

The COVID cash transfer study II: The hardship and mental health impacts of an unconditional cash transfer to low-income individuals *

Brian Jacob
Natasha Pilkauskas
Elizabeth Rhodes
Katherine Richard
H. Luke Shaefer

April 25, 2022

Abstract

This paper reports findings from a randomized controlled trial of a one-time, \$1,000 unconditional cash transfers to low-income households in October 2020. We use a combination of administrative and survey data collected six weeks post-treatment to examine four pre-registered hypotheses: impacts on material hardship and mental health in the full study sample as well as among a very low-income sample. We find no effects of the cash transfer on any of the pre-specified outcomes, or other exploratory outcomes. We explore various explanations for these null results and discuss implications for future research on unconditional cash transfer programs.

*Special thanks to all partner organizations who made this study possible, including GiveDirectly, Propel, Y Combinator Research, Stand for Children and the Schusterman Family Philanthropies. Particular thanks to Karen Ann Kling, Alex Nawar, Michael Cooke, Farheen Rizvi, Miriam Laker, Jeff Kaiser, Cherie Chung, Jeremy Guardiola, Marty Moore, Kaitlin Raimi, and Jonah Edelman. Greg Duncan, Jonathan Morduch, and Johannes Haushofer provided excellent feedback on design and implementation.

1 Introduction

While the U.S. social safety net has expanded in recent decades, access to cash assistance for low-income families has declined sharply since the 1990s (Parolin (2021); Shaefer et al. (2020)). Unconditional cash assistance has garnered increasing interest as a means to alleviate hardship and poverty in recent years (Genetian et al. (2021)). Despite this growing attention, many questions remain regarding how to best design such transfers and what impacts they might have in the U.S. context. Most evaluations of unconditional cash transfers have been conducted outside the U.S., provide recurring cash payments, or involve extremely large transfer amounts relative to average annual income, making such programs difficult to replicate in the U.S. (Bastagli et al. (2016); Davis and Handa (2015)).

This paper reports on the second of two experimental studies of unconditional cash transfers to low-income households in the U.S. In response to the economic hardship caused by COVID-19 pandemic, in March 2020 the charitable organization GiveDirectly (GD) began providing low-income households an unconditional cash transfer of \$1,000 in an effort to mitigate their financial stress. As a one-time, lump-sum transfer, the payment did not impact recipients' eligibility for other public assistance benefits.

In order to shed light on the impacts of cash transfers, we partnered with GiveDirectly to conduct two randomized control trials of their program in 2020. Pilkauskas et al. (2022) describes the findings of the first experiment (RCT-1), which was conducted in May 2020 and focused on a set of low-income families with children receiving federal food assistance (SNAP benefits) across 12 states. Although the cash transfer did not have a statistically significant impact in the full sample, the intervention reduced material hardship among the most economically disadvantaged households in the study and found suggested improvements in mental health. Based on the results of this first study, we designed and carried out a second experiment (RCT-2).¹

As in the first study, the treatment consisted of a one-time, \$1,000 cash payment offered to a set of families receiving SNAP in September 2020. However, we refined the study design to include a broader set of material hardship and mental health measures, and an expanded sample of lower-income families: including families both with and without children, focusing on respondents living in 1,265 high-poverty zip codes, and expanding the sample to include 49 states, D.C. and Puerto Rico.

In contrast to expectations, we find no discernible impacts on any of the outcomes we measure for either the full sample or any of the pre-specified subgroups. As we describe below, the treatment and control groups are balanced on a large set of baseline covariates. The point estimates for our primary material hardship outcomes are very small (around 0.01-0.02 SD) and we have the statistical power to rule out effects as small as 0.07 SD for the full sample and 0.10 SD for the extremely low-income sample.

This paper describes the results of this experiment in detail, and examines a variety of potential explanations for the null results. We conclude by exploring reasons for the differing results across experiments, including differences in sample construction, recruitment tactics, timing of data collection, and underlying material hardship sensitivity. The results across both experiments suggest that low-income families in the U.S. face a high degree of financial complexity that is difficult to capture in standard survey measures, and likely made more so during a global pandemic with an unprecedented policy response. More research concerning the effects of cash is needed, especially work that considers deeply what outcomes are most appropriate for such studies, and that vary the timing and size of transfer payments.

¹This study was pre-registered with the [AEA Pre-Trial Registry](#).

2 How Might a \$1,000 Unconditional Cash Transfer Affect Material Hardship of Low-Income Households in the U.S.?

Unconditional cash transfer programs are rare in the United States so there is little prior research to directly inform our expectations for the GiveDirectly transfers.² In the international context, evidence suggests that unconditional cash transfer programs can reduce poverty, increase food expenditure and improve mental health (Bastagli et al. (2016), Oosterbeek et al. (2008)). However, these transfers are often much larger than the \$1,000 transfer provided in our study.

Quasi-experimental studies of *recurring* cash transfers in the U.S. find they are associated with positive outcomes. For example, studies of casino dividends disbursed among U.S. tribes show that receipt of the cash payments is correlated with improvements in health outcomes, educational attainment, and parent-child interactions and reductions in risky behaviors in adolescence (Akee et al. (2010)).³ There are several ongoing experimental studies of recurring cash payments to low-income families, including OpenResearch’s basic income experiment, Baby’s First Years, Compton Pledge and Chelsea Eats. The first published findings from this research indicates that at age one children in low-income families who received \$1,000 cash assistance each month exhibited greater brain activity in areas correlated with language, cognition and socio-emotional wellbeing than children in the control group (Troller-Renfree et al. (2022)). However, the recurring nature of these payments is quite different than the one-time transfer provided in our context.

The best evidence on unconditional *lump sum* cash transfers in the U.S. comes from quasi-experimental studies of the Earned Income Tax Credit (EITC). The EITC literature suggests that a \$1,000 increase is associated with modest benefits in some areas, but no effects in other areas. For example, Batra et al. (2022) find \$1,000 (approximately 4% of annual income) is associated with a very small decline (less than 1% of a standard deviation) in short-term food insecurity. Another study found a \$1,000 increase in the EITC had no effect on extreme housing hardships like homelessness, but was associated with a 15% decline (2 percentage points) in moving in with others (Pilkauskas and Michelmore (2019)). The EITC has been linked with few short-term effects on self-reported health and mental health (Collin et al. (2021)), but positive longer-term effects on mental health (Evans and Garthwaite (2014)).

Other evidence comes from the Women’s Employment Study, which followed a set of current and former welfare recipients from 1997-2003 (Sullivan et al. (2008)). When they consider a measure of average income over the entire study period, the authors find that a 10% increase in income (roughly \$1,900) decreased material hardship by 1.1 percentage points or 3.4%. Holding constant average income, however, they find little evidence that material hardship is associated with transitory changes in income. The authors suggest that this weak relationship may be due to the fact that families use informal, typically unmeasured resources to buffer income shocks.

²One experimental study of a one-time lump-sum unconditional cash transfer to low-income families with children in the United States is currently in progress (Hauser et al. 2020). Lottery studies are also unanticipated cash transfers; however, most of the research in this area is focused on employment outcomes and very large cash transfers due to data limitations (e.g., Imbens et al. (2001); Golosov et al. (2021)).

³The Canadian child benefit – a monthly cash payment to families with children – is linked with a variety of positive outcomes (e.g., Milligan and Stabile (2011)). On the other hand, Jacob et al. (2015) find little effect of cash on child outcomes among recipients of Section 8 housing assistance in Chicago.

Taking this evidence as a whole, we anticipated that a one-time \$1,000 unconditional cash transfer would modestly reduce but not eliminate material hardship. Based on some of the prior literature and our earlier study (Pilkauskas et al. (2022)), we expected larger effects for the lowest income households, which motivated our decision to look separately at families earning less than \$500 per month. For this group, the study transfer represented roughly 75% of annual income, while for the study households reporting more than \$500 in monthly earnings, the transfer was closer to 5% of annual income.

3 The Intervention

As part of its charitable mission and in response to COVID-19, in March 2020 the nonprofit organization GiveDirectly started providing a set of low-income families in the U.S. a one-time, unconditional, lump-sum cash transfer of \$1,000 to help mitigate the financial stress associated with the pandemic. GiveDirectly identified and recruited recipients through a free mobile application (app) called Fresh EBT (now called Providers) that helps users manage their public assistance benefits. The payments were completely unanticipated and families could use the money as they chose. As a one-time, lump sum gift to families, this aid did not impact other public assistance benefits.⁴

The study described in this paper examines the impact of one set of cash disbursements in September 2020. At the time of this study, roughly 3-4 million individuals were actively using the Fresh EBT app, most of whom were receiving federal food assistance through the Supplemental Nutrition Assistance Program (SNAP) program.⁵ Individuals were not required to participate in the research project, or fulfill any other obligations, in order to receive the cash payment.

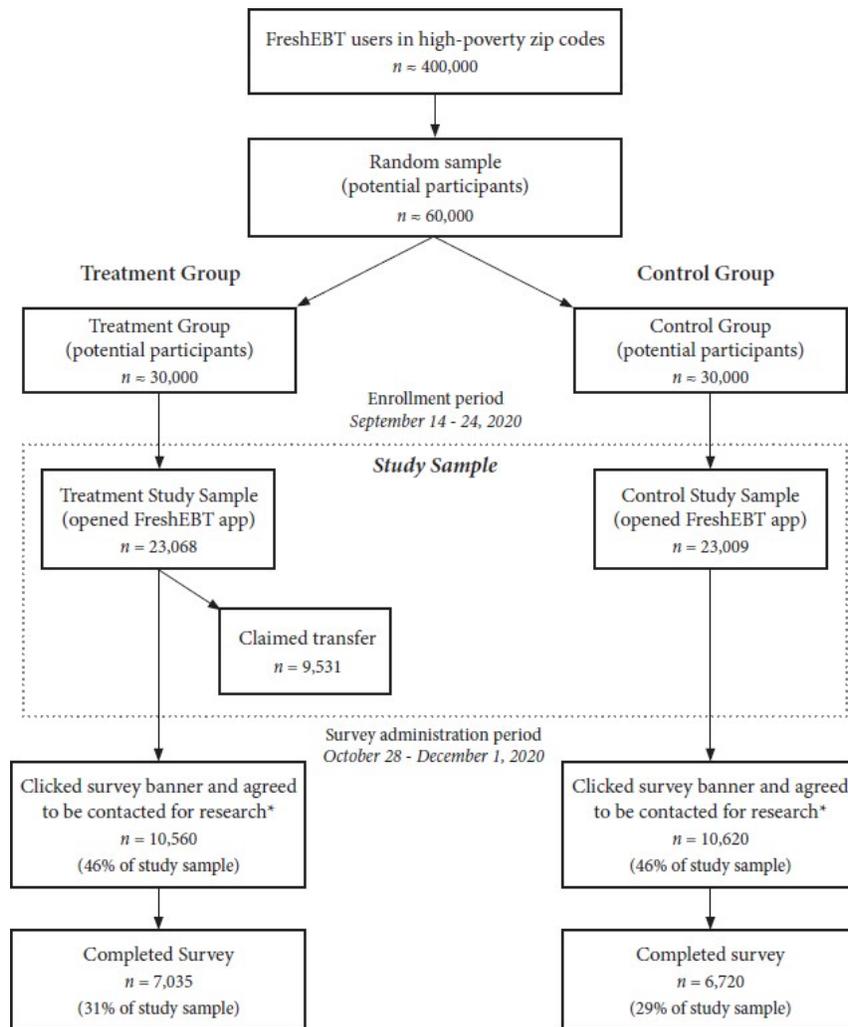
3.1 Randomization

The sample for the present study was drawn from a set of families receiving SNAP and using the Fresh EBT application in summer 2020. In an effort to target its fall 2020 campaign to the most disadvantaged individuals, GiveDirectly's fall intervention focused on zip codes with poverty rates of at least 35% (as measured in the 2018 American Community Survey, five-year moving average estimates).

Among the roughly 400,000 Fresh EBT users in high-poverty zip codes, we randomly assigned 60,000 users in equal proportion to treatment and control groups, stratifying on six groups defined by whether or not the individual began using the app prior to February 1, 2020 and a three-category measure of the generosity of the social welfare benefits provided by the state in which the individual lives (as measured by the maximum TANF benefits for a three-person household in the state). We used the re-randomization method outlined in Lock and Rubin (2012) to ensure covariate balance for all variables available prior to treatment assignment. Table A.1 lists all of the variables used in the re-randomization process. To test for statistically significant imbalance on covariates, we re-randomize treatment assignment 1,000 times and compute the proportion of T statistics that are larger than our observed T statistic for each bivariate regression. The resulting proportions, or permutation p-values, are listed in column 4 of A.1 and 2. We are able to reject imbalance on the majority of characteristics, with no single permutation p value greater than 10 percent.

