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

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

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 individuals 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 employment outcomes.

[*]Assistant Professor, University of Toronto. eva.vivalt@utoronto.ca, http://evavivalt.com
[†]Assistant Professor, University of Illinois at Urbana-Champaign
[‡]Associate Professor, University of California, Berkeley
[§]Assistant Professor, University of Michigan Ross School of Business
[¶]Research Director, Open Research Lab

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, however, 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 Sustained Unconditional Cash Transfers and Employment Outcomes

Receiving sustained unconditional cash transfers could affect both labor market supply and labor market outcomes.

There have been many studies of labor supply elasticities; Keane (2011) and McClelland and Mok (2012) provide recent reviews focusing on responsiveness to wages and taxes. Studies of

[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.

the Earned Income Tax Credit and Negative Income Tax experiments tend to find small, negative effects on labor supply (e.g. Eissa and Liebman 1996, Price and Song 2017). Most meta-analyses of cash transfer programs in a developing country context also find mixed, sometimes negative effects (Kabeer and Waddington 2015, Bastagli et al. 2016).

In contrast to much of the existing literature, our focus is on a relatively young sample whose career trajectories may be more malleable. Some of the labor supply effects previously observed in other studies were driven by older adults retiring sooner as a result of the transfers (Price and Song 2017), an effect that is unlikely to occur in our setting. We may expect to see greater career changes, potentially including some individuals leaving employment to pursue further training or some electing to stay unemployed longer while searching for a better job rather than settling for an available but less desirable job.

This study also has the opportunity to speak to an on-going debate in the literature as to whether expansions of the social safety net lengthen unemployment but ultimately result in better job matches between job seekers and employers. This literature has historically focused on changes in the generosity of employment insurance, but similar arguments could apply to job search under the increased security of monthly cash transfers. The literature, mostly from other countries (e.g. Germany, Austria), is mixed, with Centeno (2004), Caliendo, Tatsiramos and Uhlendorff (2012), and Nekoei and Weber (2017) finding that more generous benefits enable better jobs, while another strand of the literature finds no such effects (e.g. Card, Chetty and Weber 2007, Lalive 2007, van Ours and Vodopivec 2008). In addition to drawing from other countries with more generous social safety nets, past papers in this literature have often had limited information on job quality, inferring job quality from income or the duration that the post-unemployment job was held. In contrast, we will have access to a rich array of variables we can use to identify quality of employment.

To capture possible shifts in job quality, we thus ask very detailed questions about the nature of respondents' current jobs, paying particular attention to their stability and formality. We also focus on whether their jobs are "high quality" as indicated by the hourly wage and work hours available, non-wage benefits such as whether training is offered by an employer, and a variety of self-reported measures of quality of work life.

Given the importance that has been placed in the literature on disability and the role of disability in long-term unemployment outcomes (Autor and Duggan 2003), we also probe whether respondents are disabled and, if so, if they expect to return to work in the future. This could potentially serve as a valuable explanatory variable. Importantly, at the end of the program, after the transfers are stopped, we will also follow up to see what the likelihood is of a participant applying for disability or other benefits; however, these post-program outcomes will be detailed in a separate pre-analysis plan. We additionally catalog barriers to employment that could be mitigated by a cash transfer, such as whether the respondent has had to miss work due to illness or lack of transportation. These outcomes may be more immediately responsive to cash transfers than disability.

The security of regular cash transfers could help to make self-employment more attractive, so we will also consider entrepreneurial activity.

We will also analyze human capital formation: i.e., if we do observe recipients working slightly less, what are they doing with their time instead? Are they pursuing formal or informal education to gain skills which could help them obtain better jobs in the future? The effects of sustained cash transfers on future labour market outcomes could help to address concerns about the cost of extending such transfers as a social policy (Hoynes and Rothstein 2019). Other outcomes that we are measuring in separate pre-analysis plans that can contribute to our understanding of long-run effects are child development outcomes, time use outcomes, outcomes predictive of long-run health, and outcomes relating to social problems that are very expensive to address such as homelessness, drug use and crime.

Due to the literature typically finding large empirical differences in employment by sex, we will disaggregate each family's index and several key outcomes noted in the text by this variable.[22]

[22] Unless otherwise noted, those self-identifying as "other" at baseline will be pooled with those self-identifying as

These disaggregated results will be presented as main results along with estimates pooled across sex. For some outcomes flagged in the text, we will separately consider heterogeneity by age or whether the participant had a college degree at baseline.

7 Employment Outcomes: Survey Measures

This section describes the survey measures that will be used in each family of outcomes.

As noted, apart from considering the measures in each family separately, we will report a single index for each family (following, e.g., Finkelstein et al. 2012). On occasion, an outcome measure will be pre-specified to be excluded from this index as it may be useful descriptively but not make sense to include in an index.

In addition to considering families of outcomes through indices, we will report results tables that disaggregate by each outcome measure. We will focus particular attention on three outcomes critical for policy decisions: the employment rate, annual hours worked, and annual earnings. We will also include discrete choice questions and construct models of job search and selection to test whether individuals find better jobs if they can search longer or if this is associated with negative outcomes. Approximately 30% of our sample is unemployed at baseline, making progression from unemployment particularly important.

We will also consider several outcomes that are descriptive and will not be included in any family index: type of employer (e.g. private sector), industry of work, and job title. Industries can be linked through NAICS codes and job titles through SOC and OCC codes to data from the Bureau of Labor Statistics (BLS) and the Economic Census to explore whether there is growth in occupations with higher or lower average monthly unemployment rates or a lot of turnover, as represented by the share of hires and separations out of total employment.

