no response after 6 weeks, we put contact on hold for two months before making another call and sending another text. If there was still no response, interviewers continued to call and text at least once per month until 3,000 participants had been enrolled.[10]

The in-person enrollment proceeded as follows:

- SRC staff first explained the purpose of the cash assistance program and the program pro-

[9]When conducting any such estimation, our estimand will be the average treatment on treated effect (ATT), weighting to the sample actually targeted for enrollment in the program. We had originally planned to conduct pooled analyses that estimated treatment effects by pooling our main analysis with an analysis that compared this "administrative data control" group to the treatment group that received the cash assistance. However, due to many participants having either very low or very high probabilities of assignment to the administrative data control group and the lower than anticipated take-up rate of the study among those assigned to the group targeted for in-person enrollment (due in part to COVID-19, which required enrollment to be done over the phone rather than in person), we do not plan to pursue this estimator for our final analysis. Our power calculations indicated that it would only increase our statistical precision by approximately 2%.

[10]Depending on response rates after the two-month break, interviewers in some cases attempted to reach individuals by visiting their home up to three times. In-person outreach stopped in March 2020 due to the COVID-19 pandemic.

cedures. Everyone was informed that they will receive "$50 or more" each month for three years and that the specific amount will be randomly assigned, but the fact that some participants will receive $1000 each month was not disclosed. This reduces the likelihood that the control group will know they are in the control group, as that knowledge may change their behavior in ways that would bias the results (including differential take up or attrition and a negative reaction to learning one is receiving less than others). Additionally, we did not want the prospect of a large cash transfer to coerce anyone into participating in the study.

- Individuals who agreed to participate in the program were enrolled in accordance with the procedures established by the non-profit organizations implementing the program.
- SRC staff then explained the purpose of the research and the study procedures.
- The explanation included the incentive structure for participation in research activities: $50 each for completing in-person baseline, midline, and endline surveys, $15 for each mobile baseline survey, $10 for each short monthly survey, and $10 per month for completing short activities on a mobile app. These incentives are taxable (unlike the cash assistance gifts), so we will send participants a 1099 if the participation incentive payments exceed $600 per calendar year, although we intend to keep incentives under the threshold.

During study enrollment, the enumerators:

- Obtained informed consent and contact information for friends and family that can help us locate the participant if we cannot reach them.
- Collected names and demographic information for other members of the household and a description of their relationship to the participant, to help document spillover effects.
- Helped the participant install the custom mobile app and showed participants how to use it, if the participant had a smartphone and consented to using a mobile app.

- Administered the first and most comprehensive baseline survey, including collecting biomarkers (height, weight, and blood pressure).

- Helped the participant set up direct deposit for the research incentive payments. If the participant already had a bank account, the interviewer logged in to a custom-built payments processing system and allowed the participants to verify their bank account information. If participants did not have a bank account, they were given the option of opening an account at Chime Bank, an online bank with no monthly fees, no minimum balance, and no overdraft fees. If they chose this option, they received a Visa debit card in the mail within 7 business days.

4.3.1 Changes to Enrollment in Response to COVID-19

Enrollment began in October 2019, and 1,317 individuals were enrolled and completed the in-person baseline survey by March 14, 2020. On March 15, 2020, the University of Michigan imposed restrictions prohibiting all in-person research activities in response to the COVID-19 pandemic. All outreach was suspended and no enrollments were conducted for approximately six weeks. During that time, we worked with SRC to make the necessary adjustments so that interviewers could enroll participants and administer the baseline survey over the phone. With the exception of biomarkers and the cognitive tasks, all other data could be collected over the phone. Enrollments resumed in late April and all remaining participants were enrolled remotely by October 6, 2020. Ultimately, 44% (1317) of enrolled individuals were enrolled via an in-person baseline survey and 56% (1683) were enrolled via phone.

4.4 "Long Baseline"

Enrollments took place over a 12 month period (the "long baseline"). During this time, random assignment to treatment had not yet taken place; all participants who had been enrolled were

receiving the control group cash assistance gift of $50 per month. In the month after a participant was enrolled, we administered three additional waves of web-based baseline surveys, notifying participants by text and email. These "mobile baselines" allowed us to collect data on outcomes that were not included in the in-person baseline. We also began distributing short web-based surveys each month that took approximately 10 minutes to complete. The purposes of these surveys are 1) to gather additional pre-treatment data to increase the precision of the estimates, and 2) to identify individuals likely to attrit from the study under the $50 condition.

The desire to identify participants likely to attrit is primarily driven by concerns over differential attrition. As previously noted, the 1970s NIT experiments were plagued by differential attrition. Differential attrition also seems likely *ex ante*; even though participants will continue receiving their $50 (in the control group) or $1,000 (in the treatment group) monthly payments regardless of whether they participate in all of the surveys, individuals receiving $1,000 per month may nevertheless be significantly more responsive than those receiving only $50. In case this differential attrition occurs, we hope we can identify a large subsample *ex post* that did not exhibit differential attrition, as defined by their *ex ante* responsiveness. For example, we might conclude: "We see differential attrition on average, but among those who answered at least 2 of the 3 pre-randomization baseline surveys, we do not." We will not, however, exclude any participants from randomization or change the probability of assignment to the treatment group based on whether they continue responding to surveys during the "long baseline."

4.5 Randomization 2: Treatment and Control Groups

After all 3,000 individuals had been enrolled, we randomly assigned them to the **"treatment"** ($1,000 per month) and **"program control"** (remain at $50 per month) groups.

We used blocked and clustered random assignment as follows:

1. *Clustering.* We first formed clusters of individuals based on information that a small num-

ber of study participants knew each other. We placed individuals who reported knowing each other into the same cluster, such that they would always receive the same treatment assignment.

2. *Selecting the Waitlist.* We next selected a stratified random sample of 300 individuals in each state to be placed in a waitlist group. Only individuals not in a cluster with other individuals were eligible for this waitlist group. Within this waitlist group in each state, we formed 10 blocks of 30 observations, blocking on a number of pre-treatment characteristics. We then placed the observations on the waitlist in order such that each 10 observations contained one randomly sampled observation from each of the 10 blocks.

3. *Blocking.* We next "collapsed" the data to the cluster level to conduct a cluster-level random assignment. (The vast majority of individuals are in a cluster of size one with no other observations, but around a dozen clusters were of size two or three.) We then formed blocks of clusters as follows. We first formed strata based on race/ethnicity, income group, and state; any clusters with more than one individual within them were placed in their own strata. Within these strata, we formed blocks of three based on several dozen pre-treatment covariates using the `blockTools` package in R. When the number of clusters in a strata did not evenly divide into three, there were either one or two leftover clusters in a strata after the first round of blocking. We then conducted a second round of blocking for these leftover clusters, again forming blocks based on a set of pre-treatment covariates using `blockTools`.

4. *Random Assignment: blocks.* Within each block of three, we selected one of three observations to be in the treatment group and placed the remaining two in the program control group. Given that the number of clusters did not evenly divide into three, within the final block we sampled from the vector $\{0, 0, 1\}$ without replacement to assign treatment within the final block.

5. *Random Assignment: waitlist.* After the first random assignment, we computed the number

of *individuals* (not clusters) in each state that had been placed in the treatment group. Because the clusters are not of equal size, the number of individuals placed in the treatment group during the first random assignment step varies by randomization. We then calculated how many remaining individuals N from the waitlist would need to be placed into the treatment group in order for 1/3 of each state to be in the treatment group. For example, our target was to place 501 participants in one state (1/3 of the 1503 enrolled) into the treatment group; if 401 participants had been randomly assigned to the treatment group in the first randomization, we would place 100 of the state's 300 observations on the waitlist into the treatment group.

Recall that the waitlist had already been placed in a random order within each state. To select the individuals on the waitlist that would be initially placed in the treatment group, we simply selected the top N individuals on the waitlist.

6. *Re-randomization.* After conducting a randomization, we conducted a series of balance checks across several dozen pre-treatment covariates. Each pre-treatment covariate was associated with a different p-value floor, with covariates we deemed to be more important assigned a higher floor. We rejected any randomization where the p-value on a t-test was below the p-value floor for any of the individual variables. We also conducted an F-test for the joint significance of all of the same set of pre-treatment variables by outcome area and rejected a randomization if the p-value on any of these F-tests was over 0.25.

Through simulation, we verified that this procedure resulted in all observations having an exactly 1/3 probability of being in the treatment group.