⁴Families received the transfer through a payout application called Hyperwallet. GiveDirectly continued to provide unconditional cash aid to needy families in the U.S. through October 2021. See [GiveDirectly Project 100](#) for more information about the program.

⁵At this size, Fresh EBT serves a meaningfully large fraction of the SNAP recipient population which included 22.2 million households as of April 2020.



* Roughly 3,000 people in each study sample group never opened the app during the survey period and were never exposed to the survey banner.

Figure 1: RCT2 Enrollment and Survey Administration

3.2 Implementation of the Intervention

GiveDirectly invited the first cohort of treatment group members to claim their \$1,000 cash transfer on September 14, 2020.⁶ Treatment users who logged into the app were shown a banner within the user interface with the headline “\$1000 CASH GIFT from GiveDirectly”, followed by “Claim your **\$1000 cash gift** from the non-profit **GiveDirectly** for families in financial need. Enrollment will end soon, and there is limited space available. Tap here to sign up for \$1000.” On September 18, Fresh EBT sent push notifications to increase take up among individuals who started the enrollment process but had not completed and to notify users who had not logged into the app when the banner was visible. Enrollment closed on September 24, although individuals who had started the enrollment process before it closed were given until September 29 to finish the process.

Of the 60,000 individuals initially assigned to treatment or control conditions, roughly 46,000 logged onto the app at least once during the intervention enrollment period (September 14-24), and thus had the opportunity to participate in the study (either as treatment or control). These potential viewers make up the study sample. Among the 23,068 treatment users who logged on during this window, 56% (n=12,845) clicked on the GiveDirectly banner to read about the opportunity. Among that group, 74% (n=9,531) eventually claimed the award. 2,071 did not complete the enrollment process because they did not agree to the terms (n=446) or failed to finish the necessary steps (n=1,625). 1,243 individuals finished the enrollment process but never took the subsequent steps necessary to set up their HyperWallet accounts. Hence, only 41.3% of the 23,068 users assigned to the treatment group ended up receiving the cash award. Figure 1 outlines the flow of the study, from initial assignment through survey completion.

4 Data and Outcomes

We used both administrative and survey data in the study. From the Fresh EBT app, we obtain information on the individual’s state, zip code, year of birth, preferred language (e.g., English or Spanish), date started using the app, average amount of time spent on the app and most recent SNAP benefit amount. This section describes our survey and the key outcome measures we utilized in the analysis. For a more detailed discussion of the survey administration, see Appendix 1. The survey instrument is available in Appendix 2. Appendix 3 provides more details on the cleaning of survey data and creation of variables used in the study.

4.1 Survey

On October 28, 2020, roughly 6 weeks after the treatment group was first able to claim their transfers, Fresh EBT placed a notification in the app inviting treatment and control group members who had logged onto the app during the enrollment window to participate in a 10-minute survey about their experiences during the COVID-19 pandemic in exchange for a \$10 gift card. The banner contained no mention of the GiveDirectly cash transfer program. Users who clicked on the banner were directed to the consent form and then the survey.⁷

⁶To avoid over-enrolling individuals in the transfer program, Fresh EBT initially showed the banner to a limited number of treatment group members. The first cohort included roughly 13,000 members of the treatment group. The second and third cohorts (containing approximately 5,000 individuals each) were added on September 22, and the final cohort of roughly 7,000 was added five hours before enrollment closed on September 24

⁷The first question requested contact information so we could follow up if respondents consented to participate but did not finish the survey. The referral link from the Fresh EBT app included the user’s Fresh EBT ID, which enabled us to merge the respondent’s treatment status and other administrative data provided by Propel and GD with survey responses.

The survey remained open until December 1, 2020. The notification was no longer displayed in the app once an individual started the survey, but it appeared on the home screen of the app for all nonresponders in the sample throughout the survey administration period. We sent regular email reminders and eventually increased completion incentives to \$20 in order to encourage responses. In total, 13,755 individuals responded to the survey, including 7,035 individuals from the treatment group (representing 31% of the treatment group study sample) and 6,720 individuals from the control group (representing 29% of the control group study sample). On average, individuals in our treatment group responded to the survey 39 days after claiming their cash transfer.

4.2 Material Hardship and Economic Wellbeing Outcome Measures

Our primary measure of material hardship is a composite created from sixteen binary (yes or no response) questions that ask individuals whether they have experienced various types of hardship, ranging from not being able to pay the full amount of various bills, to whether they worry about running out of food or have skipped medical care because of the cost (following prior research with the Survey of Income and Program Participation [SIPP] and Fragile Families and Child Wellbeing Study [FFCWS]).

We follow the approach outlined in [Anderson \(2008\)](#) to create composite measures from these individual survey items. First, we code each item so that an affirmative response corresponds to increased incidence of hardship; next, we standardize each individual item within the domain, using the mean and standard deviation of that outcome for the control group. We then create a weighted average of these standardized outcomes, weighted by the inverted covariance matrix of the transformed outcomes in the domain. Instead of replacing with zero, we replace any missing responses with the average response to items within the domain; only respondents who answer at least two-thirds of all items within a domain are included in the composite index. Finally, we standardized the weighted average of items using the mean and standard deviation of the control group so that each index can be interpreted in control standard deviation units.

To explore the effects of the cash transfer on economic well-being more generally, we examine a series of secondary measures. First, we use the USDA’s six-item food insecurity module, to construct a three-value food insecurity scale and examine whether treatment influences the likelihood of a family experiencing low or very low food insecurity. Next, we examine treatment effects on levels of debt related to basic necessities, using a composite measure of total debt amounts (back owed) for rent/mortgage, gas/electric and phone/internet bills. We refer to this composite as “core” debt. We also explore effects on a broader composite measure of debt that further includes reported money owed on other formal loans (e.g., car payments), informal loans (e.g., from friends or family), medical debt, and back owed debt on credit cards. Our goal with these measures is to shed light on survey respondents’ household liquidity, and the degree to which a cash transfer is used to pay down debt balances. Our last exploratory outcome relating to economic well-being is a composite measure of hardship avoidance - that is, behaviors employed by families to increase economic security. This measure includes information about whether and how often families used food bank resources, sold household items to make ends meet (and how much money was earned through these sales), as well as whether households cut back their spending to make ends meet.

4.3 Mental Health Outcome Measures

To measure mental health challenges, we use well validated instruments, the Generalized Anxiety Disorder two-item scale (GAD-2) for anxiety ([Kroenke et al. \(2003\)](#); [Spitzer et al. \(2006\)](#); [Plummer et al. \(2016\)](#)) and the Patient Health Questionnaire two-item (PHQ-2) for depression ([Kroenke et al. \(2001\)](#); [Kroenke et al. \(2003\)](#)). We create a single composite mental health measure that combines the anxiety and depression

items following the approach described above.

To further explore effects of our cash transfer on mental health challenges, we analyze two additional measures: the Stress Overload Scale Short (S-SOS, [Amirkhan \(2018\)](#)) and the Somatic Symptom Scale (SSS-8, [Gierk et al. \(2015\)](#)). The SOS is a ten-item scale motivated by psychological theory that stress becomes destructive when individuals are exposed to demanding events without adequate resources to meet those demands. The SSS consists of eight items used to measure physical manifestations of psychological stress. We construct composite indices for both SOS and SSS following the method used for our primary outcomes.

4.4 Other Exploratory Outcomes

We also examine treatment effects on a few exploratory outcomes: child behavior problems, parenting problems and intimate partner conflict. Effects on child behavior problems are measured using an eleven item instrument concerning child behavior over the prior two weeks (adapted from the child behavior checklist; Achenbach 2001). We ask about four externalizing behavior items (fight, argue, destructive, disobedient), four internalizing behavior items (concentration, confusion, sadness/depression, quiet/withdrawn), and two measures of prosocial behavior (nice to others, helpful). We also included one item to assess trouble sleeping. Parenting practices are evaluated using a four-item scale that asks about parenting behavior over the prior two weeks in terms of negative behaviors (or harsh parenting, yelling/hitting or spanking) and positive behaviors (time assisting with school work, time relaxing/playing/etc; derived from the FFCWS). Finally, intimate partner conflict is assessed by a six-item instrument asking how frequently a respondent's partner or spouse encourages/helps, listens when needed, insults or criticizes, withholds contact with friends or family, is physically violent, or is sexually coercive (derived from the Conflict Tactics Scale).

4.5 Other Behavioral Responses to Transfer

Finally, we examine potential behavioral responses to the cash transfer. Thus, we study whether the transfer had any effect on public assistance use (Unemployment Insurance, Temporary Assistance to Needy Families (TANF), Supplemental Nutrition Assistance Program (SNAP), Social Security (SS) and Disability Assistance programs (SSI/SSDI/other disability)) and total dollars from public assistance. We also explore whether the transfer had any effects on labor market participation (employed at least part-time, number of hours worked in the last week, monthly earnings, whether someone looked for work if they were unemployed, and number of search tasks in the last 30 days).

In order to measure changes in consumption, we asked respondents to report spending on items in 18 different categories that were based on questions asked in the Survey of Household Economics and Decision-making (SHED) as well as prior research. These categories include rent, utilities, transportation, childcare, food and education. See full survey in Appendix 2 for more details. Unfortunately, the low quality of the responses we obtained preclude us from saying much about the impact of the intervention on consumption. For example, nearly 20% of respondents indicated spending nothing at all in the prior month, and 30 percent of respondents indicate that they spent money in three or fewer categories in the prior month, which seems implausible given the 18 potential categories, many of which involve expenditures that are common even for low-income households. See Appendix 4 for more details.

5 Empirical Analysis

By randomizing treatment status, we remove issues of endogenous selection into treatment. This allows us to estimate treatment effects as a simple difference of means between treatment and control. To increase statistical power, we control for respondent characteristics captured in administrative data prior to the study as well as some time-invariant background characteristics collected in the survey (see Table A.1). To obtain the Intent-to-Treat (ITT) effect of the cash transfer, we regress each of our primary outcomes on a treatment indicator and our set of controls where i denotes individuals and g denotes group (treatment or control) and c denotes zip codes.

$$Y_{i,g,c} = \theta_{ITT} Z_{i,g,c} + X_{i,g,c} + \varepsilon_{i,g,c} \quad (1)$$

Recall that individuals in the randomly assigned treatment group were shown a banner on the Fresh EBT app during the enrollment window. Only a subset of these individuals clicked on the banner to learn more, and only a subset of those followed through with the necessary steps (e.g., completion of waiver form) to receive the cash aid.⁸ For this reason, we also estimate the Local Average Treatment Effect (LATE) of receiving a \$1,000 cash transfer, which we obtain using the two-stage-least-squares framework below. In the first stage, we regress an indicator D for whether the respondent received the \$1,000 cash payment on the randomly assigned measure of treatment (Z).

$$D_{i,g,c} = \pi Z_{i,g,c} + X_{i,g,c} + \varepsilon_{i,g,c} \quad (2)$$

The second stage will be:

$$Y_{i,g,c} = \theta_{LATE} D_{i,g,c} + X_{i,g,c} + \varepsilon_{i,g,c} \quad (3)$$

6 Statistical Inference

Because we used the re-randomization method outlined in Lock and Rubin (2012) to assign individuals to treatment, we use a randomization inference approach when we estimate treatment effects. Using the analysis sample of 13,755 respondents who completed the survey, we randomly re-assign each individual's treatment status using the same stratification as the original assignment. We then calculate the covariate balance between our pseudo treatment and control groups. If the balance meets the criteria we used in the original randomization, we estimate a pseudo treatment effect using the equations shown above. This will yield a pseudo point estimate. We conduct 1,000 such re-randomizations and obtain up to 1,000 pseudo point estimates. By placing our actual treatment effect point estimates in the distribution of pseudo point estimates, we calculate empirical p-values for the estimates.⁹

7 Results

We begin by examining the effects of the cash transfers on our two primary outcome measures (material hardship and mental health challenges) in Table 3. The top panel shows estimates for the full sample while the bottom panel shows estimates for what we describe as the “very low-income sample” - respondents who report monthly income of less than \$500, which is approximately the bottom half of the reported income distribution in our sample. The standard errors shown in parentheses adjust for clustering at the zipcode

⁸As noted above, roughly 41% of the group exposed to the treatment banner successfully claimed their transfer. Among the set of survey respondents included in the analysis sample, 68% of the treatment group successfully claimed their transfers (N=4,825).