While the following sections focus mostly on the questions asked in the SRC surveys, as we ex-

"female" for analyses.

pect these surveys to have the highest response rate, related mobile survey questions are sometimes available and will be used for additional power.

All variables will be converted to unconditional variables wherever possible and exceptions flagged in the text. For example, if a question about wages is not asked to someone who is unemployed, we will consider the unemployed person to have wages of $0 for the sake of analysis. This approach gives us the unconditional impact on wages, avoiding selection. In this example, it will also be important to consider effects on employment for the sake of interpretation.

7.1 Family 1: Labor Supply Elasticity

We will ask respondents to provide details of their current employment and work histories at baseline/midline/endline, as well as asking related questions, described below.

1. Labor supply elasticity: a) whether the respondent is currently employed (G2: 1=working for pay, 0=unemployed and looking for work or unemployed and not looking for work) (*); b) how many hours the respondent is working total, across their main job and any other jobs (I7+J9* for employed; 0 for unemployed) (*). While the items above refer to SRC survey questions, as described elsewhere in the pre-analysis plan the preferred data source will be administrative records. In this case, we will consider both tax return data and UI data. Both will be presented, though the tax return data will be preferred (except for analyses requiring quarterly data).[23] In exploratory analysis, we will also consider the respondent's reasons for unemployment, if unemployed (H2). B21 will be used in an alternative specification for item b); it contains less information but may be less sensitive to outliers.

As an exception to our typical protocol, we will also present estimates of the income elasticity

[23]Finally, it should be noted that while the SRC data is very comprehensive, in a handful of cases individuals answered whether they were employed in the mobile baseline survey (work_1) but not in the SRC baseline. In the rare cases in which individuals answer a mobile baseline/midline/endline survey question but not the corresponding SRC question, their answer to that mobile baseline/midline/endline survey question will be combined with the SRC data before analysis.

of labor supply (η) and estimates for the extensive (η_e) and intensive (η_i) margin separately from the index. For greater comparability with past studies on the NIT experiments and the EITC (e.g., Eissa and Hoynes 2004, Robins 1985), our preferred measures for this purpose will be the average annual hours worked and the employment rate; we will also plot the quarterly employment rate to highlight any trends.

We will calculate the income elasticity of labor supply as $\eta = \frac{Y}{L}\frac{\partial L}{\partial v}$ where Y is net-of-tax total income, L is labor supply, and v is virtual income, taking a conventional approach (e.g., Rothstein 2010). We will also calculate the extensive and intensive elasticities $\eta_e = \frac{Y}{p}\frac{\partial p}{\partial r}$ where p is participation and r is non-labor income and $\eta_i = \frac{Y}{h}\frac{\partial h}{\partial v}$, where h is labor hours. In considering income changes, we will consider changes from baseline income.

Of the two lottery studies that are perhaps the closest to this paper, Cesarini et al. (2016) finds support for a "separate spheres" household model (Lundberg and Pollak 1993) while Golosov et al. (2021) cannot reject a unitary household model. We will present elasticities both using individual-level data on the participant and their partner, if any, as well as household-level income. The pre-analysis plan on intrahousehold outcomes describes how the estimates of partner income, labor force participation and labor hours will be measured in detail; we will prioritize administrative data for income and participation and baseline/midline/endline survey measures for labor hours, analogous to how we construct participant outcomes.

Recall that in our setting, participants were eligible to apply for the program based in part on falling below a household income threshold, and we randomized which individual in the household the recruitment mailer was directed to. However, it is still possible that who applied with the household, and hence receives the transfers, is non-random.[24] This is important because it means that if we observe differences in responses between participants and their partners, it may be due to selection. We will test whether who we addressed the mailer to predicts who applied. We will

[24] If multiple adults within the same household applied, one would have been randomly selected to participate since one individual from each address was randomly selected to form the active survey group, but this was fairly rare.

also ideally test for baseline differences between the participants and their partners using tax return data for married spouses filing jointly, as this would provide the most accurate source of data on participants' spouses. As a fallback measure, we will consider whether participants reported that they were the primary breadwinner or caregiver in the household, based on survey data. These analyses will be considered secondary.

In the unitary household model, household income should not depend on which individual in the household participates in the program. As an exploratory analysis, we can examine to what extent this appears to be the case by testing whether dual-earner households respond differently depending on the participant's sex or relative wage, following Golosov et al. (2021).

In our setting, it is possible that other household members' income or labor supply may react to the transfers. Further, we are following participants over an extended period of time, and household composition and membership could change over time, either exogenously or as a result of the transfers. We will test both of these hypotheses (i.e., that there may be individual income or employment spillovers to others in the household, particularly for romantic partners, and that household composition may change as a result of the transfers, again with special focus on romantic partners). Details are provided in the pre-analysis plan on intrahousehold outcomes. Assuming we do not observe significant income or employment spillovers, we will prefer to treat partners' income as exogenous and include it in non-labor income, along with other transfers; if we do observe differences, we will prefer the unitary household model. In either case, we will also present these elasticities by subgroup: participants who are single household heads at baseline; participants who are married or cohabiting at baseline; participants with children at baseline; and participants without children at baseline (the latter two groups overlap with the first two groups). All these subgroup analyses will be exploratory.

Other non-labor income, in either model, includes government transfers and gifts. The cash transfers were designed to have minimal impact on benefits and are not taxable. Nonetheless, if we know that government transfers were mechanically reduced for any participants as a result of the transfers, we will consider the net amount they received rather than the gross (i.e., cash transfer minus reduction in benefits). To further maximize comparability with existing studies, we will ideally use total income net of capital gains from administrative data (McClelland and Mok 2012). While the main results will use net-of-tax income, we will also present results using pre-tax income given that the differences observed in Cesarini et al. (2016) and Golosov et al. (2021) depended in large part on which metric was used.