4.6 Intervention

After random assignment, participants in the treatment and control groups will be notified about the amount of the cash transfer they will receive each month and the schedule for disbursements.

The intervention in this study is an exogenous increase in income in the form of unconditional cash transfers. The transfers ($50 monthly for the program control group and $1,000 monthly for the program treatment group) will be delivered by the implementing non-profit organizations via direct deposit to the participants' bank accounts.[11] All participants will be notified monthly when the payment is deposited into their account.

Receipt of the treatment transfers and the nominal transfer for the control group is not conditional on participation in any of the research activities and individuals can use the money however they choose. Note that the transfers are provided as a gift from a non-profit organization and will not be subject to income tax.

4.6.1 Waitlist

We originally expected that some participants may not have wished to receive the $1,000 per month transfer (e.g., because they did not feel comfortable taking money they did not "earn," or because it affects their eligibility for other benefits). During the first three months of the program, if any individuals assigned to the treatment group refuse the $1,000 per month transfer, we therefore originally planned to go to the next person on the randomized waitlist in their state and offer that person the transfer instead. In practice, though, we only enrolled one person from the waitlist, in order to replace on participant assigned to the treatment group who was removed from the program for violating program rules. We therefore ignore the waitlist in our estimation strategy and analyze the experiment using intent-to-treat, following the original random assignment (since the compliance rate with the treatment was 99.9%).

[11] The implementing partner organizations work with participants who do not have a bank account and who decline to or are unable to open a Chime account to ensure that they are able to receive direct deposits via a reloadable debit card or payment transfer app.

4.7 Outcome Measurement

4.7.1 Monthly Surveys

We plan to use Qualtrics to conduct monthly web-based surveys. Participants will be notified by a text message and an e-mail containing a personalized link to the survey, and we will ask them to complete the questionnaire at their convenience within 2 weeks. We will send reminders to nonresponders, and $10 will be deposited to participants' bank accounts immediately upon completion. We plan to keep the surveys very short to reduce fatigue.

Maintaining regular contact allows us to identify changes in employment, housing, education, and other variables for which a change will trigger an additional module asking about the reasons for the change and collecting new data on relevant measures (e.g., housing quality following a move, job satisfaction and earnings for new job, etc.). We will spread the modules to be administered less frequently across months to keep the length fairly consistent. Questions pertaining to variables with higher likelihood for measurement error or misreporting due to difficulty remembering will be asked more frequently.

If we see large differential attrition from these surveys, we may abandon them and focus on collecting data during the midline and endline surveys. However, we do see the monthly surveys as an important way to maintain contact with respondents, and response rates were very high (over 90%) throughout the pilots.

4.7.2 Midline Survey

The survey firm will administer an in-person midline survey 15-18 months after the treatment group begins receiving $1000 per month.

4.7.3 Endline Survey

The survey firm will administer an in-person endline survey towards the end of year 3, several months before the cash transfers will end. Respondents in the treatment group may behave differently during the last few months of the program in anticipation of the payments ending, so we will conduct this survey a bit early, starting at 2.5 years into the program and ending at least 3 months before the transfers cease. We hope to conduct long-run follow ups in the future after the program has ended to observe whether effects persist.

4.7.4 Administrative Data

We will gather a variety of administrative data which is described in more detail below.

4.8 Mobile Phone Application

Participants have the option to download a mobile phone application created for the study. We will use this mobile app for both passive and active data collection for consenting participants. We will administer 2-4 short activities each month through the app; participants who choose not to or are unable to download the app will be able to complete these activities via a web interface. From the subset of participants who consent to share anonymized location data, we will passively collect GPS location and accelerometer data from the participants' phones that we can connect to other data sources to potentially improve the precision of our estimates.

5 Estimation

We measure each outcome at multiple time periods and observe outcomes across many different substantive areas. To reduce measurement error and the number of hypotheses we test, we pre-specified the following estimation procedure, drawing on Anderson (2008), Finkelstein et al. (2012), and Guess et al. (2023). Table 2 provides an overview.

5.1 Pooling items across time

Our primary hypotheses rely on effects of the treatment on versions of the items[12] that are pooled across time. In order to pool across time, we average individual outcomes across the study period, placing greater weight on later time periods.[13] If we have no measures of an item within a particular time period (e.g., year 2, at midline, etc.) for an individual but do have measures of that item at other time periods, we will replace that item's outcome for that individual at that time period with the treatment-arm-specific mean, following e.g. Kling, Liebman and Katz (2007).

Because we observe higher response rates and less differential attrition in the midline and endline surveys than in the mobile monthly surveys, we perform this aggregation twice: (1) we compute one set of estimates using data only from the midline and endline surveys and (2) we compute another set of estimates also including data from the mobile surveys. The latter set of estimates places 70% of the weight on the midline and endline surveys and 30% of the weight on the mobile surveys when items are observed in both, although some items are only observed in the mobile surveys.

[12] In some cases, what we consider a single item might be a composite based on multiple survey questions. Any such combination of multiple survey questions into a single item would be pre-specified in the topic-specific section of this PAP below.

[13] For the midline and endline estimates, we place 70% of the weight on the endline and 30% of the weight on the midline. For the monthly surveys, we place 50% of the weight on surveys conducting in the final year, 30% of the weight on surveys conducted in the second year, and 20% of the weight on surveys conducted in the first year. When aggregating the monthly surveys by year, average within item within respondent across all non-missing responses for that year. We code each respondent's response to each item within each year as missing only that item is never observed for that respondent within that year.

We refer to these individual outcomes as *primary items* when we pre-specified that they represented primary outcomes. For example, one of the items is a dummy variable which is set to one if the individual has recorded over three moves within the past year. To estimate treatment effects on the individual items, we first average the item across time using the procedure described in the previous paragraph.

5.2 Item-level effects

The first step in our estimation is to estimate the effects on the individual items, either pooled across time or at individual times.

To estimate these estimates, following Bloniarz et al. (2016), we first predict the item (or item pooled across time) using the Lasso to select pre-treatment covariates that predict the item. Next, to estimate the treatment effects we use OLS, regressing the item on a treatment indicator and the Lasso-selected pre-treatment covariates (in order to increase precision), with clustered standard errors. This yields a treatment effect estimate for every individual item (both at each time and pooled across time).

5.3 Component-level effects

Drawing on Anderson (2008), to reduce the number of primary hypothesis tests we conduct, we group primary items into *components*. (Some items are labeled as secondary or tertiary and are not in any components.) Which items were assigned to which components was pre-specified. For instance, we categorized the dummy variable for whether participants moved over three times within the past year in the 'Excessive Residential Mobility' component.

To estimate the effects on each component, we use seemingly unrelated regression with clustered standard errors. This procedure allows the standard errors to reflect the correlation between the estimates in each of the constituent regressions.[14]

[14]This is similar to the 'pooled OLS' approach described in Finkelstein et al. (2012), although in simulations con-

In particular, at each time period and separately on pooled versions of the items when pooling across time periods, we first estimate a system of equations using all of the item-level regressions described in the previous subsection. We then estimate an average of the estimated treatment effect from each of these regressions, weighting the estimate on each item equally. When taking this average, we rescale each item's estimates and standard errors so that the estimates are in terms of standard deviations (by dividing the estimates and standard errors by the standard deviation in the control group for that item). The effect at the component level can therefore be interpreted as the average effect on standardized versions of the individual items within that component.

We separately estimate effects on components using only data from the midline/endline and when also including data from the mobile surveys. Note that because some items were not asked on the midline/endline, and others were not asked on the mobile surveys, the estimates from the midline/endline-only specification and the midline/endline and mobile survey specification will not be directly comparable, as the midline/endline-only specifications may contain different items.

5.4 Family-level effects

Components are also grouped into *families*. For example, we categorized the 'Excessive Residential Mobility' component within the 'Housing Hardship' family.

To estimate effects on individual families, we use seemingly unrelated regression to test the hypothesis that the average of all the component-specific treatment effects is zero. In particular, we place all the individual regressions for the specific items within all the components within the family into one seemingly unrelated regression system. We then estimate all the component-level estimates as described in the previous section, and finally take the average of these component-level estimates. The effect at the family level can therefore be interpreted as the average effect on the components.

Which items are placed into which components and which families are listed in the topic-

ducted when preparing our pre-analysis plan we found that it had better statistical power for our data structure.