⁹Note, we do not adjust for multiple hypothesis testing as we find no significant results.

level. The p-values shown in square brackets are based on the randomization inference method described above.

All of the estimates in both the full and very low-income sample are small and not statistically different from zero, suggesting the cash transfer did not have a discernible impact on either our measures of material hardship or mental health challenges. Note that the lack of significance is not due to a lack of statistical power. Given our large sample size and covariates, even with the relatively low treatment take-up, the minimum detectable effect sizes of the LATEs for the full (very low-income) samples were roughly 0.06 (.085) SD for both the material hardship and mental health outcomes. Appendix Table A.2 presents LATE estimates for each of the 15 binary outcomes that go into the hardship index. The results do not reveal impacts on any of the individual items. Given the null results on our primary outcome domains, it is unsurprising that we find no significant effects on several secondary outcome domains: parenting problems, child behavior problems or partner conflict (see Appendix Table A.3).¹⁰

Of course, it is possible that the cash transfer influenced the households' economic or psychological well-being in ways that are not captured by our main indices. Table 4 shows results for several related outcome measures described above: core debt (which includes rent/mortgage, gas/electric and phone/internet), total debt (which includes core debt plus loan debt, medical debt and back credit card debt), hardship avoidance and food insecurity. We find no significant effects on any of these measures for either the full or very low-income sample. Table 5 presents estimates for several secondary health outcomes, including a binary indicator for anxiety, a binary indicator for depression, a composite measure of stress (the S-SOS scale) and a composite measure of somatic symptoms associated with stress and/or mental health problems (the SSS-8). Again, we find no significant effects.

In summary, it appears that the \$1,000 cash transfer did not have a statistically significant or substantively important influence on any of a wide range of economic, social and psychological outcomes. The null effects for the very low-income sample are puzzling given the results of our earlier results in RCT-1 (Pilkas et al. (2022)). In the next sections, we explore potential explanations for these null results, including changes to labor supply, benefit utilization and savings. We then turn to a variety of robustness checks to assess if the null findings persist when we define the very low-income sample differently and consider differential response rates.

7.1 Behavioral Responses to Cash

While the cash transfer was not associated with significant reductions in self-reported material hardship, amounts of debt or mental health measures, it may be the case that transfer receipt altered behavior in other dimensions. For example, perhaps treatment families responded to the cash transfer by seeking fewer forms of public assistance or working less. However, Appendix Table A.4 shows the treatment had no impact on the likelihood of reporting benefit receipt or the total dollar amount of benefits received. Moreover, Appendix Table A.5 shows the treatment had no impact on employment, hours worked or monthly earnings. Interestingly, we find that receipt of the cash transfer is associated with a 6 percentage point increase in the likelihood of searching for a job, more than a 10% increase from the control mean of 48 percent.

It is possible that generous federal support provided during COVID (see above) enabled families in the study to save their cash transfer. If they chose not to spend the extra money, we would not be surprised to find little effect on material hardship. While we asked respondents about amounts spent across a variety of

¹⁰The minimum detectable effect LATEs for these outcomes are as small as 0.08 (0.11) SD for the full (very low-income) sample.

categories, the resulting data suffered from several data quality issues, which precluded us from drawing quantitative conclusions about the effects of transfer receipt on spending. However, we also asked treatment respondents how they spent their transfers by selecting from a list of categories.¹¹ Table 1 reports the proportion of respondents that indicated using their transfer on each type of spending. Responses indicate that only 5.5 percent report using their transfer to put money into a savings account, in comparison to 93 percent reporting using their transfer to pay bills. As such, it appears that the vast majority of respondents spent their transfers as opposed to saving them.

Table 1: How Treatment Group Said Spent Transfer

	(1) % Spent Transfer on Category
Pay bills	92.8
Pay rent	54.4
Buy food	63.2
Home repairs or items	13.5
Buy children’s clothing	48.8
Buy other things for children (toys, books)	18.3
Put money into a saving account	05.5
Pay for health care	03.5
Pay loans	09.9
Buy other necessities	39.1
Pay for entertainment	04.1
Observations	4,483

7.2 Effects of Place

Relative to our first experiment, we targeted the second experiment to SNAP users living in high poverty zipcodes where baseline rates of hardship may be elevated. We also stratified treatment assignment to ensure balanced proportions of treatment and control respondents by state-level of government benefit generosity. In order to examine treatment effect heterogeneity across geographic areas, we estimate additional models in which we split participants into those living in relatively higher poverty zipcodes (defined as those with poverty rates of at least 41%) and lower poverty zipcodes (defined as those with poverty rates from 35 to 40%). In addition, we estimate separate models on participants living in states with low, medium and high social benefit levels (as measured by the maximum TANF benefits for a three-person household in the state). The results, shown in Appendix Table A.6, show no clear pattern.

7.3 Robustness

7.3.1 Alternate Subgroup Definitions

Following the procedure used in our first study (RCT-1), we defined a very low-income subgroup as households that reported less than \$500 of monthly earnings in the prior month. However, as there are multiple

¹¹Note that respondents could select multiple categories.

ways to measure economic well-being, there may be alternative ways to define our subgroups that better capture underlying economic well-being of families. As noted in our pre-analysis plan, we hypothesized that the cash transfer would have a larger impact on households with fewer economic resources (Jacob et al. (2020)). We focus these robustness checks on the material hardship outcome because this was the most robust finding in our first study (RCT-1).

First, among those reporting less than \$500 in monthly earnings, there may be important variation in terms of how long their earnings had been below this threshold. To approximate for this variation, we split our full study sample into three groups based upon when they report first receiving SNAP benefits. The first three columns of Appendix Table A.7 show treatment effects on our material hardship composite for those that were receiving SNAP immediately prior to the onset of COVID (column 1), those that report they had received SNAP in the past but were not receiving these benefits immediately prior to the onset of COVID (column 2), and those that started receiving SNAP for the first time ever as a result of COVID (column 3). We expect that those who were receiving SNAP before COVID might represent a more “permanent” very low-income group, the second group is a more “precarious” group and the final group may be experiencing real economic hardship for the first time. The middle group, which recently received SNAP as a result of COVID but had received food assistance at some point in the past, shows a marginally significant reduction in material hardship due to transfer receipt. There are no significant findings for the other two groups.

Next, we incorporate additional information that may be predictive of material hardship when defining subgroups. Prior research shows that multiple factors are associated with a family’s experience of material hardship, including not only earned income, but also benefit receipt, household composition, demographics and others (Rodems and Shaefer (2020)). Ideally, one might test whether the cash transfer reduces material hardship the most among households who, in the absence of the intervention, would have been predicted to experience the most material hardship.

To examine this, we approximate for levels of hardship by predicting material hardship using data on the control group from our first study, RCT-1 and a set of covariates collected in both studies. We then apply the estimated coefficient of each covariate to the same variables in RCT-2 in order to predict an approximate level of hardship in the current study (RCT-2) and split the sample into three groups based on the predicted distribution of material hardship. Within each tercile, we estimate the LATE on material hardship, using the same control variables as all other regression models in the current study (RCT-2). Note that this avoids the well-known bias that arises due to endogenous stratification (Abadie et al. (2018)).¹² Columns 4-6 of Appendix Table A.7 show results from this exercise. Across groups, we estimate very small treatment effects, with only the middle tercile showing any statistical significance. Taken together, the results in Appendix Table A.7 suggest that our null results are not simply the result of how we defined the economically disadvantaged subgroup.

7.3.2 Differential Response

While the survey response rates were approximately balanced across treatment and control groups, it is possible that our results may be biased by compositional differences between treatment and control samples that arose due to differential survey response. As is the case in many experiments, the initial survey response rate in this study was substantially higher among the treatment group. One week after the opening of the survey, the response rate among the treatment group was roughly 6 percentage points higher than among the control group: 24% versus 18%. We sent additional reminders to the control group and offered them a

¹²In theory, one should adjust the standard errors of these estimates to account for the uncertainty introduced by using estimated coefficients to create the subgroups. Given the overall lack of significance, this would not change our inferences.

larger incentive and obtained response rates of roughly 30% in both groups.

To determine whether this additional, differential recruitment may have changed the sample in ways that biased our results, we examine a set of individuals who responded to the survey during the first week it was offered and thus received identical recruitment messages and incentives. The treatment and control groups in this “early survey response” set look comparable in terms of observable characteristics, and may be more similar along unobservable dimensions (say eagerness to participate in surveys). Even when we limit the sample to respondents exposed to identical recruitment tactics, we do not find any significant impacts on material hardship in the full or very low-income sample (see Appendix Table A.8).

In sum, our robustness checks confirm our main finding that the second cash transfer (RCT-2) had little to no effect on material hardship, reported labor supply, benefit utilization or savings behavior. These null effects are robust to various subgroup analyses. In the next section of the paper, we explore possible hypotheses for why we found sizable, significant effects for the very low-income subgroup in our first intervention occurring in May 2020, but no evidence that the cash transfer impacted outcomes in the second experiment.

8 What might explain the different results across the two studies?

The results from the second experiment (RCT-2) differ from our first experiment (RCT-1) in which we found the cash transfer reduced material hardship and improved mental health for very low-income households. The null results in RCT-2 are particularly puzzling as we had expected to find even larger effects in this second study, both because of changes we made to the survey to better capture material hardship as well as the fact that there were fewer federal benefits available in September 2020 compared with May 2020. Additionally, we expected to find larger effects in our second intervention, in part because we targeted treatment to low-income zip codes where baseline hardship was more likely to be elevated. In this section, we explore several potential explanations for the differences across our two experimental studies.

8.1 Hypothesis 1: Observable Sample Differences

Our first potential explanation for the different results across the two studies involves differences in sample composition (see Table 2). RCT-1 was limited to families with children in a relatively small set of predominantly Midwestern states. RCT-2 drew participants from a wider set of states and included households without children, but limited participation to households living in zip codes with poverty levels greater than 35% in the 2018 American Community Survey five-year moving average estimates.

To what extent could the observable sample differences across the two studies explain differences in the treatment effects? To explore this question, we first create samples from RCT-1 and RCT-2 that are limited to households with children, no SSDI or SSI income, residing in one of the RCT-1 study states and within a high-poverty zip code. Appendix Table A.9 provides summary statistics for these comparison samples.¹³ Even in these comparison samples, there are some differences between households in RCT-1 and RCT-2. For example, RCT-2 respondents have been using the Fresh EBT app longer, are more likely to be Hispanic or White (and less likely to be Black), have lower levels of education and are more likely to be married than those in RCT-1. Looking at the very low-income comparison samples from RCT-1 and RCT-2 (columns 2

¹³Note, for RCT-1 we focus on the one-month survey (one month post transfer rather than three months post transfer) as the timing of this follow-up is similar to the timing between the cash transfer and the survey in RCT-2.

Figure 2: Sample Composition Differences Between RCT-1 and RCT-2

	<i>RCT-1</i>	<i>RCT-2</i>
<i>State</i>	IN, KY, LA, ME, MI, MN, NC, NH, NM, OH, PA, RI.	49 states, Washington D.C. and Puerto Rico in 1,265 high-poverty zip codes
<i>Family type</i>	Families with children	All low-income individuals/families
<i>Receive SSI</i>	Excluded	Included
<i>To make comparison sample</i>	Constrain to high-poverty zip codes included in RCT-2	Constrain to 1) families with children 2) not on SSI/SSDI 3) in RCT-1 sample states excluding NH, RI and ME because very limited RCT2 sample coverage

and 4 of Appendix Table A.9), we also find some remaining differences.

To account for the remaining differences across samples, we reweight RCT-1 (RCT-2) households in proportion to how closely they resemble RCT-2 (RCT-1) households. Specifically, we stack all observations from the comparison samples of both studies and estimate a probit model where the outcome is an indicator for being part of one of the RCTs and the predictors include characteristics likely unaffected by treatment including race, sex, education, age, household composition, per capita household income in 2019, and an indicator equal to one if a family used SNAP before COVID. We also include the zip code characteristics listed in Appendix Table A.1. We then use the predicted probabilities from these models to weight our treatment effect estimates. For the purpose of comparability, we create a slightly different version of our material hardship measure in RCT-2 that more exactly matches the composite used in RCT-1 and present LATE estimates for both studies. In addition, we use an identical set of covariates across studies, as described in the table notes.¹⁴

Table 6 compares treatment effects across the two experiments for the very low-income group. Column 1 presents estimates from the original samples, illustrating the significant impact of the cash transfer in RCT-1 as well as the small and non-significant result for RCT-2. Note that the estimate of 0.018 for RCT-2 differs slightly from the 0.017 shown in Table 3 for the reasons described above. Column 2 shows results for the comparison samples. Specifically, we estimate the treatment effects for RCT-2 respondents weighting by the predicted probability of being in the RCT-1 sample. We repeat this analysis predicting the probability that RCT-1 individuals would have been included in RCT-2 and then re-estimate the effects for the RCT-1 sample weighting for the RCT-2 predicted probability.