We will also consider the minimum and maximum work hours an individual reports (G11, G12, workhrs_1 and workhrs_2). However, these outcomes will be exploratory.

We will conduct additional exploratory analyses to consider whether there is heterogeneity in the first two outcome measures (1a and 1b) by age (under 30 or 30 and over), sex (female/other or male), or education (with or without a college degree) at baseline. We will also consider heterogeneity by whether the participants had randomly received three or more mailers recruiting them to the study (as an indicator of more intense recruitment)[25] and whether the participant was recruited through the FreshEBT app, though we do not expect this to be significant given the very small subset of participants recruited through this means. It should be emphasized that these are exploratory analyses. The heterogeneity analyses by age, sex or education will be secondary, and the latter tests tertiary.

7.2 Family 2: Employment Preferences and Job Search

This family of outcomes captures whether individuals prefer more, less, or different work. The results of this family of outcomes should be interpreted carefully and in conjunction with results from other families of outcomes. For example, searching for more work is not unambiguously a positive outcome: it could indicate work effort, but it could also indicate inadequacy of one's current position.

1. Preferences for employment: a) whether a respondent is employed or if unemployed would

[25] We can also consider the incentive randomly offered, but there is not much variation in this.

prefer to be working (H1b) (1 if so, 0 if not); b) how much work they want (here we will combine responses for those who are employed, who answered "more", "same" or "fewer" (I8) with responses for those who are unemployed, who answered "full-time", "either" or "part-time" (H4); by default, if someone is unemployed and wanting any kind of work, that counts as "more" and if they are not seeking a job that counts as "same"). The responses to items a) and b) will be also considered separately descriptively for those who are unemployed, despite selection into employment; for item b), we will revert to using a categorical scale that takes the value 3 if H4="full-time", 2 if H4="either", 1 if H4="part-time" and 0 for those unemployed and not seeking employment. Selection will be addressed in this separate analysis using a Heckman correction. As a robustness check, we will also use jobsearch_2 from the mobile surveys for item b). We will additionally descriptively consider reasons why individuals left their previous jobs (leftjob_2 and newjob_3).

2. Active search: a) if employed, whether the respondent is seeking a new or additional job (G7 and G7a); if unemployed, whether the respondent is seeking a job (G2: 1=unemployed and looking for work) (*); b) how many job applications they made (jobsearch_7, G7b, or 0 if no job search); c) if they were looking for a job in the last year (jobsearch_1); d) how many interviews they had in the last three months (jobsearch_8 or 0 if no job search); e) number of actions taken to search for a job (jobsearch_6 or 0 if no job search). As a robustness check for unemployed respondents, G2 in item a) will be replaced by H3, which asks about actions taken in the last 4 weeks to look for new work; this measure might be less prone to social desirability bias, but G2 is more comparable to the question asked of individuals who are employed (G7 and G7a). Analysis of the baseline data suggests high correlation between H3 and G2. Regarding item b), jobsearch_7 asks about the last 3 months while G7b asks about the last 4 weeks; to put them on a comparable time scale, the latter will be scaled by 13/4. Since G7b would have only been asked of those employed at the time of the survey, to mitigate selection the main specification will restrict attention to jobsearch_7,

with the SRC data being pulled in for robustness checks. It should be emphasized that less active search could be as a result of having obtained a better job, so the interpretation of this outcome depends on the trajectory of employment and quality of employment families. We will present items a) and b) for those who are employed and unemployed separately as a descriptive measure outside the index, to help us interpret the main results.

7.3 Family 3: Duration of Unemployment

We may expect sustained unconditional cash transfers to result in a longer period of time spent searching for employment, potentially resulting in better matches. Thus, in this section, a lengthy period of unemployment may be a positive outcome if respondents are also searching for jobs and, ideally, ultimately able to find better employment (including self-employment). The questions asked in this family can be thought of as indicating the duration of unemployment, but do not, by themselves, determine whether the observed effect should be considered a positive or negative outcome; for that, one should consider the selectivity of job search, employment trajectory and quality of employment families.

1. Duration of unemployment: a) average duration of non-employment (*). This will be constructed using a SRC survey question (H1a) and mobile survey questions (leftjob_1* and newjob_2*) in the preferred specification, counting each spell of non-employment only once (this is important, for example, if respondents answer multiple surveys about the same period of non-employment; the longest continuous reported spell would then be used) and averaging across spells of non-employment. newjob_2* gives duration of non-employment for those who have found a new job, but if someone left a job without finding a new one then leftjob_* will indicate the time of non-employment so far;[26] b) longest period of non-employment since baseline (constructed using H1a, leftjob_1* and newjob_2*) (*). For the

[26]Since this family considers unemployment directly, we do not have to worry about selection into employment/unemployment in combining SRC and mobile survey data, as we do in sections 7.2 and 7.4.

job search index, we will restrict attention to question a), with someone with no periods of non-employment being counted as having "0" average duration of non-employment. As a robustness check, we will use only SRC survey data, but as this survey is asked relatively few times, it is more likely some individuals will not report ever being non-employed during this period. Finally, we will also consider results for item a) considering unemployment rather than non-employment (i.e. considering duration of non-employment for those who are searching for work (G2). Though this is an important outcome and will be considered as a primary item in the FDR corrections, it will not be included in the index to avoid double-counting. If an individual is not searching for work, their duration of unemployment will be considered to be 0 in this alternative measure.