Table 2: Item Hierarchy

Level of Aggregation	Example	Estimation Approach
Primary Items, Pooled Across Time	Dummy Variable for Over 3 Moves Within 1 Year	Average the item measured at midline, endline, and in years 1, 2, and 3, placing more weight on later years and on midline/endline. Estimate treatment effects with OLS with clustered standard errors, controlling for Lasso-selected pre-treatment covariates.
Components, Pooled Across Time	Excessive Residential Mobility	Average treatment effect estimates on standardized versions of constituent items using Seemingly Unrelated Regression (SUR).
Family, Pooled Across Time	Housing Hardship	Average the treatment effect estimates on constituent components using Seemingly Unrelated Regression (SUR), taking the regressions on the components as inputs and weighting all components equally.
Secondary Items	Number of Moves in Past Year	Average item measured at midline, endline, and in years 1, 2, and 3. Estimate treatment effects with OLS with clustered standard errors, controlling for Lasso-selected pre-treatment covariates.
Tertiary Items	Number of Moves in Past 5 Years	Average item measured at midline, endline, and in years 1, 2, and 3. Estimate treatment effects with OLS with clustered standard errors, controlling for Lasso-selected pre-treatment covariates.
Heterogenous Treatment Effects	Effect on Excessive Residential Mobility Component by pre-treatment poverty category	Separately estimate treatment effects among each subset.

specific portion of the pre-analysis plan below.

5.5 Heterogeneous Treatment Effects

Given the sample size and the many hypothesis tests we already plan to conduct, we are concerned about statistical power. Therefore we pre-register that all heterogeneous treatment effect estimates will be considered exploratory unless explicitly pre-specified otherwise. Pre-analysis plans for some outcome areas may specify hypothesis tests for heterogeneous treatment effects and note them as exploratory or non-exploratory.

We plan to explore heterogeneity in treatment effects more thoroughly in a separate paper. We will pursue at least three approaches.

5.5.1 Needs and Priorities

First, one may imagine that individuals have different preferences and circumstances that would contribute to how they use the transfers and their subsequent outcomes. For example, one individual may prioritize paying off debts, while another may prioritize going back to school. It is possible that there are some outcomes for which we do not observe statistically significant effects on the whole sample, but which are significant for subgroups that prioritize those outcomes. More broadly, it is possible that the program had larger benefits when evaluated from the perspective of participants' own preferences. In evaluations of the Chicago Resilient Communities Pilot (CRCP) program and the Cook County Promise Pilot (CCPP) program, respondents were asked at enrollment to identify their top needs and priorities.[15] We did not ask these questions at baseline for this study, but we added them at the final month of the treatment period, and at this time we asked them to both identify their current needs and priorities and how they think they would have answered at baseline. We will use these responses to construct estimates that weigh those factors that respondents prioritize more highly as we do for the CRCP and CCPP evaluations.[16] As a robustness check, we will construct alternative weights based on a smaller set of questions respondents were

[15]These needs and priorities that participants of CRCP and CCPP were asked about included: saving money; getting a new car or repairing my current car; finding a new place to live; finding a new job or getting promoted to a better position; being able to pay my bills; buying things I need (such as more or better food, clothes, etc.); reducing stress or anxiety; getting health, dental, or mental health care; paying off debts; buying things for my children (including gifts, child care, and health care costs); finishing my education or getting more education or training; growing my family; improving my romantic relationships or finding a new partner; finding more ways to relax or have fun; helping my community (for example, church or neighborhood); and helping my family members.

[16]If we observe no differences between treatment and control in terms of which outcomes participants prioritize in the final treatment period, we will take that as an indication that it is reasonable to use these weights as a proxy for what individuals would have answered had they been asked the question before treatment. However, we will prefer answers to the question about how they would have ranked items if they had been asked about them at baseline.

asked about at baseline.[17][18]

5.5.2 Deficits at Baseline

Bearing in mind that baseline rates of an outcome variable are often very predictive of treatment effects (Vivalt 2020), and based on the theory that individuals may be more likely to put their efforts into improving areas in which they are relatively deprived, we will seek to construct an estimate of what areas individuals might have prioritized by considering which outcomes in the CRCP/CCPP "needs and priorities" module they had particularly low levels of at baseline.[19] We can then use these priorities to weight each individual's outcomes, similar to our approach in CRCP/CCPP, and to conduct heterogeneity tests.

We can also apply the same approach to the CRCP and CCPP data, for those outcomes for which we have baseline measures, to gauge how well our predictions of what respondents might find most important align with respondents' own priorities. Since policymakers often have to determine how to target programs without knowing individuals' own preferences, this exercise could also be of independent interest in learning about how well such an approach might work.

[17] Including their preferences over finding another job, pursuing more education, starting a business, getting pregnant, and whether they plan to move.

[18] While these questions map to relatively specific things like "saving money", "finding a new place to live" or "paying off debt', individuals may also have broader preferences, such as preferences over improving their health or their financial security. We asked a broader rankings question that can help us synthesize results if we see different effects across different topic areas. This broader ranking question relates to items in Benjamin et al. (2014), so we can observe how different our respondents' preferences are and whether treatment appears to change their preferences. In particular, we include some items that were highly rated in Benjamin et al. (2014) ("your health", "your financial security", "the amount of time you have to do the things that you like doing"*), some medium-ranked items ("your physical safety and security", "the quality of your romantic relationships"*, "your ability to have and raise children") and some low-ranked items ("feeling part of your community"*, "your material standard of living", "the overall quality of your experience at work"). Items with an asterisk are not asked in the exact same way in Benjamin et al. (2014) and will be excluded in a robustness check. We can compare the relative ranks our respondents put on these measures compared to the weights Benjamin et al. (2014) found via a discrete choice experiment and, importantly, test whether treatment appears to have changed those weights. Seven of these items also relate to other satisfaction questions asked at baseline, so that we can distinguish between cases in which people report different ranks because they improved along that dimension since baseline so are relatively satisfied vs. cases where people report different ranks because their preferences changed.

[19] Though we note that it is also possible that a participant having low levels of a variable at baseline instead signals that they are not interested in improving it.

5.5.3 Data-Driven Approach

To further explore heterogeneity and ability to target, we will use a data-driven machine learning approach in which each site's variables are used to predict the best targeting of the program for a certain outcome in the other sites, using the full set of variables available across all sites. The best targeting of the program will likely depend on the outcomes prioritized. At a minimum, we will consider minimizing negative effects on employment rates and income, increasing educational attainment, and improving health outcomes observed in administrative data.

Should we obtain forecasts from policymakers or other experts about how they think different subgroups or samples may react to treatment, we will contrast targeting based on forecasts with targeting based on a data-driven approach.

Finally, we will leverage a partial pooling analysis, described further in a later subsection of this plan, to characterize treatment effects for different quantiles of the data, as in Meager (2019). For this analysis, we will particularly focus on the 10th, 25th, 50th, 75th and 90th quantiles of the data for the continuous outcome variables for which we elicited forecasts.

Again, these more detailed analyses will be performed separately and are not the focus of this pre-analysis plan.

5.6 Robustness Checks

We perform a number of robustness checks to help assess the importance of several empirical concerns.

To assess the impact of the presence of modest differential attrition, we perform the following robustness checks:

- A differences-in-differences analysis in which we compare changes since baseline across the treatment and control groups.
- Bounding analysis to assess differential attrition

- Run the 'midline and endline only' specifications, but subset to cases where the outcome is non-missing at both midline and endline.

For any items marked as susceptible to outliers in the topic-specific component of the pre-analysis plan below, we will perform the following robustness checks:

- Estimate effects using median regression.
- Estimate a version of the estimates where we winzorize the outcomes at the 99th percentile.

5.7 False Discovery Rate Adjustment

Following Guess et al. (2023), the family-, component-, and item-level estimates will be placed into tiers for the purpose of multiple comparison adjustment. We use Benjamini and Hochberg (1995)'s false discovery rate adjustment to compute q-values; following Benjamini and Hochberg (1995) we do this within families of outcomes. We place our hypothesis tests into tiers (denoted K0, K1, K2, K3, and K4) as follows, corresponding with our prioritization of the tests:

- K0: Family-level estimates pooled across time. The q-values for these items will be computed using all the K0 items across families in a paper.
- K1: Component-level estimates pooled across time. The q-values for these items are computed using the K0 and K1 items in the outcome's same family.
- K2: Primary item-level estimates pooled across time. The q-values for these items are computed using the K0, K1, and K2 items in the outcome's same family.
- K3: All other estimates ("exploratory" tier). This includes family-level, component-level, and item-level estimates which are computed within each time period, estimates on items pre-specified as secondary or tertiary, and all tests of heterogenous treatment effects, as well

as descriptive analyses. The q-values for these items are computed using the K0, K1, K2, and K3 items in the outcome's same family.