Results shown in Panel A of Table 6 indicate that the effect of the cash transfer in RCT-1 becomes less beneficial for very low-income respondents as they more closely resemble the RCT-2 population. The treatment effect estimated with the original RCT-1 sample (-0.158) almost completely vanishes when we use the sample restricted to most closely resemble the RCT-2 population (-0.049). This suggests that sample differences and treatment effect heterogeneity may explain the differences between the first and second experiment. Unfortunately, the estimates are too imprecise to statistically distinguish them from each other.

¹⁴In practice, because there was virtually 100% take-up in RCT-1, the ITT equals the LATE. However, as discussed earlier, there was substantial noncompliance in RCT-2. Because [Pilkauskas et al. \(2022\)](#) report estimates based on a sample of individuals who completed both the one-month and the three-month survey, the estimates reported here differ slightly from those in the report on RCT-1. See Appendix 3 for the crosswalk between the material hardship items used in RCT-1 and RCT-2, along with how we adjusted the RCT-2 items to match those in RCT-1.

However, the pattern of results for RCT-2 does not tell the same story. We find very little difference between estimates using the original RCT-2 sample and the sample that is restricted and reweighted to resemble the RCT-1 study population. In both cases, the estimates are small and statistically indistinguishable from zero. Of course, the standard errors on all of these estimates are quite large, and so the conclusions here should be taken with a large grain of salt. Nevertheless, these results suggest that adjusting for observable differences between study samples would not have impacted null findings in the second experiment.

8.2 Hypothesis 2: Unobservable sample differences

The results presented above provide some evidence that observable differences across the RCT-1 and RCT-2 participants could be partly responsible for the different effects. Yet participants in RCT-1 and RCT-2 may have differed in ways that are harder to observe, stemming from the procedures used to recruit and survey respondents in each study. As explained in Appendix 5, individuals in RCT-1 might be described as “eager beavers” - that is, users who were quite engaged with the app and who were inclined to engage with content such as banner ads and survey opportunities. Relative to RCT-1, we suspect that the enrollment process in RCT-2 resulted in less eager participants - that is, individuals who may have seen the enrollment and/or survey banners multiple times and received push notifications before deciding to investigate the GiveDirectly offer and/or respond to our survey.

While there is no way to directly measure the “eagerness” of study participants, we conducted a subgroup analysis for RCT-2 intended to focus on individuals who we believe may be most eager and, in this way, be more similar to those in RCT-1. Specifically, we limit the analysis to users who responded to the follow-up survey within five hours of logging into the app after the survey banner was visible, excluding individuals who first logged onto the app after receiving a push notification. While our estimates for this relatively small sample are less precise than our baseline estimates, we do not find any evidence to suggest that these eager respondents in RCT-2 benefited from the cash transfer. Using control variables common across both studies and a comparable material hardship index, the point estimate (s.e.) for the material hardship outcome for the full group of eager respondents ($n=2,936$) is 0.050 (0.051); the corresponding estimates for the very low-income eager respondents ($n=1,600$) is 0.017 (0.071).

8.3 Hypothesis 3: Impacts dissipated by time survey fielded

In RCT-1, those who were treated completed our one-month survey 27 days after claiming their transfer payments on average, and approximately 90 percent of those who were treated completed the survey between 20 and 33 days after claiming their transfer (see Appendix 5 for additional survey details). On the other hand, of the 4,825 treatment users in the RCT-2 analysis sample who successfully cashed out their transfer from Hyperwallet, 12.8 % responded to our survey within 30 days of cashing out, 64.5 % responded to our survey within 30 to 45 days ($N=3111$), and 22.7 % responded after 45 days of cashing out. The average length of time between cash receipt and survey response was 39 days.

Given the low levels of financial resources in our study sample, it is possible that benefits of the cash transfer could have dissipated by the time some of our respondents completed the survey in RCT-2. If we limit the analysis to RCT-2 individuals who completed the survey within the first week of administration, the average length of time between cash receipt and survey response for treatment compliers is 36 days, and 15.6 % responded within 30 days of cash receipt. Table A.10 shows that the results for this sample are virtually identical to the original sample (for full as well as low-income groups). It is also worth noting that although time between the cash transfer and the measured outcomes might explain the null findings in

RCT-2, in RCT-1 we also conducted a follow-up survey three months post-transfer and continued to find a link between treatment and reduced hardship and mental health problems for the very low-income group.

8.4 Hypothesis 4: Effects of differential priming

As described in Appendix 5, in RCT-1, the research project was closely connected to the intervention in the minds of participants. Immediately after being invited to participate in the study and providing their contact information, treatment individuals were offered the \$1,000 cash award. Both treatment and control groups were contacted via email later that evening to take the baseline survey, and then contacted 30 days later to take the one month survey. While we did not inform participants that a goal of the study was to evaluate the impact of a cash transfer, it seems likely that the treatment group individuals may have made this connection in their minds. If this were true, the results in RCT-1 could have suffered from bias, although the direction of bias is not clear ex-ante. On the one hand, treatment households may have felt compelled to minimize their reported hardship to ensure the intervention was a success (i.e., a standard Hawthorne effect). On the other hand, it is possible that treated households believed they would be eligible for additional money if they reported greater hardship in the follow up survey, although we have no way to test for this potential bias in our data.

8.5 Hypothesis 5: Changes in underlying relationship between income and material hardship between two experiments

Poor families in the U.S. experienced many changes during the summer and fall of 2020 as a result of the continuing COVID pandemic, state responses to the pandemic, and federal intervention intended to provide support to individuals and businesses. For this reason, it is possible that the relationship between income and material hardship may have dissipated by fall 2020. To examine this possibility, Table 7 presents OLS estimates of the relationship between material hardship and income using control households in the two studies.

In column 1, we see a similar relationship between earnings and material hardship across the two studies. In RCT-1, an additional \$1,000 is associated with .202 SD lower hardship; in RCT-2, the relationship is .166 SD. Column 2 restricts the analysis to control households in the comparison sample (described above). Here the relationship appear *stronger* in RCT-2. To explore the relationship between material hardship and a more inclusive measure of income resources, we calculate each household's financial resources as the sum of reported monthly labor earnings and reported benefit amounts from unemployment insurance, SNAP, cash assistance, Social Security, disability income and regular financial assistance such as child support or alimony. Columns 3 and 4 suggest that using this broader resource measure, the relationship between additional income and material hardship is stronger for respondents in RCT-1. For example, among control households in the comparison sample, an additional \$1,000 of financial resources is associated with .226 SD and .164 SD lower material hardship in RCT-1 and RCT-2 respectively. Overall, these results do not provide compelling evidence that the difference in results across studies is due to a change in the underlying relationship between income and material hardship.

9 Conclusion

This paper reports on the second of two randomized control trials evaluating a one-time \$1,000 unconditional cash transfer program administered by GiveDirectly and conducted via the Fresh EBT (now called Providers) application. In our first study, while there were no statistically significant impacts in the full sample, the intervention reduced material hardship among the most economically disadvantaged households in

the study and suggested improvements in mental health. In the second study, we find no discernible impacts on any of the outcomes we measure for either the full sample or any of the pre-specified subgroups.

The results from these studies do not lead to a straightforward conclusion about the efficacy of a modest, one-time cash transfer in mitigating hardship and improving mental health. It remains possible that a one-time cash transfer of \$1,000 is not enough to move our outcomes of interest even in the short-term. It is also possible that the effects of the one-time transfers were diluted because of the unprecedented expansion of cash transfers by the federal government in response to the pandemic. Finally, it may be that the cash transfers did provide true benefits to families, but our survey measures were not able to capture these effects.

One thing remains clear: low income families in the U.S. face a high degree of financial complexity that is difficult to capture with standard measures. Going forward, we suggest that it is essential for the research and policy communities to consider several issues related to unconditional cash transfers. First, we encourage those who design programs and those who evaluate them to think deeply about what outcomes we would expect cash transfers of various sizes and forms (one-time or recurring) to impact. The nature of the desired outcome should determine the design of the cash transfer. For example, if the goal is to impact child development, it is likely that one-time payment is insufficient. On the other hand, if the goal is to reduce the likelihood of homelessness, a well-targeted one-time transfer may have a large effect.

Second, we believe it is essential to move beyond survey-based measures because of the concerns related to survey response bias. In our own work, we are exploring the use of credit bureau data and administrative data on earnings and social benefit receipt. If researchers continue to use surveys, it would be valuable to collect very general subjective wellbeing measures - e.g., asking respondents at random times to describe how they are feeling, and how happy or satisfied they are with life - as well as the type of material hardship measures studied in our study.

Third, we urge researchers to think creatively about how best to measure outcomes in cash transfer studies. Given the flexibility of cash transfers, we propose the exploration of an “adaptive outcomes” approach. If delivering transfers as cash means that households can allocate aid towards their greatest need, leading to heterogeneous spending patterns, then researchers might seek out ways to assess what sample members identify as their area of greatest need, and then assess progress in meeting that need.

In our view, future research would test a series of cash transfers of varying sizes, where some recipients receive lump sum transfers and some receive recurring payments. For example, one treatment arm might receive a \$6,000 lump sum while a second treatment arm might receive a \$1,000 payment made over 6 months. Another treatment arm might include a lump sum payment followed by recurring payments to test whether the two forms of transfer work better in tandem. Researchers should then examine both administrative and survey outcomes, including an outcome identified by participants (prior to random assignment) as most salient.

This is an exciting time for the evaluation of cash transfer programs in the United States. There is considerable non-experimental evidence that cash transfers can be an effective way to address economic hardship in the U.S., and a small number of ongoing experimental studies that will soon inform the debate. Yet the findings from our partnership with GiveDirectly suggests that there is a long way to go in understanding the best ways to design and experimentally evaluate cash transfers programs. We hope our work will help guide policymakers and researchers in this endeavour.

References

- Abadie, A., Chingos, M. M., and West, M. R. (2018). Endogenous stratification in randomized experiments. *The Review of Economics and Statistics*, C:567–580.
- Akee, R., Copeland, W., Keeler, G., Angold, A., and Costello, J. (2010). Parents' incomes and children's outcomes: A quasi-experiment using transfer payments from casino profits. *American Economic Journal: Applied Economics*, 2:86–115.
- Amirkhan, J. H. (2018). A brief stress diagnostic tool: The short stress overload scale. *Assessment*, 25:1001–1013.
- Anderson, M. 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*, 103:1481–1495.
- Bastagli, F., Hagen-Zanker, J., Harman, L., Barca, V., Sturge, G., Schmidt, T., and Pellerano, L. (2016). Cash transfers: what does the evidence say? a rigorous review of programme impact and of the role of design and implementation features.
- Batra, A., Karasek, D., and Hamad, R. (2022). Racial differences in the association between the u.s. earned income tax credit and birthweight. *Women's Health Issues*, 32:26–32.
- Collin, D. F., Shields-Zeeman, L. S., Batra, A., Vable, A. M., Rehkopf, D. H., Machen, L., & Hamad, R. (2020). Short-term effects of the earned income tax credit on mental health and health behaviors. *Preventive medicine*, 139, 106223.
- Davis, B. and Handa, S. (2015). How much do programmes pay? transfer size in selected national cash transfer programmes in sub-saharan africa. *Innocenti Research Brief*.
- Evans, W. N. and Garthwaite, C. L. (2014). Giving mom a break: The impact of higher eite payments on maternal health. *American Economic Journal: Economic Policy*, 6:258–290.
- Gennetian, L. A., Shafir, E., Aber, J. L., and Hoop, J. D. (2021). Behavioral insights into cash transfers to families with children. *Behavioral Science and Policy Ass*, 7.
- Gierk, B., Kohlmann, S., Toussaint, A., Wahl, I., Brünahl, C. A., Murray, A. M., and Löwe, B. (2015). Assessing somatic symptom burden: A psychometric comparison of the patient health questionnaire-15 (phq-15) and the somatic symptom scale-8 (sss-8). *Journal of Psychosomatic Research*, 78:352–355.
- Golosov, M., Graber, M., Mogstad, M., and Novgorodsky, D. (2021). How americans respond to idiosyncratic and exogenous changes in household wealth and unearned income.
- Imbens, G. W., Rubin, D. B., and Sacerdote, B. I. (2001). Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players.
- Jacob, B., Kapustin, M., and Ludwig, J. (2015). The impact of housing assistance on child outcomes: Evidence from a randomized housing lottery. *The Quarterly Journal of Economics*, pages 465–506.
- Jacob, B., Pilkauskas, N., Rhodes, E., Richard, K., and Shaefer, H. L. (2020). The covid cash transfer study: The impacts of an unconditional cash transfer on the wellbeing of low-income families.
- Kroenke, K., Spitzer, R. L., and Williams, J. B. W. (2001). The phq-9 validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16:606–613.