7.4 Family 4: Selectivity of Job Search

This family of outcomes only applies to those who are searching for a job, so will be considered separately. The main specification will rely on mobile survey data (jobsearch*) which is available both for employed and unemployed participants; in robustness checks, we will also pull in the corresponding SRC questions (i.e. H6 and H6a, H7-H11) for those who are unemployed, as we expect a higher response rate to the SRC surveys, though there may be selection in who is unemployed. For those who are unemployed, in exploratory analysis we will also consider the SRC survey measures separately, bearing in mind there will be selection in employment status that we will address using a Heckman correction. In addition, we will descriptively consider the conditions under which an unemployed respondent would take a job (e.g. high income potential, potential for advancement, convenient location, etc.) (H5). This measure will not be included in the index, as it is exploratory. Further, as a participant's subjective expectations as to when they would find a job may be incorrect, this outcome will also be presented separately as a secondary outcome and not combined with selectivity in the index.

1. Selectivity of job search: a) if searching for a job, how long the respondent would be willing to search (H6 and H6a and jobsearch_4); b) willingness to make sacrifices for a job (e.g. take a low-paying job, take a job in a different field, take a job with a long commute, etc. We will count the total number of sacrifices the individual says they are willing to make as a negative indicator of selectivity) (H7 - H11 and jobsearch_5*); c) if searching for a job, ln income of jobs applied for, as gauged by mapping job titles of jobs applied for to the BLS Occupational Outlook Handbook (jobsearch_9*) (if multiple job titles are provided, the ln of the average income will be used); d) willing to take any job (H4a). For items combining a mobile survey jobsearch* question with SRC data, it should be noted that the SRC questions were only asked to those who were unemployed at the time of the survey. To mitigate selection issues, the main specifications will rely on the jobsearch* data, with the SRC data for those who are unemployed pulled in for robustness checks.

2. Expectations: a) if searching for a job, perceived likelihood of finding a job they would accept in 6 months (H12 and leftjob_3, counting each spell of unemployment only once; as above, the main specification will rely on the leftjob_3 variable, with the SRC data for those who are unemployed pulled in for robustness checks).

3. Reservation wage: a) the participant's reservation wage (H5b, H5c). This will not be available for everyone and hence excluded from the index.

In addition to building an index as described above, we will also ask some questions that will be analyzed through discrete choice models. First, we will model selectivity of job search. We will ask respondents about the jobs they have applied for and match the job titles they enter with job titles in the BLS Occupational Outlook Handbook, which provides characteristics of those jobs such as median pay, job outlook, and on-the-job training, all of which we will assume independently enter into participants' utility function. We will then model their choice of jobs to apply for as a discrete choice problem in which they select which jobs to apply for among those which they might realistically attain based on their previous education and work experience. (For each job title, the BLS also contains details on typical entry-level years of education required, which we have data on for study participants.)

Second, we will separately estimate a discrete choice model of job selection in order to better understand their preferences and their job search strategy. Job selection may differ from job search because job search is a function of both participants' expectations of employer preferences and participants' own preferences, while job selection more narrowly captures participant preferences (though respondents may still select jobs in part because of their expectations for future employment). For this second component, similar to Mas and Pallais (2017) and Maestas et al. (2018), we will provide respondents with repeated sets of two hypothetical jobs with different randomly assigned attributes to choose between. We asked individuals to provide responses to two sets of discrete choice questions, whose attributes and levels are listed in Table 2. Two sets of questions were used for robustness since we are interested in how respondents think about jobs with a low or high number of hours per week, as in the first experiment, but jobs with more comparable hours may be easier for individuals to compare, and the second experiment allows us to estimate finer gradations of individuals' willingness-to-pay for amenities. We expect to place more emphasis on the results of the second experiment, which also was asked more frequently.

The analysis of job offer acceptances and hypothetical choices will allow us to construct willingness-to-pay estimates, i.e., how much individuals would be willing to trade off in terms of wages for each non-wage benefit:

$$WTP = -\beta_n/\beta_w$$

where β_n is the coefficient on each non-wage benefit, in turn, and β_w is the coefficient on wages. The estimates of the discrete choice model for job applications will help us understand how transfers affect job seekers' job application strategies. We did not include attention checks in the interest

Table 4: Attributes and Levels for the Job Preferences Discrete Choice Experiment

Attribute	**Levels**
Experiment #1	
Work from home	Required; allowed, but not required; not allowed
Flexible schedule	Can set your own schedule; fixed
Number of hours per week	10; 20; 30; 40; 50; 60
Hourly wage*	$7.25; $9; $11; $13; $15; $17; $19; $21; $23; $25
Experiment #2	
Work from home	Allowed; not allowed
Flexible schedule	Can set your own schedule; fixed
Number of hours per week	30, 35, 40
Hourly wage	$14, $14.50, $15, $15.50, $16, $16.50, $17, $17.50, $18, $18.50, $19, $19.50, $20, $20.50, $21

* The minimum wage in one of the states was $7.25 per hour and the minimum wage in the other $13 per hour. For realism, respondents in the two states saw different options, according to their state, in Experiment #1. However, the amounts were otherwise designed to align as closely as possible.

of time, however, some of the options can be reasonably thought of as dominating the others. In particular, if a job has a flexible schedule, in which work from home is allowed, and it pays more (in terms of hourly wage * hours per week) while having fewer hours per week than the other option in the choice set, we will consider it to dominate the other option. For a robustness check, we will restrict attention to those participants that choose this option when it is available to them.