- K4: Any post hoc comparisons conducted after filing these pre-analysis plans (e.g., in response to referee comments). The q-values for these items are computed using the K0, K1, K2, K3, and K4 items in the outcome's same family.

For example, when estimating the FDR-adjusted q-values for the primary item-level estimates (tier K2), we include the p-values for the treatment effects on every primary item in the entire paper (regardless of which family and component it is in or whether it is based on midline/endline surveys only or not), as well as the p-values from all the components and families.

In some cases, the plan for a family may deviate from this. For example, in some families, only one item is pre-specified to be included in the index for a given component, or only one component for the family. In such cases, the FDR adjustments will done on one fewer "level" (e.g., if only one item is in a component, it will not be adjusted with K2, as it would already be adjusted at the K1 level for the component. If only one component is in a family, that component will be counted as K0, primary items counted as K1, secondary items as K2, etc.). For some families, there will also be a distinction between the secondary and tertiary items. In these cases, secondary items that are pooled across midline/endline and monthly surveys and secondary items that are pooled across midline/endline surveys only will be prioritized above other K3 and K4 items; this effectively pushes other K3 items to K4 and K4 items to K5. These cases will be flagged in the text.

Table 3 summarizes all of our estimates and the FDR tiers we place them in.

We will treat ordinal outcomes as continuous by default.

Table 3: FDR Tiers

	Pooled Across Midline/Endline and Monthly Surveys	Pooled Across Midline/Endline Surveys Only (Omitting Monthly Surveys)	Estimates At Each Time Period (e.g., at midline, in year 2, etc.)
Family	K0	K0	K3
Primary Components	K1	K1	K3
Primary Items	K2	K2	K3
Secondary Items	K3	K3	K3
Tertiary Items	K3	K3	K3
Heterogeneous treatment effects	K3	K3	Not calculated
Any post-PAP tests	K4	K4	K4

5.8 Attrition

We will present a set of results correcting for differential attrition. We will check for balance in attrition rates using the same set of covariates that we used to test for balance at randomization.

We will conduct two-stage sampling for midline and endline data collection to minimize attrition-related bias by concentrating resources and efforts on a randomly chosen subset of the cases that are the most difficult to reach (and adding weights accordingly). We will also keep track of the number of contacts required to reach each participant for each survey. We will consider using the randomly assigned intensive follow-up and number of contacts required to reach each participant to construct attrition adjusted treatment effect estimates.

5.9 Characterizing "Treatment" of Control Group Participants

Not all eligible respondents who complete the online eligibility screener will be randomly selected to participate in the program and study. As a result, we have access to an additional "control" group of individuals who consented to passively provide administrative data but will not be contacted by

the research team. Using this "administrative control" group can help us shed light as to whether the program has any effects on the "program control" group, either as a result of the $50 monthly payments, the survey incentives, or the act of completing surveys themselves. We will use this group to characterize any such effects on outcomes measured using administrative data that might be present in the program control group.

These hypothesis tests are not the main focus of the paper so they are not subject to a multiple comparisons adjustment.

5.10 Elicitation of Forecasts

We will be eliciting forecasts for several key outcomes on the Social Science Prediction Platform. We expect to receive forecasts from other researchers, those working in policy or non-profit organizations, and the general public. These forecasts can help in gauging the novelty of our results (DellaVigna, Pope and Vivalt 2019). There are not currently standard ways of presenting comparisons of *ex ante* forecasts with research results, but we anticipate including some comparisons, if only in an appendix. In comparing our research results to the *ex ante* forecasts, we will focus on comparing our results to the predictions of researchers in economics unless otherwise specified. The outcomes that we will forecast are indicated with an asterisk in the section on outcomes below.

5.11 Partial Pooling Analysis

There are some outcomes which may be relatively low-powered. For example, relatively few individuals may have direct engagement with the criminal justice system. To improve power, we will use a Bayesian hierarchical model to partially pool results across this study and several other programs we are evaluating, namely, the Chicago Resilient Communities Pilot program and the Cook County Promise Pilot program. This analysis will also help us understand and characterize potential heterogeneity in treatment effects.

The logic underlying Bayesian hierarchical models is that each study may be partially informative as to what another study will find, and so partial pooling may improve estimates. For our analyses, rather than use summary measures like treatment coefficients and standard errors, we will use individual-level data where possible in order to take advantage of covariates that vary at the individual level.[20] For continuous outcomes, the default model and priors will be as the main model in Meager (2019); for binary outcomes, we will use a logistic regression model.[21] We will also examine the sensitivity of the results to alternative priors.

The main outcomes considered in this analysis will be the items that are collected in all studies and denoted with an asterisk, for which we are collecting *ex ante* priors. Due to the potential for differential attrition in survey-based measures, the outcomes available in administrative data will be prioritized. In one set of robustness checks we will use these expert forecasts to inform the priors used in the mixed model.

We expect the partial pooling approach to be implemented in a separate paper following the conclusion of all the programs, although depending on the arrival of administrative data relative to the program end dates, it is possible that the analysis will begin with only two of the three programs' data.

6 On Income and Health

The correlation between health and income is one of the most persistent and long-standing observations in the social sciences. This relationship between socio-economic status and health holds across every countries, within age and other demographic groups, and for every almost every measure of health considered (Evans, Wolfe and Adler (2012)). In the United States, a man at age

[20] For some administrative data, it may not be possible to pool individual-level data across sites, in which case we will revert to pooling treatment coefficients and standard errors.

[21] In the case of rare events, we will compare our individual-level logistic model results with what one would obtain from a model based on the study-level log odds ratio point estimates and standard errors should we observe events affecting $<1\%$ of the study sample.

40 at the bottom 1 percent of the income distribution can expect about 15 fewer more years of life than a man at the top 1 percent of the income distribution, and this disparity appears to be growing over time (Chetty et al. (2016)). Furthermore, the relationship between household income and health emerges early in childhood and has been posited as one of the pathways through which intergenerational disadvantage is transmitted (Case, Lubotsky and Paxson (2002)).

Understanding the relationship between health, income, achievement, and the persistence of disadvantage across generations is key in designing effective policy interventions to improve economic mobility and relieve poverty. However, despite its clear importance, we still know relatively little about what drives this relationship and how health may be improved. For example, life expectancy among the poor is not strongly correlated with some leading explanations for socioeconomic disparities in health, such as access to health care, environmental quality, or labor market opportunities (Chetty et al. (2016)). It could be the case that poor health causes low income, or that poverty itself causes poor health, or that a third factor results in both poor labor market outcomes and poor health.

The current evidence relies primarily on quasi-experimental analysis to shed light on the relationship between income and health. Cesarini et al. (2016) explores variation in wealth driven by lottery winnings in a large Swedish administrative dataset. They find little association between lottery winnings and longevity among adults, although they do detect some small reductions for adults in the consumption of drugs that treat mental health disorders. Among the children of lottery winners, they find some evidence of increased health care utilization and lower obesity risk. These results suggest that the causal impact of wealth on health is relatively modest in Sweden. However, given the strong social safety net in Sweden, including publicly-funded universal health care and schooling, these results may not generalize to a setting like the United States where access to health care services is less equal.

Quasi-experimental studies set in the United States have generated mixed results. Evans and Garthwaite (2014) looks at changes in the Earned Income Tax Credit that increased payments to

women with two or more children relative to women with only one child. The authors find evidence that the self-reported health of the women who received higher payments improved, and reduced the incidence of risky biomarkers for this group. In contrast, Snyder and Evans (2006) studied changes in social security payments to the elderly and found that higher payments are actually associated with lower life expectancy. Other studies looking at very short-run changes in income also find perverse effects (Evans and Moore (2011), Gross and Tobacman (2014)).

7 Health Outcomes

This section describes the survey, biometric, and administrative measures that will be elicited to measure the effect of basic income on health.

In addition to the component measures listed below, we will report a single index for each family (following, e.g., Finkelstein et al. (2012)). This index measure will test the hypothesis of whether there was a change overall among these measures of within this family.

Some of the components to the index are not easily directly interpretable (e.g., those based on clinical scales used to diagnose mental illnesses). To make interpretation easier, we may also report coefficients on re-coded items within the component, even if these do not contribute to the overall index. These re-coded items will not themselves contribute separately to the index.