- Kroenke, K., Spitzer, R. L., and Williams, J. B. W. (2003). The patient health questionnaire-2: Validity of a two-item depression screener. *Medical Care*, 41:1284–1292.
- Milligan, K. and Stabile, M. (2011). Do child tax benefits affect the well-being of children? evidence from canadian child benefit expansions. *American Economic Journal: Economic Policy*, pages 175–205.
- Oosterbeek, H., Ponce, J., and Schady, N. (2008). The impact of cash transfers on school enrollment.
- Parolin, Z. (2021). Decomposing the decline of cash assistance in the united states, 1993 to 2016. *Demography*, 58:1119–1141.
- Pilkauskas, N., Jacob, B., Rhodes, E., Richard, K., and Shaefer, H. L. (2022). The covid cash transfer study: The impacts of an unconditional cash transfer on the wellbeing of low-income families.
- Pilkauskas, N. and Micheltore, K. (2019). The effect of the earned income tax credit on housing and living arrangements. *Demography*, 56:1303–1326.
- Plummer, F., Manea, L., Trepel, D., and McMillan, D. (2016). Screening for anxiety disorders with the gad-7 and gad-2: a systematic review and diagnostic metaanalysis . *General Hospital Psychiatry*, 39:24–31.
- Rodems, R. and Shaefer, H. L. (2020). Many of the kids are not alright: Material hardship among children in the united states. *Children and Youth Services Review*.
- Shaefer, H. L., Edin, K., Fusaro, V., and Wu, B. C. P. (2020). The decline of cash assistance and the well-being of poor households with children. *Social Forces*, pages 1000–1025.
- Spitzer, R. L., Kroenke, K., Williams, J. B., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The gad-7. *Archives of Internal Medicine*, 166:1092–1097.
- Sullivan, J. X., Turner, L., and Danziger, S. (2008). The relationship between income and material hardship. *Source: Journal of Policy Analysis and Management*, 27:63–81.
- Troller-Renfree, S. V., Costanzo, M. A., Duncan, G. J., Magnuson, K., Gennetian, L. A., Yoshikawa, H., Halpern-Meekin, S., Fox, N. A., Noble, K. G., Farah, M., and Luby, J. (2022). The impact of a poverty reduction intervention on infant brain activity. *PNAS*, 119.

Table 2: Balance Table of Survey Respondent Characteristics

	Control Mean (1)	Treatment Mean (2)	T Stat (3)	Perm P Val (4)	Std Diff (5)
N	6,720	7,035			
Claimed Transfer	0	4,825			
Zipcode % BA or More	14.525	14.670	0.839	0.437	0.014
Zipcode Median HH Income	25,961	25,975	0.140	0.866	0.002
Zipcode Pop Density	9,353	8,976	-1.307	0.270	-0.022
Zipcode % Black	46.835	0.000	0.109	0.904	0.002
Zipcode % Hispanic	28.966	28.745	-0.379	0.663	-0.006
Metro Area	0.891	0.887	-0.817	0.403	-0.014
Per-capita COVID deaths	2,119	2,056	-1.066	0.332	-0.018
Days using Fresh EBT	387.534	382.768	-0.750	0.569	-0.013
Number of App Uses	182.073	183.233	0.273	0.803	0.005
SNAP Benefits	291.306	297.249	1.302	0.249	0.028
Missing Benefits	0.371	0.368	-0.342	0.741	-0.006
Spanish Language	0.096	0.104	1.607	0.081	0.028
Age	36.806	36.749	-0.282	0.969	-0.005
Michigan	0.138	0.127	-1.829	0.083	-0.031
New York	0.108	0.105	-0.513	0.719	-0.009
Ohio	0.100	0.099	-0.267	0.748	-0.005
Pennsylvania	0.086	0.089	0.769	0.388	0.013
Puerto Rico	0.066	0.068	0.541	0.574	0.009
Texas	0.065	0.060	-1.257	0.144	-0.021
Hispanic	0.277	0.283	0.766	0.380	0.014
White Non-Hispanic	0.158	0.141	-2.584	0.024	-0.045
Black Non-Hispanic	0.515	0.528	1.476	0.117	0.026
Female	0.854	0.850	-0.781	0.735	-0.014
Less than High School	0.385	0.392	0.820	0.161	0.015
High School Diploma	0.265	0.254	-1.492	0.300	-0.026
Some College	0.133	0.128	-0.925	0.194	-0.016
Associates Degree or More	0.084	0.079	-1.017	0.396	-0.017
Age	36.83	36.80	-0.171	0.969	-0.003
Household Size	3.515	3.599	2.140	0.026	0.037
Total Number of Kids	1.797	1.842	1.488	0.090	0.027
Lives with Spouse	0.195	0.191	-0.515	0.613	-0.009
Lives with Own Children	0.747	0.751	0.570	0.551	0.010
Lives with Children of Others	0.086	0.091	1.174	0.237	0.020
Lives with Related Adults	0.178	0.193	2.184	0.036	0.038

Note: See Appendix Table 1 for a complete list of variables and Appendix 4 for a list of which variables were used to randomize RCT2, as controls in RCT2, and as controls in RCT1. Treatment and control means are calculate for the full sample. Each characteristic is regressed on a indicator for treatment status, with standard errors clustered at the zipcode level; column (3) reports the T-statistic for each bivariate regression; column (5) reports the treatment coefficient, divided by the standard deviation of each characteristic among the control group. Column (4) displays the proportion of simulated T statistics that are larger than the observed T statistic of each bivariate regression. We produce simulated T statistics by re-randomizing treatment assignment according to original assignment rules, and computing the T statistic of each characteristic regressed on treatment status if and only if assignment conditions are satisfied. As such, the denominator for each permutation p value varies, depending on the number of satisfactory re-randomization, but on average is around 820.

Table 3: Primary Outcome Treatment Effects of Cash Transfer

	Material Hardship		Mental Health Challenges	
	(1)	(2)	(3)	(4)
<i>Full Sample</i>	ITT	LATE	ITT	LATE
Treatment	0.005	0.007	-0.018	-0.026
Zipcode Cluster Robust SE	(0.016)	(0.023)	(0.018)	(0.026)
Randomization Inference P Value		[0.99]		[0.99]
Observations	13,738	13,738	13,718	13,718
Control Mean	0.000		0.000	
Control Complier Mean		0.039		0.025
<i>Very Low-Income Sample</i>	ITT	LATE	ITT	LATE
Treatment	0.012	0.017	-0.011	-0.017
Zipcode Cluster Robust SE	(0.023)	(0.034)	(0.024)	(0.035)
Randomization Inference P Value		[1.00]		[0.99]
Observations	7,260	7,260	7,242	7,242
Control Mean	0.121		0.030	
Control Complier Mean		0.170		0.046

Note: Columns 1 and 3 present the effects of Intent-to-Treat (ITT) specifications, while columns 2 and 4 show Local Average Treatment Effects (LATE) obtained by instrumenting for treatment assignment with an indicator equal to one if a user successfully claimed their cash transfer. The top panel shows effects for the full sample of survey respondents, while the bottom panel restricts to those users reporting no more than 500 dollars of monthly earnings in the prior 30 days. For the LATE specifications, we also show a “randomization inference p-value” which are computed as described in the main text and indicate the proportion of “pseudo treatment effects” that are larger than our observed effect. To contextualize effect magnitudes, we present the control group mean for ITT models, and the control complier mean for LATE models. The latter statistic is obtained by instrumenting for treatment assignment with an indicator equal to one if a user did not claim a transfer and estimating effects on each outcome variable only for the subset of users that did not cash out transfer payments. The control complier mean tells us the average outcome for control group members who would have plausibly claimed their transfers if offered. All models use the same set of control variables as outlined in Appendix 1. Standard errors in parentheses and are clustered at the zipcode level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: LATE Estimates of Exploratory Economic Well-being Measures

	Hardship Avoidance	Core Debt \$	Total Debt \$	Low Food Security	Very Low Food Security
	(1)	(2)	(3)	(4)	(5)
Full Sample					
Treatment	-0.012 (0.025)	-17.128 (33.231)	-8.389 (319.101)	0.010 (0.011)	-0.010 (0.012)
Observations	12,892	13,755	13,755	13,755	13,755
Control Complier Mean	-0.004	863.124	5,625.94	0.352	0.482
Very Low-Income Sample					
Treatment	0.013 (0.035)	-20.819 (49.011)	93.080 (413.330)	-0.003 (0.016)	0.003 (0.017)
Observations	6,803	7,269	7,269	7,269	7,269
Control Complier Mean	-0.071	876.694	4,721.89	0.348	0.509

Note: Coefficients are Local Average Treatment Effects (LATE) obtained by instrumenting for treatment assignment with an indicator equal to one if a user successfully claimed their cash transfer. The top panel shows effects for the full sample of survey respondents, while the bottom panel restricts to those users reporting no more than 500 dollars of monthly earnings in the prior 30 days. Column 1 shows results for a composite index of the following items: an indicator equal to one if a user obtained free food from a food bank or pantry, the number of times a user obtained free food from a food bank or pantry, an indicator equal to one if a user cut their spending in the past 30 days to make ends meet, the number of times a user sold a household item for cash in the past 30 days, and the reported income earned from sold items (top coded at \$ 2,500). Columns 2 and 3 show dollar amounts of reported debt. We group rent, phone and utilities debt as “core” items, and include medical, car, family loans, credit card and other forms of debt in our “total” debt measure. Finally, to measure food insecurity, we create a 6 item scale equal to the number of food-related hardships (an indicator if meal size was cut or meals were skipped, an indicator if a user reports having cut or skipped meals at least four times in the past month, an indicator if couldn’t afford to eat balanced meals, an indicator if food did not last and there was not money to buy more, an indicator if user reported eating less due to financial constraints, and an indicator if a user reported being hungry due to insufficient food. Users who report experiencing 3 or 4 of these hardships are considered to have “Low food security,” while users who report experiencing 5 or 6 of these hardships are considered to have “very low food security.” Standard errors in parentheses and clustered at the zipcode level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: LATE Estimates of Exploratory Mental Health Measures

	SOS Index	SSS Index	Depression	Anxiety
	(1)	(2)	(3)	(4)
<i>Full Sample</i>				
Treatment	0.000 (0.025)	0.020 (0.025)	-0.017 (0.013)	-0.010 (0.012)
Observations	13,513	13,397	13,755	13,755
Control Complier Mean	0.021	-0.012	0.478	0.464
<i>Very Low-Income Sample</i>				
Treatment	-0.013 (0.034)	0.003 (0.036)	-0.007 (0.017)	-0.015 (0.017)
Observations	7,123	7,071	7,269	7,269
Control Complier Mean	0.035	0.002	0.487	0.481

Note: Coefficients are Local Average Treatment Effects (LATE) obtained by instrumenting for treatment assignment with an indicator equal to one if a user successfully claimed their cash transfer. The top panel shows effects for the full sample of survey respondents, while the bottom panel restricts to those users reporting no more than 500 dollars of monthly earnings in the prior 30 days. Columns 1 and 2 are as described in the text. Column 3 shows effects on the probability that a user indicated they felt down, depressed or hopeless, or indicated feeling little pleasure or interest in doing things, more than half the days last month. Column 4 shows effects on the probability that a user indicated being unable to stop worrying, or indicated feeling anxious or on edge, more than half of the days last month. Standard errors in parentheses and are clustered at the zipcode level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Comparison of Treatment Effects in Very Low-Income Samples

	Original Experiment	Comparison Re-weighted Sample
	(1)	(2)
<i>Panel A - RCT 1 Respondents</i>		
Treatment	-0.158*** (0.038)	-0.049 (0.094)
Observations	2805	510
Control Complier Mean	0.248	0.214
<i>Panel B - RCT 2 Respondents</i>		
Treatment	0.018 (0.034)	0.097 (0.077)
Observations	7266	1568
Control Complier Mean	0.162	0.183