For both the model of job search and the model of job selection, we will use the experimental variation provided by the transfers to consider whether the cash transfers affect respondent choices.

In exploratory analysis, we will also consider whether estimated time and risk preferences appear to be associated with greater selectivity or duration of job search and whether these estimates depend on the log of baseline household income.

7.5 Family 5: Employment Trajectory

This family captures the respondents' employment trajectory. We will ask respondents to provide information on their current main job at baseline as well as their most recent job that they are not doing anymore, conditional on their having that other job within the past year (if employed) or within the past three years (if unemployed). This means we will be able to make three types of comparisons for the first five components in this family (the latter components do not depend on one's past job):

1. Comparing pre- and post-treatment trendlines. Changes in the trend lines will be exploratory only as they are based on the relatively small and likely selected set of people who worked at baseline and had another job within the past 12 months (in particular, considering variables L1a, L5, L3b, L2, L3, and L3a).

2. Investigating changes from respondents' most recent job using data on their current job if employed or most recent job if unemployed but having held a job in the last three years.

3. Investigating changes from respondents' current job (at baseline) treating those who are unemployed as having responses of 0 for all items except 1b), 1c) and 1e), 4) and 5), where it would not make sense and these items would need to be excluded.

Our preferred approach will be #3, to conserve sample size. Still, results for other approaches may be presented as exploratory, and results from #1 may be particularly interesting. Where applicable, we will also plot trajectories graphically.

1. Trajectory of main occupations: occupation will be elicited through job title (I5 and J5b; the main analyses will restrict attention to the primary job) and matched to BLS data from the Occupational Outlook Handbook. The following statistics will then be included as part of the index: a) ln of median pay (i.e. ln(follow-up median pay - baseline median pay); b) job outlook (%); c) typical entry-level years of education (12 = high school diploma or equivalent, ...); d) on-the-job training (0 = none, ...). A secondary analysis will weight a respondent's jobs by the number of hours worked at each job, however, this analysis has the limitation that factor variables would be hard to interpret. Finally, as a further robustness check, we will use measures from mobile surveys, in which every time an individual reports

getting a new job we ask its title (newjob_4), for matching with BLS data, as well as asking how it compares to their old job on a number of dimensions (newjob_17*). As before, we will focus on individuals' main job.

2. Trajectory of earnings and benefits: a) ln of total annual work-related earnings (I7, I9, I10_s_1 through I15b for the primary job, and equivalent questions for each non-primary job or 0 for those who are unemployed); b) number of benefits (count of I16__s_* and corresponding benefits for each other job), weighted by the number of hours worked at that job (I7 and J9* or 0 for those who are unemployed).

3. Trajectory of work hours: a) hours worked at the respondent's current jobs (I7 and J9* or 0 for those who are unemployed).

4. Trajectory of nature of work: a) whether the respondent was a regular employee or a temp or contract worker at their main job (I20).

5. Trajectory of stability of work: a) duration of each current job (I9, J10*), weighting responses for each job by hours worked at that job (I7 and J9*).

6. Aspirations: respondents will be asked to provide the job title they hope to have in 10 years' time (G15). This job title will be matched to BLS data and the index previously described will be re-computed for their aspirational job.

7. Subjective career prospects: a) whether the respondent thinks their career prospects have improved over time (Likert scale) (G14).

Outcomes 6 and 7, being subjective but very important, will be considered separately and not in the index.

The above measures consider the SRC survey data, but we will also consider the corresponding mobile survey-based questions where available, in an alternative specification. Further, we will separately report the extent of churn in jobs (job turnover), leveraging the frequent reporting of changes of jobs in the mobile trigger surveys (newjob*, leftjob*, incomchg*, jobintro_1_1_2, jobintro_1_2_2) in conjunction with the SRC survey data, being careful to avoid double-counting (e.g. through the mobile and SRC surveys asking about overlapping time periods).

It should be noted that a respondent's most recent job that they no longer hold may have been either a main job or a secondary job. This does not pose a problem so long as we consider all jobs that an individual might hold at a given point in time. However, we cannot straightforwardly compare measures beginning with "I" (current main jobs), "J" (current secondary jobs) and "L" (most recent job they no longer hold) without further assumptions. We will approach this challenge by two approaches: 1) conservatively comparing only like to like across surveys (e.g., "I" and "J" to "I" and "J" measures, ignoring "L" measures); 2) classifying each job into bins based on hours worked (e.g., less than 10 hours of work per week or 10 or more hours of work per week; less than 20 hours of work per week or 20 or more hours of work per week) and making comparisons only among jobs in the same bin. The conservative approach will be preferred for the main specification.[27]

Reasons for leaving past jobs will also be descriptively considered (L7, L7a, leftjob*).

We will separately consider whether there is any heterogeneity in the trajectory of employment outcomes by age and whether the participant had a college degree at baseline, but this will be exploratory.

7.6 Family 6: Quality of Employment

This section captures various factors often associated with quality of employment. Unlike the other families, this family of outcomes will focus exclusively on those who are employed and will not be combined with other questions among those not employed. Recognizing that there will

[27] In the case of conflicting reports of employment or unemployment at a given time period, we will present bounds based on the best and worst case scenario.

be selection, we will use a Heckman correction to account for this selection and present results with and without this correction to aid interpretation. Some of the outcomes will have their sign reversed prior to combination in an index; this is noted in the text.

1. Adequacy of employment: a) if the respondent is employed part-time in their main job but would prefer to work full-time (negative) (I24); b) respondent would prefer to work more hours in their current main job (negative) (I8); c) the number of jobs held by the respondent apart from their main job (negative) (G4); d) the reason provided for the number of jobs (G5). When constructing an index, the last question will be considered descriptive. We will also count the total number of W-2s and 1099s from administrative records (also asked in M3 and M5) in an alternative specification.