7.1 Family 1: Use of Office-based and Surgical Care

We will collect data in the in-person baseline, midline, and endline on the use of outpatient care. We group outpatient utilization separately from inpatient and emergency department use because predictions from the existing literature are straightforward for outpatient care: we expect use of such care would increase or stay the same. Inpatient and emergency department care, on the other hand, may increase or decrease with income; for example, if income allows participants to get more timely outpatient care, they may have less need of emergency department services.

Within the family of outpatient care, we will examine six groups of outcomes. Each group of outcomes consists of one or more outcome measures. The outcomes are:

1. Primary Care: a) has the respondent seen a general doctor or primary care doctor in the last 12 months;* b) the number of office visits in the last 12 months, c) had a usual source of care other than the ER, d) had a personal doctor or health care provider.

2. Specialist and Surgical Care: a) has the respondent seen a specialist in the last 12 months b) has the respondent had any surgery or surgical procedures, either inpatient or outpatient, c) the respondent had any visits to a mental health professional in the last 12 months.

3. Dental Care: a) has the respondent seen a dentist in the last 12 months, b) number of visits.

4. Preventive Care: In the last 12 months, has the respondent had a) a flu shot or nasal spray, b) a cholesterol test, c) PAP test (women only).

5. Total spending on health care excluding insurance premiums (derived from expenditure diary).

7.2 Family 2: Use of Hospital Care

We hypothesize that inpatient and emergency department (ED) care might increase if the additional income allows recipients to be able to afford more of such care, or it may decrease if recipients are able to access more timely care that is a substitute for inpatient and ED care. Our inpatient and emergency department care family is composed of two outcomes.

1. Hospital care: a) any hospital visit in the last 12 months, b) number of hospitalizations in the past 12 months.*

2. ED care: a) any ER visit in the last 12 months, b) number of ER visits in the past 12 months.*

Note that we will also link recipients to administrative data on hospital and ED use when possible. These linkages are discussed in the administrative data subsection below.

7.3 Family 3: Access to Health Care/Financial Strain

An important avenue through which income may affect health is by improving access to needed care or by reducing the financial strain placed on a household that needs medical care due to an illness or injury. This family is composed of the following outcome measures:

1. Insurance coverage: Whether or not the respondent has health insurance coverage*

2. Financial burden: a) how worried is the respondent about paying for medical costs b) did the respondent ever skip other bills to pay health bills, c) medical debt

3. Access to Care: a) did the respondent ever need medical care but could not afford it, b) did the respondent ever need dental care but could not afford it, c) usual source of care is a hospital ER, d) use alternative therapies to save money, d) did the respondent ever need mental health care but could not afford it.

4. Prescription Drug Access: a) did the respondent ever skip medication or take less medicine to save money, b) did the respondent ever delay filling a prescription to save money, c) did the respondent ask for a lower cost medication to save money.

As a separate secondary outcome we will also report the effect of the payments on the total amount spent on insurance premiums.

7.4 Family 4: Unhealthy Behaviors

Another avenue through which income may affect health is by changing the propensity of an individual to engage in unhealthy behaviors such as excessive drinking and smoking. The existing

literature on this is mixed, with clear negative correlations arising between income and, e.g., smoking (Casetta et al. (2017)), but with quasi-experimental evidence documenting perverse short-term effects of income receipt on such behaviors (e.g., Gross and Tobacman (2014)). These questions will be asked during the in-person baseline, midline, and endline, with some questions being asked through a mobile surveys depend on time available for the in-person surveys. This family is composed of the following outcome measures:

1. Alcohol interference:[22] a) in the past 30 days, how many days did you have at least one drink of any alcoholic beverage such as beer, wine, a malt beverage or liquor b) number of drinks per day on days when drinking, which we will recode as total number of drinks in past 30 days combining with responses to a),[23] c) number of days having 4 drinks or more d) did drinking or hangovers interfere with responsibilities at school, on the job, or at home, during past 12 months, e) during past 12 months, did drinking cause arguments or serious problems with family, friends, neighbors, or co-workers, f) in past 12 months, were you ever under the influence in a situation where you could get hurt.

2. Smoking behavior: a) do you currently use tobacco products, b) do you smoke cigarettes every day, c) number of cigarettes you smoke on a typical day.

3. Drug use: a) During the past 30 days, on how many days did you use marijuana? b) During the past 30 days, on how many days did you use painkillers that had not been prescribed to you, such as fentanyl, Oxycontin, Percocet, Vicodin, morphine or codeine, c) During the past 30 days, on how many days did you take or use illegal drugs such as heroin, cocaine, crack, etc, d) During the past 30 days, on how many days did you use a sedative such as Xanax or Valium? e) An indicator of illicit drug use in any category.

[22] Some of these questions were used in the Moving to Opportunity Interim Evaluation.

[23] In cleaning the initial year of survey data we noticed a small fraction of respondents (about 0.1 percent) inputted both 30 days of drinking and 30 drinks per day. We suspect this may have been inputted in error, so we have recoded these values at the 99th percentile of the distribution.

7.5 Family 5: Health Investments

Another way in which income could affect health outcomes is by changing the the health investments made by recipients. The health investments family is composed of the following outcome measures:

1. Exercise: a) number of times exercising in last week* b) amount of time exercising in last week (survey), c) minutes reported spent exercising or doing physical activity on time diary days (time diary).

2. Sleep: a) On average, how many hours of sleep do you get in a 24-hour period? b) how do you rate your sleep quality overall? c) Quantity of sleep derived from time diary.

We will treat amount of time exercising and minutes spent exercising as potentially susceptible to outliers.

7.6 Family 6: Self-reported Physical Health

This family will incorporate several measures of self-assessed physical health. These measures were drawn from validated health survey measures such as the SF-12 (developed by the RAND Medical Outcomes Study see, e.g., McHorney et al. (1992)), as well as previous studies such as the Oregon Health Insurance Experiment. We include the following outcomes:

1. Health poor/fair/good/very good/excellent.

2. Does health limit ("no, not limited at all", "yes, limited some of the time", "yes, limited a lot"): moderate activities, climbing several flights of stairs.

3. Ranging from "all of the time" to "none of the time"), in the past 4 weeks: did you accomplish less than you would like due to physical health, were you limited in the kind of work or other activities you could do as a result of your physical health, how much of the time

has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)

4. Ranging from “not at all” to “extremely”, during past 4 weeks, how much did pain interfere with normal work (“not at all” to “extremely”).

5. Physical Health Not Good: During the past 30 days, how many days was your physical health not good?

Each bullet point listed above will be a component inputted into the physical health index, oriented such that higher values indicate better health, with ordinal responses treated as continuous for the purpose of the index (e.g. Health poor, fair, good, very good, excellent will be coded as 1,2,3,4, and 5). In order to improve our ability to interpret these results meaningfully, we will also report, as secondary outcomes, the following indicator variables derived from the above:

- Health is very good or excellent.*
- No health limitations related to moderate activities or climbing stairs.
- Pain interferes not at all or only a little bit.
- Physical or emotional problems interfere “none of the time” or only “a little bit.”

These binary outcomes will not themselves be inputted into the index.

Finally, for the duration that COVID-19 remains a serious public health issue, we will ask about whether the participant has been diagnosed with or experienced the symptoms related to COVID-19. This will be reported as a separate exploratory measure if COVID-19 remains a significant illness for the majority of the study period.

7.7 Family 7: Self-reported Mental Health

Due to time constraints and participant burden, some of these classical self-reported health scales have been shortened. We will include components from the SF-12 (developed by the RAND Medical Outcomes Study see, e.g., McHorney et al. (1992)). The Kessler 6 is composed of 6 questions that measure general psychological distress (Kessler et al. (2002)). The Perceived Stress Scale is a widely-used measure of stress that was developed in Cohen, Kamarck and Mermelstein (1983). In addition, we include measures used in the Oregon Health Insurance Experiment for comparability. All survey outcomes will be collected during the in-person surveys, with the exception of the last two outcome measures, which will be collected at baseline using a mobile survey. This family is composed of the following outcome measures:

1. Ranging from "all of the time" to "none of the time," in the past 4 weeks: did you accomplish less than you would like due to emotional problems, did you do work or other activities less carefully as the result of emotional problems, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.) ("all of the time" to "none of the time").