Note: Coefficients are Local Average Treatment Effects (LATE) obtained by instrumenting for treatment assignment with an indicator equal to one if a user successfully claimed their cash transfer. All samples are restricted to users reporting monthly earnings less than or equal to \$500 in the prior month. The respondents in RCT-1 are those who responded to the one-month survey. Comparison samples refer to columns 3 and 6 of Table 5. Coefficients differ from other tables for two reasons: 1) our outcome variable is a version of the material hardship index made to be comparable across experiments (see Appendix 3 for details), and 2) all models use a set of covariates available in both studies. To obtain propensity scores used to re-weight samples, we run a probit model of an indicator for inclusion in each respective comparison sample on a set of observable characteristics including application utilization, demographic and household composition, zipcode covariates and state fixed effects. Standard errors are in parentheses and clustered at the zipcode level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Comparison of Anticipated Effects of Additional Income on Material Hardship

	Earnings		Simulated Resources (Earnings + Benefits)	
	(1)	(2)	(3)	(4)
<i>RCT-1 Respondents</i>	Control	Comparison Control	Control	Comparison Control
Effect of Additional \$1000	-0.202*** 0.020	-0.124 (0.076)	-0.203*** (0.018)	-0.226*** (0.044)
Observations	2394	389	2180	317
Control Mean	932.210	822.084	851.094	975.696
<i>RCT-2 Respondents</i>	Control	Comparison Control	Control	Comparison Control
Effect of Additional \$1000	-0.166*** 0.015	-0.205*** 0.026	-0.127*** 0.012	-0.164*** 0.025
Observations	6679	1644	6714	1647
Control Mean	713.658	752.078	1423.259	1392.539

Note: Coefficients are obtained via ordinary least squares regressions of monthly resources on the comparable material hardship composite. All models are run on control respondents only. For RCT1, the simulated resources measure is regressed on the material hardship composite obtained in the three month survey, but the earnings measure is regressed on the material hardship composite obtained in the one month survey. For RCT2, all outcomes and regressors come from the single follow up survey, occurring 4-6 weeks after the intervention. All models include a set of controls common across experiments. For RCT2 models, we use a version of the material hardship composite made to comparable across experiments. Standard errors are clustered at the zipcode level. See Appendix 4 for a description of how the simulated resources measure is computed.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendices

Jacob, Brian, Pilkauskas, Natasha, Rhodes, Elizabeth, Richard, Katherine and Shaefer, Luke (2022). “The COVID Cash Transfer Study II: The Impacts of an Unconditional Cash Transfer on the Wellbeing of Low-Income Families.” *National Tax Journal*, 75(3)

Table A.1: Balance Table of Survey Respondent Characteristics

	Control Mean	Treatment Mean	T Stat	Perm P val	Std Diff
	(1)	(2)	(3)	(4)	(5)
N	6,720	7,035			
Aggregate and Administrative Variables					
Zipcode PCT BA or More	14.525	14.670	0.839	0.437	0.014
Zipcode Median HH Income	25,961.410	25,974.754	0.140	0.866	0.002
Zipcode Median Rent	747.407	745.092	-0.747	0.490	-0.013
Zipcode PCT Below FPL	41.489	41.410	-0.776	0.425	-0.013
Zipcode Pop Density	9,353.430	8,976.383	-1.307	0.270	-0.022
Zipcode PCT Black	46.835	46.896	0.109	0.904	0.002
Zipcode PCT Asian	1.921	1.854	-0.996	0.436	-0.017
Zipcode PCT Hispanic	28.966	28.745	-0.379	0.663	-0.006
Metro Area	0.891	0.887	-0.817	0.403	-0.014
Micropolitan Area	0.030	0.029	-0.117	0.932	-0.002
Commuting Zone	0.025	0.030	1.813	0.073	0.032
Small Rural Area	0.042	0.043	0.193	0.844	0.003
Missing RUCA	0.012	0.011	-0.529	0.447	-0.009
Per capita COVID deaths	2,119.410	2,056.174	-1.066	0.332	-0.018
Days using Fresh EBT	387.534	382.768	-0.797	0.569	-0.013
Had Account Before 2/2020	0.559	0.550	-1.123	0.164	-0.018
Number of App Uses	182.073	183.233	0.312	0.803	0.005
SNAP Benefits	291.306	297.249	1.274	0.249	0.028
Missing Benefits	0.371	0.368	-0.345	0.741	-0.006
Spanish Language	0.096	0.104	1.672	0.081	0.028
Lives in Low Sample Mass State	0.005	0.004	-1.820	0.057	-0.025
Alabama	0.020	0.018	-0.643	0.526	-0.013
Arkansas	0.002	0.002	0.060	0.932	0.001
Arizona	0.012	0.013	0.558	0.617	0.009
California	0.036	0.035	-0.326	0.746	-0.006
Connecticut	0.004	0.006	0.618	0.616	0.021
Delaware	0.002	0.002	0.239	0.972	0.001
Florida	0.039	0.045	1.703	0.081	0.034
Georgia	0.039	0.042	0.573	0.567	0.011
Illinois	0.039	0.041	0.682	0.492	0.011
Indiana	0.010	0.010	0.231	0.835	0.004
Kentucky	0.029	0.027	-0.369	0.726	-0.008
Louisiana	0.033	0.033	0.085	0.940	0.001
Maryland	0.013	0.011	-1.381	0.163	-0.024
Michigan	0.138	0.127	-1.519	0.083	-0.031
Minnesota	0.005	0.005	0.281	0.818	0.005
Missouri	0.016	0.021	2.087	0.028	0.037
Mississippi	0.027	0.031	1.039	0.279	0.022
North Carolina	0.007	0.008	1.477	0.144	0.015
Nebraska	0.001	0.002	1.074	0.405	0.014
New Jersey	0.014	0.011	-1.702	0.088	-0.025
New Mexico	0.017	0.017	-0.115	0.913	-0.002
New York	0.108	0.105	-0.577	0.719	-0.009
Ohio	0.100	0.099	-0.284	0.748	-0.005
Oklahoma	0.008	0.006	-1.484	0.138	-0.020
Pennsylvania	0.086	0.089	0.917	0.388	0.013
Puerto Rico	0.066	0.068	0.526	0.574	0.009
South Carolina	0.004	0.004	0.137	0.939	0.002
South Dakota	0.002	0.002	0.335	0.724	0.004
Tennessee	0.018	0.019	0.643	0.532	0.008
Texas	0.065	0.060	-1.402	0.144	-0.021
Virginia	0.003	0.004	1.075	0.286	0.016
Wisconsin	0.026	0.024	-1.707	0.083	-0.013
West Virginia	0.004	0.006	0.937	0.367	0.018

Note: Treatment and control means are calculate for the full sample. Each characteristic is regressed on a indicator for treatment status, with standard errors clustered at the zipcode level; column (3) reports the T- statistic for each bivariate regression; column (5) reports the treatment coefficient, divided by the standard deviation of each characteristic among the control group. Column (4) displays the proportion of simulated T statistics that are larger than the observed T statistic of each bivariate regression. We produce simulated T statistics by re-randomizing treatment assignment according to original assignment rules, and computing the T statistic of each characteristic regressed on treatment status if and only if assignment conditions are satisfied. As such, the denominator for each permutation p value varies, depending on the number of satisfactory re- randomization, but on average is around 820.

Table A.1: Balance Table of Survey Respondent Characteristics

	Control Mean	Treatment Mean	T Stat	Perm P val	Std Diff
	(1)	(2)	(3)	(4)	(5)
Baseline Survey Demographic Variables					
Hispanic	0.277	0.283	0.758	0.380	0.014
White Non-Hispanic	0.158	0.141	-2.405	0.024	-0.045
Black Non-Hispanic	0.515	0.528	1.453	0.117	0.026
Other Race	0.050	0.048	-0.735	0.490	-0.013
Female	0.854	0.850	-0.802	0.735	-0.014
Missing Sex	0.078	0.075	-0.572	0.555	-0.010
Less than HS Diploma	0.216	0.226	1.342	0.161	0.025
HS Diploma	0.385	0.392	0.896	0.300	0.015
Some College	0.265	0.254	-1.460	0.194	-0.026
Associates Degree or More	0.133	0.128	-0.902	0.396	-0.016
Age	36.833	36.800	-0.181	0.969	-0.003
Household Size	3.515	3.599	2.200	0.026	0.037
Missing Household Size	0.001	0.001	-0.809	0.421	-0.013
Number of Kids Under 6	0.674	0.690	0.927	0.344	0.017
Number of Kids 6-12	0.667	0.691	1.508	0.097	0.025
Number of Kids 13-18	0.473	0.477	0.272	0.742	0.005
Total Number of Kids	1.797	1.842	1.517	0.090	0.027
Lives with Spouse	0.195	0.191	-0.509	0.613	-0.009
Lives with Unmarried Partner	0.094	0.100	1.089	0.289	0.018
Lives with Own Children	0.747	0.751	0.586	0.551	0.010
Lives with Children of Others	0.086	0.091	1.199	0.237	0.020
Lives with Related Adults	0.178	0.193	2.218	0.036	0.038
Lives with Unrelated Adults	0.037	0.037	0.015	0.971	0.000

Note: Treatment and control means are calculate for the full sample. Each characteristic is regressed on a indicator for treatment status, with standard errors clustered at the zipcode level; column (3) reports the T-statistic for each bivariate regression; column (5) reports the treatment coefficient, divided by the standard deviation of each characteristic among the control group. Column (4) displays the proportion of simulated T statistics that are larger than the observed T statistic of each bivariate regression. We produce simulated T statistics by re-randomizing treatment assignment according to original assignment rules, and computing the T statistic of each characteristic regressed on treatment status if and only if assignment conditions are satisfied. As such, the denominator for each permutation p value varies, depending on the number of satisfactory re-randomization, but on average is around 820.

Table A.2: LATE Estimates for Material Hardship Items in last 30 days

	Unpaid rent	Unpaid electric or gas	Unpaid phone or internet	Water shutoff	Gas/electric shutoff	Phone/internet shutoff	Worried food would run out	Food didn't last, couldn't buy more
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Full Sample								
Treatment	0.015 (0.009)	0.010 (0.009)	0.004 (0.008)	-0.001 (0.006)	-0.008 (0.007)	0.002 (0.008)	0.004 (0.007)	0.002 (0.007)
Observations	10,972	11,609	12,621	13,719	13,712	13,708	13,755	13,755
Control Complier Mean	0.369	0.409	0.291	0.125	0.205	0.431	0.825	0.769
Very Low-Income Sample								
Treatment	0.002 (0.014)	0.005 (0.013)	0.012 (0.012)	0.003 (0.008)	-0.013 (0.009)	-0.004 (0.012)	0.005 (0.009)	0.010 (0.010)
Observations	5,405	5,866	6,440	7,244	7,238	7,232	7,269	7,269
Control Complier Mean	0.457	0.457	0.33	0.13	0.226	0.463	0.841	0.791
	Forego doctor due to cost	Forego prescription due to cost	Forego dentist due to cost	Moved in with others	Stayed in a shelter	Stayed in place not meant for housing	Evicted or forced to leave your home	Transport Insecurity
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Full Sample								
Treatment	-0.006 (0.006)	-0.005 (0.007)	-0.001 (0.008)	-0.000 (0.006)	0.002 (0.004)	0.002 (0.004)	0.005 (0.004)	0.002 (0.008)
Observations	13,738	13,725	13,724	13,737	13,725	13,722	13,718	13,736
Control Complier Mean	0.184	0.218	0.313	0.169	0.051	0.063	0.075	0.426
Very Low-Income Sample								
Treatment	-0.009 (0.009)	-0.000 (0.010)	-0.004 (0.011)	0.000 (0.009)	0.011+ (0.006)	0.001 (0.006)	0.008 (0.006)	0.004 (0.012)
Observations	7,261	7,249	7,249	7,262	7,251	7,250	7,248	7,257
Control Complier Mean	0.195	0.226	0.318	0.203	0.058	0.084	0.083	0.469

Table A.3: LATE Estimates for Additional Composite Indices

	Parenting Problems	Child Behavior Problems	Partner Conflict
	(1)	(2)	(3)
Full Sample			
Treatment	-0.008 (0.030)	-0.009 (0.031)	-0.035 (0.043)
Observations	8663	8629	3703
Control Complier Mean	0.009	-0.025	0.012
Very Low-Income Sample			
Treatment	-0.014 (0.044)	0.002 (0.042)	-0.087 (0.066)
Observations	4363	4351	1638
Control Complier Mean	0.009	-0.065	0.147