2. Hourly wage: a) the respondent's hourly wage (*); if the respondent works more than one job, this is their total income from employment weighted by the number of hours worked at each job, using the income and hour questions noted in the financial health pre-analysis plan.

3. Employment quality: a) number of non-wage benefits (count of I16__s_1* and corresponding benefits for each other job, weighted by hours worked at that job (I7 and J9*); b) whether training is offered by the main employer (I17); c) whether formal training is offered at the main job (I17a*=3 or 7); d) whether training is offered during work hours (I17a*=4, 5, or 6); e) whether respondent must work a night shift / irregular shift / split shift / rotating shifts / be on-call (negative) (I22 and I23 and J18*), weighted by hours worked at that job (I7 and J9*). In an alternative specification we will consider the number of non-wage benefits provided by either the main job or other jobs (count of I16__s_1* and J9*, where each benefit is included at most once in the count) and present results descriptively by type of benefit (this enables us to consider, e.g., the share of individuals that have health insurance at either their main job or another job).

4. Quality of work life: a) how hard it is to take time off (I28), with "never tried" recoded to the

mean; b) flexibility of schedule (I28a, with answers of 1 and 5 coded as 1 and answers of 2, 3, or 4 coded as they are); c) advance notice of schedule (I27); d) whether job demands interfere with family life (negative) (I29); e) whether the respondent has decision-making input in their job (I30); f) safety and health conditions (workqual_2); g) whether the work activities are boring (negative) (workqual_3); h) whether the respondent experienced fair treatment by their supervisor (workqual_4); i) whether the job is a good fit with the respondent's experience and skills (I31); j) opportunities for promotion (I32); k) overall satisfaction (I19a); l) whether the respondent plans to leave the job in the next year (negative) (I34); m) whether the respondent faces age discrimination at work (negative) (workqual_5); n) whether the respondent faces racial or ethnic discrimination at work (negative) (workqual_6); o) whether the respondent faces sex discrimination at work (negative) (workqual_7); p) satisfaction with compensation (I18); q) satisfaction with non-wage aspects of the job (I19); r) number of stressors in their work environment (negative) (workqual_8); s) whether a scheduled shift was cancelled with less than 24 hours notice in the last month (G13 and workhrs_3). These questions apply to the respondent's current main job, and all are Likert scale measures.

5. Stability of employment: a) whether the respondent is performing contract or freelance work as their main job or other job (negative) (I6 and J8*), coding "don't know" as "no" and weighting responses for each job by hours worked at that job (I7 and J9*); b) how many months they have been employed in the past year (K1*); c) how long they have spent at their current main job or other job (I9, J10*), weighting responses for each job by hours worked at that job (I7 and J9*); d) how many jobs they have held in the past 12 months (negative) (G4); e) whether their job has a salary (1) or they are paid hourly or by task, tips, etc. (0) (I10* and J11*), weighting responses for each job by hours worked at that job (I7 and J9*); f) whether their job is a temp job (negative) (I20); g) how long the respondent expects to remain in their job (I21), taking the midpoint of each range and coding "one year or more" as 18 months. Participants will also be asked how many jobs they have held in the past 2

years (K4), which may have more variation than how many jobs they have held in the past 12 months but could potentially be less accurate; it could also include a few pre-treatment values at midline. For robustness, the index will also be calculated using K4 at baseline and endline in lieu of item d).

6. Formality of employment: a) whether the respondent reports any informal jobs such as babysitting, home repairs, home cleaning, cooking and catering, sewing (negative) (M9* excepting renting property); as a robustness check, we will consider the percent of reported income not on W-2s (negative) (using administrative records for the W2s and total income from the SRC survey as described in the financial health pre-analysis plan). As an exploratory measure not included in the index, we will also note whether the respondent reports any "gig economy" jobs such as Uber, TaskRabbit, or online surveys (M9*=9, 10, 13, 14), but we expect too few individuals will participate in one of these jobs to consider it except descriptively.

7. Relative firm rating: this measure will be constructed by linking the company names of respondents' employers to data on those employers such as from Indeed.com ratings or other sources as available. The average firm rating will be considered an alternative proxy of quality of employment, conditional on there being at least 10 reviews, but since not all firms may have a rating this measure will be presented separately and not be part of the index for this family. The number of reviews and the number of employees, if reported, will also be used to descriptively discern whether individuals are more likely to be employed at larger or smaller firms. If fewer than 10% of participants work at a firm with a rating, these outcome will not be presented.

Since individuals may take on multiple jobs in part due to need, to aid in interpretation we will construct a parallel index that focuses only on respondents' main job, ignoring any secondary jobs and not weighting responses across jobs by hours worked.

It is also possible that some people may prefer freelance or informal work. Thus, in an alternative specification we will calculate the same index excluding the stability and formality of work outcomes. We will also separately consider heterogeneity by whether the participant had a college degree at baseline, but this will be exploratory.

Finally, since we ask participants to answer discrete choice questions about their preferences for different kinds of jobs (described in section 7.4), as a robustness check we will construct a measure of quality of employment that uses their own estimated weights for different levels of different attributes. It is possible that treatment status itself affects preferences; in this case, the preferred analysis will still use the individual's own preferences, but we will also consider applying the control group's aggregate weights to estimate job quality.