2. Kessler 6. We will report the score itself as the main outcome and the sub-questions as secondary outcomes. The components are: During the past 30 days, how often did you feel a) nervous, b) hopeless, c) restless or fidgety, d) depressed e) like everything was an effort, f) worthless

3. Perceived Stress Scale (14-question scale; see appendix).

4. Anxiety: GAD-2 Anxiety Screen. We will report the score itself as the main outcome and the sub-questions as secondary outcomes. The components are: In the last 2 weeks, how often (4 point scale ranging from "not at all" to "nearly every day") have you been bothered by: Feeling nervous, anxious or on edge and Not being able to stop or control worrying,

5. Depression: PHQ-9 Depression Screen. We will report the score itself as the main outcome, and sub-questions as secondary outcomes. The components are: In the last 2 weeks, how often (4 point scale ranging from "not at all" to "nearly every day") have you been bothered by: Little interest or pleasure in doing things?, Feeling down, depressed, or hopeless?, Trouble falling or staying asleep, or sleeping too much?, Feeling tired or having little energy?, Poor appetite or overeating?, Feeling bad about yourself — or that you are a failure or have let yourself or your family down?, Trouble concentrating on things, such as reading the newspaper or watching television?, Moving or speaking so slowly that other people could have noticed? Or so fidgety or restless that you have been moving a lot more than usual?, Thoughts that you would be better off dead, or thoughts of hurting yourself in some way?

6. Mental Health Not Good: During the past 30 days, how many days was your mental health not good?

Each outcome listed in 1. above will be a separate component inputted into the physical health index, oriented in the same direction, with ordinal responses treated as continuous for the purpose of the index (e.g. Health poor, fair, good, very good, excellent will be coded as 1,2,3,4, and 5). The Kessler 6, Perceived Stress Scale, GAD-2 Anxiety Screen, and PHQ-9 depression screen will be calculated following the scoring algorithm for each of these batteries with input measures reported in an appendix and treated as secondary. In order to improve our ability to interpret these results, we will also report the following indicator variables derived from the above. These will be considered secondary outcomes:

- K6 in the "high distress" range (greater than or equal to 13).
- Perceived Stress Scale in the "high stress" range (greater than or equal to 27).*
- GAD-2 indicates anxiety (greater than or equal to 3).
- PHQ-9 in the "moderately severe" to "severe" range (greater than or equal to 15).

7.8 Family 8: Nutrition and Food Security

We will elicit information on food security using a slightly modified version of the 6-item food security module developed by the USDA. This measure has been shown to perform well with minimal respondent burden (Blumberg et al. (1999)). We will ask about both the food security of the participant and, for those with children, the food security of their children.

In addition, we will conduct 24-hour nutrition diaries using the ASA 24 tool. Respondents will receive additional compensation (amount to be determined) for completing this nutrition diary, which will take approximately 20 minutes.

Finally, we will also ask two ad hoc questions about nutrition behavior about consuming sugar-sweetened soda and eating at fast food establishments.

1. Healthy Eating Index (derived from food diaries).

2. Food Security Index: a) I worry about whether food would run out, b) the food I bought didn't last, c) I couldn't afford to eat balanced meals, d) I cut size of meals or skipped meals because there wasn't enough, e) how often did (cutting meals or skipping meals) happen, f) did you ever eat less than you should because there wasn't enough money for food.

3. Nutrition: a) during the past week, how many times did you eat at a fast food establishment b) during the past week, how often did you drink regular soda or pop that contains sugar?

We will also investigate the following outcomes, but these will be considered exploratory given the large number of outcomes and we will winsorize these at the 99th percentile given the open ended nature of the questions:

- Total calories.
- Cups of vegetables.
- Cups of whole fruit.

- Grams of sugar consumed.
- Teaspoon equivalent of "added" sugar.
- Grams of protein consumed.
- Grams of sodium consumed.
- Grams of cholesterol consumed.
- Grams of carbohydrates consumed.
- Grams of saturated fat consumed.
- Grams of alcohol.

7.9 Family 9: Biomarkers

In addition to these self-reported measures of health and behavior, we will also collect clinical measures of health using specially trained enumerators which will form their own family of outcomes. Because of the invasiveness of the procedure, we plan to wait until midline to collect blood-based biomarkers. During the midline survey, participants will receive an additional payment of $50 if they agree to participate in the blood spot collection.

We will collect the following measures. Note that each outcome measure will be reported separately (i.e., we will report the effect on total cholesterol, HDL, LDL, and the indicator of high cholesterol). Note we are pre-specifying here what we hope to collect; however, due to the expense of collecting these data we may need to collect fewer or different measures depending on funding.:

1. BMI: a) Constructed from height and weight, b) indicator that BMI is in the obese range [baseline, endline]

2. Blood pressure: a) diastolic, b) systolic, c) an indicator that the individual is in the hypertensive range (130/80 or higher), d) an indicator that the individual is in the elevated blood pressure range (120-129 systolic with diastolic less than 80) [baseline, endline]

3. A1c: a) overall, b) an indicator that an individual is in the diabetic range (A1c $\geq$ 6.5 %), c) an indicator that an individual is in the pre-diabetic range (A1c $\geq$ 5.7). collected from blood draws [endline]

4. Lipid panel: collected from blood draws; total cholesterol, HDL, LDL, an indicator that cholesterol is in the high cholesterol range. [endline]

5. C-reactive protein: collected from blood draws [endline]

6. American Heart Association "Ideal Cardiovascular Health" Index: we will combine these three biometric indicators with the survey data on exercise, smoking status, and nutrition to construct the AHA "Ideal Cardiovascular Health" Index. (Lloyd-Jones DM (2010)).

Each of the enumerated items above will be combined into its own index and FDR adjustments will be applied across these 7 summary outcomes. Individual items will also be reported to improve the interpretability of the index estimates.

7.10 Family 10: Hospital/ED Administrative Data Measures

In addition to the survey-elicited data and the biomarkers described above, we will merge individuals with administrative data on hospital and emergency department care where feasible. We will exclude hospitalizations related to labor and delivery and explore the impact of the unconditional cash transfers on fertility in a separate analysis. Administrative data on hospital and ED usage will be considered its own family of outcomes, following, e.g., the Oregon Health Insurance Experiment. If there are substantial lags in obtaining these data, we may treat this family as its own separate paper/topic or group with with a different topic.

From the hospital and ED data, we will have the following outcome measures:

1. Hospital care use: a) # of hospitalizations in the last 12 months, b) total number of hospital days in the last 12 months, c) total hospital charges.

2. ED use: a) # of ED visits in the last 12 months, b) total ED charges in last 12 months.

3. # of ED visits for non-urgent, urgent but primary-care treatable, and urgent but preventable conditions (as assigned using the NYU ED Algorithm).

4. # of Preventable Hospitalizations (as categorized using the HCUP Prevention Quality Indicator Software).

If we estimate significant effects on # of hospitalizations or # of ED visits, we will further explore visits by diagnosis group based on ICD-10 groupings. Similarly, if we find increased hospital or ED charges, we will further explore the source of charges (e.g. radiology, room and board, etc). If other types of administrative data on health care utilization are available (e.g. data on ambulatory surgeries, prescription drug usage, etc), we will analyze those outcomes in a similar way.

If we are able to obtain Medicaid claims data, and do not find meaningful changes into/out of the Medicaid program associated with treatment, we will also use claims data to characterize use of medical care among Medicaid enrollees if such data are available for linkages.

7.11 Family 11: Mortality

Our final health measure will be mortality. This will be measured by matching participants to National Death Index data or other administrative death records. We will examine:

1. Overall mortality.

2. Health care amenable mortality.

If we find a significant effect on either overall mortality or health care amenable mortality, we will also explore mortality by ICD-10 body group classification. These cause of death specific analyses will be considered secondary/exploratory.

8 Heterogeneity Analysis

All heterogeneity analyses are considered "exploratory." Subgroup analysis will be conducted by:

- Age group.
- Income at baseline (less than 100% FPL vs. 100% FPL and above).
- Baseline health "very good" or "excellent" versus "good," "fair," or "poor."
- Insurance status at baseline.
- Indication that the respondent did not get needed care due to costs at baseline.

9 Conclusion

9.1 Known Limitations

Our study has several limitations. First, the limited nature of the RCT does not permit us to simulate the macroeconomic conditions of the government introducing an unconditional cash transfer program to all residents of the United States who meet broad eligibility criteria. If recipients are spending the money helping friends and family who would receive their own cash transfer under the policy, the treatment is diluted and the likelihood of the hypothesized effects is undermined. Similarly, the dispersed sample precludes our ability to capture the multipliers and general equilibrium effects identified in the theoretical literature and observed in studies in developing countries. The dispersed study also precludes studying the effect of sustained unconditional cash transfers on

cultural attitudes towards work and other social spillovers. Despite these limitations, we selected a geographically dispersed population for several reasons. Most importantly, the intervention is very expensive and our sample size is constrained by the budget. A geographically saturated study would likely cost billions of dollars, and we would not have enough statistical power to detect effects with a geographically saturated study with our budget.