Note: Composite indices are constructed using method described in text. Low income sample includes respondents reporting earnings of less than \$ 500 in prior month. Standard errors in parentheses and clustered at zipcode level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: LATE Estimates for Benefit Receipt

	Use UI	Use TANF	Use SNAP	Use Housing Assistance	Use Social Security	Use Disability	Total Benefit \$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Full Sample							
Treatment	0.014 (0.010)	0.001 (0.009)	-0.001 (0.008)	0.000 (0.009)	-0.004 (0.006)	-0.002 (0.009)	13.044 (12.191)
Observations	13,633	13,605	13,662	13,596	13,579	13,599	13,755
Control Complier Mean	0.261	0.157	0.869	0.181	0.083	0.206	696.792
Very Low-Income Sample							
Treatment	0.007 (0.014)	0.005 (0.013)	0.006 (0.010)	-0.011 (0.012)	-0.001 (0.008)	0.010 (0.012)	6.565 (15.432)
Observations	7200	7180	7215	7172	7164	7182	7269
Control Complier Mean	0.274	0.174	0.864	0.205	0.064	0.163	633.589

Note: UI refers to Unemployment Insurance; TANF refers to Temporary Assistance for Needy Families; SNAP refers to Supplemental Nutrition Assistance Program; Disability refers to Supplemental Security Income or SSDI. Standard errors in parentheses and clustered at the zipcode level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: LATE Estimates for Labor Supply in last 30 days

	Worked Part or Full time	Hours Worked Last Week	Monthly Earnings	Job Search	Number of Search Tasks
	(1)	(2)	(3)	(4)	(5)
Full Sample					
Treatment	-0.012 (0.011)	0.462 (0.726)	-12.174 (19.362)	0.060*** (0.012)	0.023 (0.048)
Observations	13,755	4,579	13,699	13,714	6,737
Control Complier Mean	0.346	27.326	739.060	0.480	2.407
Very Low-Income Sample					
Treatment	-0.020 (0.012)	1.135 (1.272)	-0.347 (4.871)	0.062*** (0.016)	0.038 (0.065)
Observations	7269	1319	7269	7245	3952
Control Complier Mean	0.198	19.667	110.638	0.538	2.369

Note: Search tasks include looking at job postings, contacting someone about a job, applying to a job, posting a resume, having an interview, or "other." Standard errors in parentheses and clustered at zipcode level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Sample Splits on Zipcode Poverty Average and State Benefit Generosity

	Material Hardship					Mental Health Challenges				
	Lower Poverty Zipcodes	Higher Poverty Zipcodes	Low Benefit States	Medium Benefit States	High Benefit States	Lower Poverty Zipcodes	Higher Poverty Zipcodes	Low Benefit States	Medium Benefit States	High Benefit States
Full Sample										
Treatment	-0.044 (0.032)	0.065+ (0.035)	0.043 (0.042)	-0.026 (0.034)	-0.001 (0.051)	-0.003 (0.036)	-0.045 (0.035)	0.001 (0.044)	-0.040 (0.037)	-0.024 (0.054)
Observations	7,024	6,714	4,938	6,133	2,667	7,012	6,706	4,930	6,127	2,661
Control Complier Mean	0.085	-0.012	0.043	0.036	0.042	0.022	0.025	0.022	0.039	-0.013
Very Low-Income Sample										
Treatment	-0.049 (0.050)	0.098* (0.045)	0.048 (0.057)	-0.018 (0.043)	0.014 (0.096)	0.026 (0.049)	-0.040 (0.051)	0.026 (0.065)	-0.076 (0.052)	0.053 (0.063)
Observations	3,681	3,579	2,685	3,099	1,476	3,670	3,572	2,677	3,093	1,472
Control Complier Mean	0.254	0.076	0.151	0.185	0.174	0.032	0.048	0.012	0.099	-0.013

Note: Low poverty zipcodes are those with less than 40 percent of respondents living below the Federal Poverty Line in the 2018 ACS five year moving average data; high poverty zipcodes are those with more than 40 percent of respondents living below the FPL. Levels of benefit generosity are group based on the Urban Institutes Welfare Rules Database into terciles. Standard errors in parentheses and clustered at the zipcode level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Effects on Material Hardship with Alternative Subgroup Definitions

	Length of SNAP Use			LATE Estimates Joint Covariates		
	SNAP Before COVID	Prior SNAP but not Active Pre-COVID	Started SNAP due to COVID	Bottom Tercile	Middle Tercile	Top Tercile
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.028 (0.031)	-0.092+ (0.054)	0.050 (0.112)	0.012 (0.015)	0.012* (0.005)	0.017 (0.026)
Observations	8,162	2,740	871	6,901	1,304	5,533
Control Complier Mean	-0.030	0.206	0.180	-0.754	0.017	0.952

Note: Estimates in columns 1-3 from OLS regressions that include all controls. Estimates in columns 4-6 use the endogenous stratification approach described in the main text and a set of covariates common across RCT1 and RCT2.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: LATE Estimates for Identical Recruitment Sample

	Material Hardship Index	Hardship Avoidance Index	Core Debt (\$)	Total Debt (\$)	Low Food Security	Very Low Food Secu- rity	Mental Health Challenges
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Full Sample							
Treatment	0.003 (0.026)	-0.014 (0.029)	-26.274 (36.702)	-293.083 (350.944)	0.013 (0.012)	-0.008 (0.013)	-0.053* (0.026)
Observations	9815	9249	9824	9824	9824	9824	9797
Control Complier Mean	0.028	-0.009	845.549	5767.037	0.349	0.476	0.033
Very Low-Income Sample							
Treatment	0.024 (0.039)	-0.023 (0.040)	-41.001 (53.079)	-123.369 (452.549)	-0.011 (0.017)	0.013 (0.019)	-0.052 (0.038)
Observations	5216	4905	5220	5220	5220	5220	5200
Control Complier Mean	0.154	-0.037	860.218	4840.326	0.357	0.493	0.058

Note: Sample includes respondents exposed to identical recruitment methods. See Appendix 1 for a description of survey administration. Standard errors are in parentheses and clustered by zipcode.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Comparison of RCT1 and RCT2 Subsamples

	RCT-1		RCT-2	
	Comparison Sample	Very Low-Income Comparison	Comparison Sample	Very Low-Income Comparison
	(1)	(2)	(3)	(4)
N	1,033	532	3,392	1,744
Days using Fresh EBT	377.137	363.814	433.077	436.442
FEET Account Before 2/20	0.546	0.541	0.570	0.577
Number of App Uses	212.766	230.209	189.215	198.717
SNAP Benefits	348.198	372.787	363.543	367.171
Missing Benefits	0.393	0.464	0.271	0.278
Spanish Language	0.011	0.013	0.022	0.028
Hispanic	0.077	0.075	0.131	0.132
White Non-Hispanic	0.087	0.073	0.175	0.150
Black Non-Hispanic	0.815	0.825	0.644	0.664
Other Race	0.021	0.027	0.050	0.054
Female	0.952	0.953	0.922	0.904
Missing Sex	0.005	0.006	0.091	0.093
Less Than HS	0.145	0.187	0.198	0.241
HS Degree	0.429	0.459	0.413	0.431
Some College	0.308	0.278	0.283	0.250
Associates or More	0.119	0.076	0.107	0.078
Age	32.547	31.665	33.089	32.495
Missing Age	0.000	0.000	0.009	0.011
HH Size	4.062	4.030	4.078	3.991
Missing HH Size	0.005	0.006	0.001	0.001
Total Kids	2.591	2.684	2.511	2.470
Spouse	0.118	0.083	0.188	0.147
Nonspouse	0.139	0.128	0.103	0.079
Own kids	0.991	0.991	0.999	0.999
Other Kids	0.099	0.102	0.104	0.097
Relatives	0.114	0.096	0.145	0.130
Nonrelatives	0.018	0.019	0.021	0.022
Zipcode PCT BA or More	13.309	13.244	13.549	13.637
Zipcode Median HH Income	24,883.691	24,295.923	25,497.336	25,124.299
Zipcode Median Rent	708.511	689.301	724.317	718.840
Zipcode PCT Below FPL	42.364	43.385	41.563	41.907
Zipcode Pop Density	5,613.799	5,240.164	5,862.283	5,888.309
Zipcode PCT Black	64.285	64.507	56.625	57.706
Zipcode PCT Asian	1.280	1.273	1.577	1.687
Zipcode PCT Hispanic	8.174	7.700	12.375	12.266
Metro Area	0.962	0.972	0.910	0.917
Micropolitan Area	0.010	0.008	0.017	0.015
Commuting Zone	0.014	0.009	0.019	0.018
Small Rural Area	0.013	0.009	0.044	0.042
COVID deaths Jan to Jul	3,405.629	3,049.164	3,452.707	3,338.742
Indiana	0.018	0.021	0.027	0.030
Kentucky	0.038	0.049	0.066	0.068
Louisiana	0.037	0.034	0.089	0.096
Michigan	0.361	0.291	0.311	0.286
Minnesota	0.001	0.002	0.013	0.013
North Carolina	0.019	0.021	0.019	0.021
New Mexico	0.004	0.004	0.039	0.034
Ohio	0.376	0.461	0.251	0.261
Pennsylvania	0.146	0.118	0.184	0.190

Note: RCT1 Comparison Sample includes all respondents living in RCT2 zipcodes. RCT2 Comparison Sample includes all respondents with no children, not receiving SSI or SSDI, and living in RCT1 states. Both low income samples are defined as respondents reporting earning less than \$ 500 in the prior month.

Table A.10: LATE Estimates for respondents who took the survey in the first week

	Material Hardship Index	Hardship Avoid- ance Index	Core Debt (\$)	Total Debt (\$)	Low Food Security	Very Low Food Se- curity	Mental Health Chal- lenges
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Full Sample</i>							
Treatment	-0.002 (0.027)	-0.009 (0.030)	-52.611 (38.410)	-333.753 (351.761)	0.010 (0.012)	-0.004 (0.014)	-0.041 (0.028)
Observations	8,834	8,330	8,843	8,843	8,843	8,843	8,817
Control Complier Mean	0.036	-0.010	871.535	5822.235	0.352	0.472	0.030
<i>Very Low-Income Sample</i>							
Treatment	0.021 (-0.040)	-0.005 (0.041)	-70 (55.292)	-170 (477.322)	-0.011 (0.017)	0.020 (0.019)	-0.027 (-0.040)
Observations	4,665	4,381	4,669	4,669	4,669	4,669	4,650
Control Complier Mean	0.165	-0.051	893	4,920	0.355	0.489	0.053

Note: Sample includes respondents that completed follow up survey within the first week. Standard errors are in parentheses and clustered at the zipcode level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 1: Survey Recruitment and Administration

1 Recruitment in FreshEBT

On October 28, 2020, roughly 6 weeks after the first treatment group payments were claimed, FreshEBT placed a notification in the app inviting all treatment and control group members to participate in a 10-minute survey about their experiences during the COVID-19 pandemic in exchange for a \$10 gift card. Users who clicked on the banner were directed to a Qualtrics survey, and the first question requested contact information (either phone number, email address, or both) so we could follow up if respondents consented to participate but did not finish the survey. The referral link from the FreshEBT app included the user's FreshEBT ID, which enabled us to merge the respondent's treatment status and other administrative data provided by Propel and GD with survey responses.

The survey remained open until December 1, 2020. The notification was no longer displayed once an individual started the survey, but it appeared on the home screen of the app for all nonresponders in the sample throughout the survey administration period. FreshEBT sent three push notifications to individuals in the sample encouraging them to open the app and participate in the study, though whether individuals received the push notifications depended on their version of the app and notification settings within the phone's operating system. The first push went to all nonresponders on November 2, the second went only to control group nonresponders on November 11, and the third – also limited to the control group – was sent on November 23.

2 Follow up for partial responses

Since we collected respondents' contact information at the beginning of the survey, we implemented a system of regular follow up reminders and eventual escalation of incentives to maximize response rates. Participants were assigned to groups according to the day they started the survey, and every unfinished member of each group received the first reminder the day after starting the survey. The sequence of reminders alternated between emails and text messages (texts were only sent to individuals who consented to be contacted via SMS when providing their phone number) sent roughly every other day, not to exceed 8 total reminders. Individuals who started the survey earlier received all 8 reminders in the first few weeks and were not subsequently contacted, but the survey remained open and they were able to finish until it was closed on December 1. Individuals who started closer to the closure date received fewer reminders, though we switched to daily reminders during the final week given the compressed time frame. Ultimately, we obtained contact information for 21,843 of the roughly 46,000 individuals in the sample and attempted to follow up at least once with unfinished respondents.