It should be noted that this family largely relies on survey data. If the transfers cause a change in aspirations that is not met with a change in job conditions, it is possible that they may cause dissatisfaction even while the jobs stay the same or improve in quality. Therefore, care should be taken in interpreting these outcomes. In particular, a robustness check will rely on only the administrative measures.

7.7 Family 7: Barriers to Employment

This family of outcomes is only asked to those who are employed. For the main specification, we will consider unconditional outcomes (i.e., considering a respondent to miss 0 days of work due to illness if unemployed), but in exploratory analysis we will use a Heckman correction to account for selection.

1. Barriers to employment: a) whether the respondent had to miss work due to illness (G8); b) whether the respondent had to miss work due to lack of transportation (G9); c) whether the respondent had to miss work due to lack of childcare (G10).

7.8 Family 8: Disability

Disabilities can be exacerbated by certain jobs, and having the security of cash transfers could improve individuals' health. A large literature also finds that a growing number of people are claiming disability insurance, though in absolute terms this cannot account for much unemployment (Autor and Duggan 2003; Staubli 2011). For these reasons, disability is an important outcome to capture, however, we may not expect much change in disabled status over the duration of the study, so this set of outcomes will be reported separately from other barriers to employment, which one may expect to be more responsive to the treatment. The primary outcome measure considered in this family is whether the respondent does not work due to disability, and beyond our interest in this measure for its own sake, we expect this measure to be useful as a control variable in other regressions.

There is some risk that individuals report more disability even if the transfers do not have an impact on disability. For example, this could be the case if the transfers resulted in their receiving more medical care and diagnoses, or if they feel the need to justify potentially reduced labor hours with disability. Care must therefore be taken in interpreting this outcome family.

1. Disability: a) Washington Group questions on whether the respondent has a disability (negative) (items dis_1_1 - dis_1_6); b) whether the respondent has a health problem or disability that limits the kind or amount of work they can do (negative) (dis_1_7); c) how much their worst disability or health problem limits the kind or amount of work they can do (negative) (dis_2); d) how long this health problem or disability has affected the kind or amount of work they can do (negative) (dis_3).

7.9 Family 9: Entrepreneurship

In this section, we will gather data not just on entrepreneurial activities but on "precursors" to entrepreneurial activity like entrepreneurial intention and entrepreneurial orientation. We expect that more people will exhibit changes in entrepreneurial orientation or entrepreneurial intention than entrepreneurial activity, both because some barriers to entrepreneurial activity may remain even with the transfers and because some intentions may not translate to activity even in the absence of any barriers.

1. Entrepreneurial activity: a) whether the respondent ever started or helped to start a business (entrep_1); b) whether the respondent knows anyone who ever started or helped to start a business (entrep_2); c) whether a family member who started a business lives in the respondent's household (entrep_3).

2. Entrepreneurial intention: a) on a 1 to 10 scale, the respondent's interest in starting a business (entrep_4); b) on a 1 to 10 scale, how likely they think they would be to start a business in the next 5 years (entrep_6); c) whether or not the respondent has an idea for a business (entrep_5).

3. Entrepreneurial orientation: a) the respondent's risk preferences, as gauged by respondents' choice of one of six lotteries based on the flip of a fair coin (Z1), taking the midpoint of each range of consistent risk preferences; b) the respondent's self-reported willingness to take financial risks, on a 0 to 10 scale (gpref_2). The latter measure follows Falk et al. (2015), while the first follows Eckel and Grossman (2002).

We will consider heterogeneity by age and whether the participant had at least some college education at baseline, but this will be exploratory.

7.10 Family 10: Human Capital Formation

This section focuses on variables relating to formal and informal education. Receiving cash transfers may provide individuals with the security to pursue more education and cause individuals to work less in the short run in pursuing education. This set of outcomes can thus help interpret some of the changes we may see in other outcome families and also provide guidance as to whether individuals may be gaining new skills that could help them in the labor market.

1. Formal education: a) enrollment in post-secondary (associate, certificate, etc.) programs (*); b) progress in program (total years of education completed); c) completion of GED or post-secondary programs; d) enrollment intensity; e) area of concentration, classified into broad areas.[28] Our preferred specification will use administrative data, and the last two measures will be descriptive. Apart from enrollment in any post-secondary program, we will also separately consider enrollment by credential level in a secondary, exploratory analysis. We will also consider whether respondents attend a private or public school and the type of institution they attend according to the Carnegie Classification, though this will also be considered secondary. Finally, using the NCES CIP code to BLS SOC code correspondence, we will examine whether individuals appear to be receiving training for different jobs than the ones they held at baseline. This will also be considered an exploratory analysis and not included in the index.

2. Informal education: a) participation in any other education, including on-the-job training, noncredit study, private study, online courses, or training offered by other groups such as employment centers (B19a = 6, 7, school_4 = 6, 7, 8, 9, 10, 11, else 0 if no such participation); b) extent of participation in these activities (part-time = 1, full-time = 2: B19, school_2, else 0 if no such participation). We also ask for hours of schooling for those pursuing it part-time but as few do this outcome will be exploratory only. Descriptively, we also ask why individuals are pursuing education (school_5). These measures will be self-reported through surveys and any such participation over the entire study period counted.

The surveys also ask about formal education, and we will compare the survey responses with the

[28]We will largely follow Sjoquist and Winters (2015) in considering STEM, business, education, health, arts and humanities, and the social sciences, with an additional focus on vocational programs (including agriculture, natural resources, and trades and services).

administrative records. In particular, we will compare 1a, 1c and 1e with survey-based measures (B19a or school_4 = 2, 3, 4, 5; new completion of any items in B20 of 3 or higher; and B20a and B20b, respectively). If we observe differences, it may be due to bias in self-reporting, which would help us gauge how much we should trust the self-reported measures of informal education. However, informal education could also be less likely to be remembered, or alternatively could be exaggerated. As relatively few participants may indicate they are pursuing informal education during the study, these outcomes will be considered exploratory.