A second limitation is the time-bound nature of our treatment. The 3-year timespan of the intervention is obviously not the same as a perceived long-term guarantee, and individuals may behave differently knowing that the transfers are time-limited (Hoynes and Rothstein 2019). Nevertheless, a study at the scale proposed in this analysis plan will allow us to provide timely evidence to inform ongoing policy debates and future research on this topic.

References

Akee, Randall K Q, William E Copeland, Gordon Keeler, Adrian Angold and E Jane Costello. 2010. "Parents' Incomes and Children's Outcomes: A Quasi-Experiment Using Transfer Payments from Casino Profits." *American Economic Journal Applied Economics* 2(1):86–115.

Akee, Randall, William Copeland, E Jane Costello and Emilia Simeonova. 2018. "How Does Household Income Affect Child Personality Traits and Behaviors?" *American Economic Review* 108(3):775–827.

Anderson, Michael L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association* 103(484):1481–1495.

Banerjee, A, R Hanna, G Kreindler and B A Olken. 2017. "Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs." *The World Bank Research Observer* 32(2):155–184.

Baumol, William J. 1974. *The Journal of Human Resources* 9(2):253–264, title = An Overview of the Results on Consumption, Health, and Social Behavior.

Becker, G S. 1965. "A Theory of the Allocation of Time." *The Economic Journal* 75(299):493–517.

Ben-Shalom, Y, R Moffitt and J K Scholz. 2012. An assessment of the effectiveness of anti-poverty programs in the United States. In *The Oxford Handbook of the Economics of Poverty*, ed. P Jefferson. New York: Oxford University Press pp. 709–749.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball and Nichole Szembrot. 2014. "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference." *American Economic Review* 104(9).

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal statistical society: series B (Methodological)* 57(1):289–300.

Bertrand, M, S Mullainathan and D Miller. 2003. "Public Policy and Extended Families: Evidence from Pensions in South Africa." *World Bank Economic Review* 17(1):27–50.

Bhargava, Saurabh and Dayanand Manoli. 2015. "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment." *American Economic Review* 105(11):3489–3529.

Blattman, C, N Fiala and S Martinez. 2014. "Generating Skilled Self-employment in Developing Countries: Experimental Evidence from Uganda." *The Quarterly Journal of Economics* 129(2):697–752.

Bloniarz, Adam, Hanzhong Liu, Cun-Hui Zhang, Jasjeet S Sekhon and Bin Yu. 2016. "Lasso ad-

justments of treatment effect estimates in randomized experiments." *Proceedings of the National Academy of Sciences* 113(27):7383–7390.

Blumberg, S, K Bialostosky, W Hamilton and R Briefl. 1999. "The effectiveness of a short form of the household food security scale." *American Journal of Public Health* 89:1231–1234.

Broockman, David E, Joshua L Kalla and Jasjeet S Sekhon. 2017. "The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs." *Political Analysis* 25(4):435–464.

Burtless, G. 1986. The work response to a guaranteed income: A survey of experimental evidence. In *Lessons from the income maintenance experiments*, ed. A H Xavier. Boston, MA: Federal Reserve Bank of Boston and The Brookings Institution pp. 22–54.

Case, Anne, Darren Lubotsky and Christina Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review* 92(5):1308–1334.

Casetta, Brunilda, Alejandro Videla, Ariel Bardach, Paola Morello, Natalie Soto, Kelly Lee, Paul Anthony Camacho, Moquillaza Hermoza, Victoria Rocío and Agustín Ciapponi. 2017. "Association Between Cigarette Smoking Prevalence and Income Level: A Systematic Review and Meta-Analysis." *Nicotine Tobacco Research* 19(12):1401–1407.

Cesarini, David, Erik Lindqvist, Robert Östling and Björn Wallace. 2016. "Wealth, health, and child development: Evidence from administrative data on Swedish lottery players." *The Quarterly Journal of Economics* 131(2):687–738.

Chan, Marc K and Robert Moffitt. 2018. "Welfare Reform and the Labor Market." *Annual Review of Economics* 10(1):347–381.

Chetty, R, M Stepner, S Abraham and et al. 2016. "The association between income and life expectancy in the united states, 2001-2014." *JAMA* 315(16):1750–1766.

Chetty, Raj and Nathaniel Hendren. 2018*a*. "The Impact of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *Quarterly Journal of Economics* 113(3).

Chetty, Raj and Nathaniel Hendren. 2018*b*. "The Impact of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *Quarterly Journal of Economics* 113(3).

Cohen, Sheldon, Tom Kamarck and Robin Mermelstein. 1983. "A Global Measure of Perceived Stress." *Journal of Health and Social Behavior* 24(4):385–396.

Congressional Research Service. 2019. "Real Wage Trends, 1979 to 2018." CRS Report.

Costello, E Jane, Scott N Compton, Gordon Keeler and Adrian Angol. 2003. "Relationships Between Poverty and Psychopathology: A Natural Experiment." *Journal of the American Medical Association* 290(15):2023–2029.

Danziger, Sandra K. 2010. "The Decline of Cash Welfare and Implications for Social Policy and Poverty." *Annual Review of Sociology* 36(1):523–545.

de Mel, S, D McKenzie and C Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." *The Quarterly Journal of Economics* 123(4):1329–1372.

DellaVigna, Stefano, Devin Pope and Eva Vivalt. 2019. "Predict Science to Improve Science." *Science* 366(6464):428–429.

Dorn, David, Gordon Hanson, Kaveh Majlesi et al. 2016. Importing political polarization? the electoral consequences of rising trade exposure. Technical report National Bureau of Economic Research.

Eissa, N and J Liebman. 1996. "Labor Supply Response to the Earned Income Tax Credit." *Quarterly Journal of Economics* 111:605–637.

Eissa, Nada and Hilary Hoynes. 2004. "Taxes and the labor market participation of married couples: the earned income tax credit." *Journal of Public Economics* 88:1931–1958.

Elesh, D and M J Lefcowitz. 1977. The effects of health on the supply of and returns to labor. In *The New Jersey Income-Maintenance Experiment*, ed. H W Watts and A Rees. New York, NY: Academic Press pp. 289–320.

Evans, W, B Wolfe and N Adler. 2012. "The SES and Health Gradient: A Brief Review of the Literature." in The Biological Consequences of Socioeconomic Inequalities, eds. B. Wolfe, W. Evans, and T. E. Seeman (New York, Russell Sage Foundation: 2012).

Evans, William N and Craig L Garthwaite. 2014. "Giving mom a break: The impact of higher EITC payments on maternal health." *American Economic Journal: Economic Policy* 6(2):258–290.

Evans, William and Timothy Moore. 2011. "The short-term mortality consequences of income receipt." *Journal of Public Economics* 95:1410–1424.

Fafchamps, M and S Quinn. 2017. "Aspire." *The Journal of Development Studies* 53(10):1615–1633.

Feinberg, Robert M and Daniel Kuehn. 2018. "Guaranteed Nonlabor Income and Labor Supply: The Effect of the Alaska Permanent Fund Dividend." *The B.E. Journal of Economic Analysis & Policy* 18(3):350–13.

Finkelstein, Amy and Matthew Notowidigdo. 2019. "Take-up and Targeting: Experimental Evidence from SNAP." *Quarterly Journal of Economics* 134(3).

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph Newhouse, Heidi Allen and Katherine Baicker. 2012. "The Oregon Health Insurance Experiment: Evidence from the First Year." *Quarterly Journal of Economics* 127(3):1057–1106.

Fiszbein, A, N Schady, F H G Ferreira, M Grosh, N Keleher, P Olinto and E Skoufias. 2009. "Conditional Cash Transfers : Reducing Present and Future Poverty." World Bank Policy Research Report.

Forget, Evelyn L. 2011. "The town with no poverty: the health effects of a Canadian guaranteed annual income field experiment." *Canadian Public Policy* 37(3):283–305.

Forget, Evelyn L. 2013. "New questions, new data, old interventions: The health effects of a guaranteed annual income." *Preventive Medicine* 57(6):925–928.

Greenberg, David and Harlan Halsey. 1983. "Systematic misreporting and effects of income maintenance experiments on work effort: evidence from the Seattle-Denver experiment." *Journal of Labor Economics* 1(4):380–407.

Gross, Tal and Jeremy Tobacman. 2014. "Dangerous Liquidity and the Demand for Health Care Evidence from the 2008 Stimulus Payments." *Journal of Human Resources* 49(2):1410–1424.

Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, Sandra González-Bailón, Edward Kennedy, Young Mie Kim, David Lazer, Devra Moehler, Brendan Nyhan, Carlos Velasco Rivera, Jaime Settle, Daniel Robert Thomas, Emily Thorson, Rebekah Tromble, Arjun Wilkins, Magdalena Wojcieszak, Beixian Xiong, Chad Kiewiet de Jonge, Annie Franco, Winter Mason, Natalie Jomini Stroud and Joshua A. Tucker. 2023. "Reshares on social media amplify political news but do not detectably affect beliefs or opinions." *Science* 381(6656):404–408.

Hanushek, Eric A et al. 1986. Non-labor-supply responses to the income maintenance experiments. In *Lessons from the Income Maintenance Experiments: Proceedings of a Conference Held at Melvin Village, New Hampshire*. pp. 106–21.

Haushofer, Johannes and Jeremy Shapiro. 2016. "The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya*." *The Quarterly Journal of Economics* 131(4):1973–2042.

Hausman, Jerry A and David A Wise. 1979. "Attrition bias in experimental and panel data: the Gary income maintenance experiment." *Econometrica: Journal of the Econometric Society* pp. 455–473.

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

Hoynes, Hilary and Jesse Rothstein. 2019. "Universal Basic Income in the United States and Advanced Countries." *Annual Review of Economics* 11:929–958.

Imbens, G W, D B Rubin and B I Sacerdote. 2001. "Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players." *American Economic Review* 91(4):778–794.

Jones, Damon and Ioana Marinescu. 2018. "The Labor Market Impacts of Universal and Permanent Cash Transfers: Evidence from the Alaska Permanent Fund." NBER Working Paper 24312.

Kaluzny, Richard L. 1979. "Changes in the Consumption of Housing Services: The Gary Experiment." *The Journal of Human Resources* 14(4):496–506.

Kangas, Olli, Signe Jauhiainen, Miska Simanainen and Minna Yilkanno. 2019. "The basic income experiment 2017-2018 in Finland. Preliminary results." Working Paper.

Kangas, Olli, Signe Jauhiainen, Miska Simanainen and Minna Yilkanno. 2020. Suomen perustulokokeilun arviointi. Technical report Ministry of Social Affairs and Health Helsinki: .

Keeley, M C. 1981. *Labor supply and public policy: A critical review*. New York NY: Academic Press.

Keeley, Michael C. 1980*a*. "The Effect of a Negative Income Tax on Migration." *The Journal of Human Resources* 15(4):695–706.

Keeley, Michael C. 1980*b*. "The Effects of Negative Income Tax Programs on Fertility." *The Journal of Human Resources* 15(4):675–694.

Kehrer, B H and C M Wolin. 1979. "Impact of income maintenance on low birth weight: Evidence from the Gary Experiment." *The Journal of Human Resources* 14(4):434–462.

Kessler, R, G Andrews, L Colpe, E Hiripi, D Mroczek, S Normand and A Zaslavsky. 2002. "Short screening scales to monitor population prevalences and trends in non-specific psychological distress." *Psychological Medicine* 32(32):959–979.

Kleven, Henrik. 2018. "Taxation and Labor Force Participation: The EITC Reconsidered." NBER conference presentation.

Kleven, Henrik. 2020. "The EITC and the Extensive Margin: A Reappraisal." NBER Working Paper No. 26405.

Kline, Patrick and Melissa Tartari. 2016. "Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach." *American Economic Review* 106(4):972–1014.

Kling, Jeffrey, Jeffrey Liebman and Lawrence Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1):83–119.

Lloyd-Jones DM, Hong Y, Labarthe D et al. 2010. "Defining and setting national goals for cardiovascular health promotion and disease reduction: The American Heart Association's strategic Impact Goal through 2020 and beyond." *Circulation* 121(4):586–613.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir and Jiaying Zhao. 2013. "Poverty Impedes Cognitive Function." *Science* (341):976–980.

Maynard, R. 1977. "The effects of the rural income maintenance experiment on the school performance of children." *American Economic Review* 67(1):370–375.

Maynard, Rebecca A. and Richard J. Murnane. 1979. "The Effects of a Negative Income Tax on School Performance: Results of an Experiment." *The Journal of Human Resources* 14(4):463–476.

McHorney, Colleen A., John E. Ware, William Rogers, Anastasia E. Raczek and J. F. Rachel Lu. 1992. "The Validity and Relative Precision of MOS Short- and Long-Form Health Status Scales and Dartmouth COOP Charts: Results from the Medical Outcomes Study." *Medical Care* 30(5):MS253–MS265.

McKenzie, D. 2015. "Identifying and spurring high-growth entrepreneurship: Experimental evidence from a business plan competition." World Bank Policy Research Working Paper No. 7391.

Meager, Rachael. 2019. "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments." *American Economic Journal: Applied Economics* 11(1):57–91.

Meyer, Bruce. 2010. "The Effects of the Earned Income Tax Credit and Recent Reforms." *Tax Policy and the Economy* 24(1).

Meyer, Bruce and Dan Rosenbaum. 2001. "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *Quarterly Journal of Economics* 116(3).

Miller, Cynthia, Rhiannon Miller, Nandita Verma, Nadine Dechausay, Edith Yang, Timothy Rudd, Jonathan Rodriguez and Sylvie Honig. 2016. "Effects of a Modified Conditional Cash Transfer Program in Two American Cities.".

MSchady, N and J Rosero. 2007. "Are Cash Transfers Made to Women Spent Like Other Sources of Income?" World Bank Policy Research Working Paper No. 4282.

Murnane, R, R Maynard and J Ohls. 1981. "Home resources and children's achievement." *Review of Economics and Statistics* 63(3):369–377.

Nichols, A. and J. Rothstein. 2016. The Earned Income Tax Credit. In *Economics of Means-Tested Transfer Programs in the United States*, ed. R.A. Moffit. Chicago: University of Chicago Press pp. 137–218.

O'Connor, J. Frank and J. Patrick Madden. 1979. "The Negative Income Tax and the Quality of Dietary Intake." *The Journal of Human Resources* 14(4):507–517.

Piketty, Thomas, Emmanuel Saez and Gabriel Zucman. 2019. "Distributional National Accounts: Methods and Estimates for the United States." Forthcoming, Quarterly Joural of Economics.

Rea, S A. 1977. "Investment in human capital under a negative income tax." *Canadian Journal of Economics* 10(4):607–620.

Riccio, James A and Cynthia Miller. 2016. "New York City's first conditional cash transfer program: What worked, what didn't.".

Robins, Philip. 1985. "A Comparison of the Labor Supply Findings From the Four Negative Income Tax Experiments." *Journal of Human Resources* 20(4):567–582.

Salehi-Isfahani, Djavad and Mohammad H Mostafavi-Dehzooei. 2018. "Cash transfers and labor supply: Evidence from a large-scale program in Iran." *Journal of Development Economics* 135:349–367.

Shaefer, H Luke and Kathryn Edin. 2013. "Rising Extreme Poverty in the United States and the Response of Federal Means-Tested Transfer Programs." *Social Service Review* 87(2):250–268.

Snyder, Stephen and William Evans. 2006. "The Effect Of Income On Mortality: Evidence From The Social Security Notch." *Review of Economics and Statistics* 88(3):482–495.

U.S. Department of Health and Human Services. 2016. "Poverty in the United States: 50-Year Trends and Safety Net Impacts." US Department of Health and Human Services, ASPE Report.

Vivalt, Eva. 2020. "How Much Can We Generalize From Impact Evaluations?" *Journal of the European Economics Association* 18(6):3045–3089.

Weiss, Y, A Hall and F Dong. 1980. "The Effect of Price and Income on Investment in Schooling." *The Journal of Human Resources* 15(4):611–640.

Widerquist, Karl, Robert Levine, Alice O'Conner, Harold Watts, Robinson Hollister and Walter Williams. 2005. "A retrospective on the negative income tax experiments: Looking back at the most innovative field studies in social policy." *The ethics and economics of the basic income guarantee* pp. 95–106.

Wolfe, Barbara, Jessica Jakubowski, Robert Haveman and Marissa Courey. 2012. "The income and health effects of tribal casino gaming on American Indians." *Demography* 49(2):499–524.