To form the index, we will consider whether the respondent reported receiving any education or job training (a binary variable that takes the value 1 if B18 = 1 or G6 = 1 at midline or endline or if administrative records show education) and descriptively report whether or not individuals plan to start a job training program (G6 = 2).

Finally, we will descriptively consider hours spent on school, but this will be exploratory and not in the index (school_3 if part-time according to school_2 and 40 hours a week if full-time).

We will consider heterogeneity by age, but this will be exploratory.

8 Employment Outcomes: Administrative Data Measures

In addition to the survey-elicited measures, we will also leverage administrative data, where possible. In particular, we are working to find a collaborator within the federal government to match individuals in our sample to tax return data and report back impacts on their records of aggregate income, including income as reported in W-2s, 1099s, passive income, and federal transfers. We are also seeking to use state UI data, which have the advantage that they are available quarterly. How these data will be used to capture income is discussed in a separate pre-analysis plan. We are additionally planning to use National Student Clearinghouse data for the education outcomes. If there are substantial lags in obtaining any of these data, items relying on them may be broken into a subsequent paper.

The administrative data play a direct role in outcome "a" under "Formality of employment" in the "Nature of employment" family, as described in the survey-based outcomes section. If the participant has W-2 income, this can also serve as an alternative measure of outcome "a" under "Stability of employment" in the same family. Both measures will be considered in alternative specifications, for robustness; however, the administrative data will be used in our preferred specification. Likewise, administrative data will be the preferred data source for the "Formal education" outcomes, while the survey data will provide robustness checks.

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.

Autor, D and M Duggan. 2003. "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics* 118:157–206.

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.

Bastagli, F, J Hagen-Zanker, L Harman, G Sturge, V Barca, T Schmidt and L Pellerano. 2016. "Cash transfers: what does the evidence say? A rigorous review of impacts and the role of design and implementation features." ODI Report.

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

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.

Caliendo, Marco, Konstantinos Tatsiramos and Arne Uhlendorff. 2012. "Benefit Duration, Unemployment Duration and Job Match Quality: A Regression-Discontinuity Approach." *Journal of Applied Econometrics* 28(4).

Card, David, Raj Chetty and Andrea Weber. 2007. "The Spike at Benefit Exhaustion: Leaving the Unemployment System or Starting a New Job?" *American Economic Review* 97(2).

Centeno, Mario. 2004. "The Match Quality Gains from Unemployment Insurance." *Journal of Human Resources* 39(3).

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, 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).

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.

Eckel, Catherine and Philip Grossman. 2002. "Sex Differences and Statistical Stereotyping in Attitudes toward Financial Risk." *Evolution and Human Behavior* 23(4):281–295.

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, 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.

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

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman and Uwe Sunde. 2015. "The Nature and Predictive Power of Preferences: Global Evidence." IZA Discussion Paper 9504.

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.

Golosov, Mikhail, Michael Graber, Magne Mogstad and David Novgorodsky. 2021. "How Americans Respond to Idiosyncratic and Exogenous Changes in Household Wealth and Unearned Income." *NBER Working Paper* 29000.

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.

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.

Kabeer, N and H Waddington. 2015. "Economic impacts of conditional cash transfer programmes: a systematic review and meta-analysis." *Journal of Development Effectiveness* 7:290–303.

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: .

Keane, M. 2011. "Labor Supply and Taxes: A Survey." *Journal of Economic Literature* 49:961–1075.

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.

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.

Lalive, Rafael. 2007. "Unemployment Benefits, Unemployment Duration, and Post-Unemployment Jobs: A Regression Discontinuity Approach." *American Economic Review* 97(2).

Lundberg, Shelly and Robert Pollak. 1993. "Separate Spheres Bargaining and the Marriage Market." *Journal of Political Economy* 101(6):988–1010.

Maestas, Nicole, Kathleen Mullen, David Powell, Till von Wachter and Jeffrey Wenger. 2018. "The value of working conditions in the United States and implications for the structure of wages (Working Paper No. w25204). Cambridge: ." *National Bureau of Economic Research Working Paper* w25204.

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

Mas, Alexandre and Amanda Pallais. 2017. "Valuing alternative work arrangements." *American Economic Review* 107(12):3722–59.

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.

McClelland, R and S Mok. 2012. "A review of recent research on labor supply elasticities." CBO Working Paper Series.

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.

Nekoei, Arash and Andrea Weber. 2017. "Does Extending Unemployment Benefits Improve Job Quality?" *American Economic Review* 107(2).

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.

Price, D and J Song. 2017. "The Long-Term Effects of Cash Assistance." Working paper.

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.

Rothstein, Jesse. 2010. "Is the EITC as Good as an NIT? Conditional Cash Transfers and Tax Incidence." *American Economic Journal: Economic Policy* 2(1):177–208.

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.

Sjoquist, David L. and John V. Winters. 2015. "State Merit Aid Programs and College Major: A Focus on STEM." *Journal of Labor Economics* 33(4).

Staubli, S. 2011. "The impact of stricter criteria for disability insurance on labor force participation." *Journal of Public Economics* 95:1223–1235.

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.

van Ours, Jan C. and Milan Vodopivec. 2008. "Does reducing unemployment insurance generosity reduce job match quality?" *Journal of Public Economics* 92(3-4).

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.