In addition to the follow up protocol outlined above, we employed two targeted communication strategies to improve completion rates. On November 13, we selected a random subsample of control group participants who had started but not completed the survey and made phone calls to answer any questions they had about the study and to encourage them to finish. On November 18, we sent personalized emails from a member of the research team to everyone who had completed more than 50% of the survey that mentioned their progress (either more than 50%, more than 75%, or more than 90%) with a link to resume where they left off so they could claim their gift card.

3 Increased incentives

We began offering larger incentive amounts on November 6, both to increase overall response rates and to reduce the imbalance in response rates between the treatment and control groups. No treatment group members who claimed

the cash transfer were ever offered more than the original \$10 gift card, but we started offering control group nonresponders a \$20 gift card beginning with their second reminder (3 days after they started the survey). We also offered \$20 gift cards to members of the treatment group who were not actually treated (i.e., did not claim the \$1,000 transfer from GiveDirectly) who started the survey between October 28 and November 5, but all treatment group members who started the survey on or after November 6 were offered \$10 regardless of whether or not they received the transfer.

Beginning on November 15, we experimented with offering a \$40 gift card to control group members with partially completed surveys. Individuals in the control group who started the survey between October 28 and November 11 only received the \$40 offer in the final one or two reminders, but those who started the survey on November 12-14 were offered \$40 beginning with the third reminder and those who started on November 15-16 were offered \$40 in the first reminder. Improvements in response rates with the larger incentive were minimal given the cost, however, and we opted to revert to a maximum of \$20 beginning with anyone who started the survey on or after November 17.

4 Modification to recruitment within FreshEBT

We implemented final modifications to recruitment and follow up on November 23. The push notification from FreshEBT to control group members who had not started the survey mentioned a \$20 gift card, the banner in the app was redesigned, and the control group banner offered a \$20 gift card for participation. All reminders to control group members who started the survey after the new banner went live included the larger incentive.

Appendix 2: Survey Instrument

Demographics and Basic Background [Census pulse survey/DMACS modified]

To begin, we would like to ask you about your household.

1. Please indicate who currently lives in your household: (Select all that apply, don't allow g to be selected and another; do not select both spouse and partner)
 - a) My spouse
 - b) My romantic partner
 - c) My child(ren) under age 18
 - d) Other children under age 18
 - e) Other adult family
 - f) Other unrelated adults
 - g) No one, I live alone

[DMACS, modified]

2. What is your current employment status? *Please select all that apply.*
[do not allow d and e to both be selected]
 - a) Employed full-time
 - b) Employed part-time
 - c) On leave or vacation
 - d) Temporarily laid off
 - e) Not employed but looking for work
 - f) Not employed, not looking for work
 - g) Retired
 - h) Other

IF C, D, E OR F ASK

3. When did you last regularly work for pay (like a job that lasted two weeks or more)? Month, YEAR

ASK ALL

[CPS, *modified*]

4. How many hours did you work last week at all jobs? *Please include all work you did last week, including freelance work or other informal jobs.*
[Prefill with zero for those who do not work]
5. In the last 30 days, have you looked for a new job or additional work?
 - a) Yes
 - b) No

IF YES LOOKED FOR WORK THEN ASK:

6. What have you done to find work in the last 30 days? *Please select all that apply.*
- a) Looked at job postings
 - b) Contacted someone about a job (an employment agency, employer, friend)
 - c) Applied to a job posting
 - d) Had a job interview
 - e) Posted a resume online or posted/updated information on a career networking website
 - f) Other

ALL RESPONDENTS

[RH H1, modified to assess earnings of household]

7. What was your household's total income from earnings (before tax) last month? (Your best estimate is fine. Please include wages, salary, tips, bonuses, commissions, self-employment and other income from work. Include your earnings and the earnings of other household members).
- a) No income from earnings
 - b) Less than \$200
 - c) \$200-499
 - d) \$500-999
 - e) \$1,000-1,999
 - f) \$2,000-2,999
 - g) \$3,000-3,999
 - h) \$4000 or more
8. Please tell us specifically how much your household income was last month. *Your best guess is fine.*

[Census pulse survey, modified]

9. In 2019 what was your total **household** income before taxes (please also include income from other members of your household)?
- a) Less than \$10,000
 - b) \$10,000-\$14,999
 - c) \$15,000-\$19,999
 - d) \$20,000-\$24,999
 - e) \$25,000-\$34,999
 - f) \$35,000-\$49,999
 - g) \$50,000-\$75,999
 - h) \$75,000-\$99,999
 - i) \$100,000 and above

[RH and SHED, modified]

10. In the past 30 days (that is, since late September) have you or anyone in your household received any of the following? (Select all that apply).

	Yes	No
a) Unemployment Insurance/Unemployment benefits		
b) Cash assistance or cash welfare (like TANF)		
c) Free food/meals from a food bank or food pantry.		
d) Food assistance from school (school lunch or school breakfast delivery/pick-up)		
e) Food Stamps/Supplemental Nutrition Assistance Program (SNAP)		
f) Assistance with housing payments from a government program		
g) Retirement income, including Social Security or survivor's benefits		
h) Disability income, including SSI, SSDI, or other disability benefits		
i) Regular financial assistance from someone outside the household, including child support or alimony		

[Fresh EBT]

11. Please tell us how much money your household gets from each source in the last 30 days.

Your best guess is fine.

ONLY ASK AMOUNT IF YES IN QUESTION ABOVE.

- a) Unemployment Insurance/Unemployment benefits: _____
- b) Cash assistance/cash welfare (like TANF): _____
- c) Food Stamps/SNAP _____
- d) Social Security or survivors benefits _____
- e) Disability income, SSI, SSDI _____
- f) Regular financial assistance from someone outside the household, including child support or alimony _____

IF YES TO FOOD FROM A FOOD BANK

12. How many times did you get food from a food bank or food pantry in the last 30 days?

(enter # of times/ I can't remember).

IF YES TO FOOD STAMPS

13. Did you get on food stamps due to COVID (since March 2020)?
- a) Yes, this is my first time on food stamps (I've not been on before)
 - b) Yes, I got on due to COVID, but I've been on food stamps in the past
 - c) No, I was already on food stamps before COVID hit in March 2020

Material Hardship

We are interested in some of the problems that families face making ends meet. First, we are going to ask you about some of the bills you pay.

Please think about the most recent bills you had to pay – the bills that were due in October or in late September.

BILLS & UTILITIES

14. Thinking about your most recent bill (the one due in late September or in October):
Options: Yes; No; not applicable/I don't pay this bill
- a) Did you pay the full amount of your rent or mortgage payment?
 - b) Did you pay the full amount of your gas, oil, or electric bill?
 - c) Did you pay the full amount of your phone or internet bill?
15. In the last 30 days (since late September) were any of the following services cut off? Options: Yes, No.
- i) Your water
 - ii) Your gas/oil or electricity
 - iii) Your phone or internet bill

HOUSING

ASK IF PAY RENT BILL

16. Do you owe any back rent or mortgage (rent or mortgage you owe but have not paid)?
- d) Yes
 - e) No

IF YES HAVE BACK RENT/MORTGAGE ASK:

17. How much do you owe in back rent or mortgage? *Your best guess is fine.* _____

ELECTRIC OR GAS

ASK IF PAY GAS/ELECTRIC BILL

18. Are you behind on your gas, oil or electric bill?
- a) Yes
 - b) No

IF YES ASK BACK PAYMENT

19. How much do you owe in back gas, oil, or electric bills? *Your best guess is fine.*

PHONE

SKIP IF N/A/DO NOT PAY THIS BILL IN PHONE/INTERNET QUESTION ABOVE

20. Are you behind on your phone or internet bill?
- a) Yes
 - b) No

IF YES ASK BACK PAYMENT

21. How much do you owe in back phone or internet bills? *Your best guess is fine.*

FOOD:

Now we are going to ask you about your food situation in the last 30 days.

22. In the last 30 days (since late September), how true were each of the following statements for you or your household:

Often true, sometimes true, never true.

- a) I worried whether our food would run out before we got money to buy more.
 - b) The food that we bought just didn't last, and we didn't have money to get more.
 - c) We couldn't afford to eat balanced meals.
23. In the last 30 days, did you or other adults in your household ever cut the size of your meals or skip meals because there wasn't enough money for food?
- Yes, No

IF YES

24. In the last 30 days, how many days did this happen? _____ days

ASK ALL

25. In the last 30 days (since late September) Yes/No

- a) Did you ever eat less than you felt you should because there wasn't enough money for food?
- b) Were you every hungry but didn't eat because there wasn't enough money for food?

MEDICAL

26. In the last 30 days (since late September), did any of the following things happen to you or someone in your household?

Responses: yes; no

- a) Someone needed to see a doctor or go to the hospital but could not because of the cost?
- b) Someone needed to get a prescription filled but could not because of the cost?
- c) Someone needed to go to the dentist but could not because of the cost?

27. Does your household owe money for a medical or hospital bill?

- a) Yes
- b) No

ASK IF YES OWE MEDICAL BILLS

28. How much do you owe in medical bills? *Your best guess is fine.*

HOUSING

ASK ALL RESPONDENTS

29. In the last 30 days (since late September), have any of the following things happened to you, even for one night:

Responses: Yes/No

- a) You moved in with other people because of financial problems?
- b) You stayed in a shelter?
- c) You stayed in another place not meant for regular housing like an abandoned building or an automobile?
- d) You were evicted or your landlord forced you to leave your home or apartment for not paying the rent or mortgage?

TRANSPORTATION SECURITY

30. In the last 30 days (since late September), were you forced to miss an appointment, skip going somewhere, or miss work because you did not have a way to get there?

Yes, No

31. Do you own a car?

- a) Yes
- b) No

IF YES TO CAR

32. Are you behind on your car payments?

- a) Yes
- b) No
- c) I do not have car payments

IF YES TO CAR PAYMENTS

33. How much do you owe in back car payments? *Your best guess is fine.* _____

Mental Health

ASK ALL RESPONDENTS

[PHQ-2/ GAD-2]

34. In the past 2 weeks, how often have you been bothered by the following problems? Responses: Not at all, several days, more than half the days, nearly every day, prefer not to say

- a) Little interest or pleasure in doing things
- b) Feeling down, depressed or hopeless
- c) Feeling nervous, anxious or on edge
- d) Not being able to stop or control worrying [SOS-S

Item]

35. During the past two weeks, have you felt:
Not at all, a little bit, somewhat, quite a bit, a lot

- a) ...inadequate?
- b) ...swamped by your responsibilities?
- c) ...that the odds were against you?
- d) ...that there wasn't enough time to get to everything?
- e) ...like nothing was going right?
- f) ...like you were rushed?
- g) ...like there was no escape?
- h) ...like things kept piling up?
- i) ...like just giving up?
- j) ...like you were carrying a heavy load? [SSS-8

Somatic Symptom Scale]

36. During the past 2 weeks, how much have you been bothered by any of the following problems?

Not at all, a little bit, somewhat, quite a bit, very much

- a) Stomach or bowel problems (nausea, indigestion, constipation, diarrhea)
- b) Back pain
- c) Pain in your arms, legs or joints
- d) Headaches
- e) Chest pain or shortness of breath
- f) Dizziness
- g) Feeling tired or having low energy
- h) Trouble sleeping

Consumption/Spending

[YCR, DMACs, Covid 19, Johannes – amalgamation, modified]

37. We would like to know more about how households have been spending their money. Did your household spend any money in the last 30 days (since late September) on the following items? If so, how much?

(Prefill with zero \$\$ so that they can skip as needed).

	\$ Amount
Food at home (like groceries, excluding your EBT benefits)	
Food outside of home (take out, restaurants)	
Rent or mortgage (include any late fees paid)	
Phone, internet or cable	
Utilities (gas, electric, water, etc.)	
Transportation (gas, parking, tolls, public transit, taxis or car services)	
Vehicle purchase or repairs	
Childcare (daycare, nanny, babysitter)	
Other child expenses (like clothes or toys)	
Clothing/shoes for adults or other personal care item (like shampoo, cosmetics, laundry)	
Entertainment or recreation (Netflix, magazines)	
Household repairs or maintenance	
Other household items like appliances, furniture or electronics	
Your education (books, tuition, excluding loan repayments)	
Your children's education (books, tuition)	
Give financial loans or gifts to family or friends	
Deposit money into a savings or retirement account	
Medical expenses (including healthcare premiums, medications, doctors bills, etc.)	

LOANS/OTHER DEBT

38. In the last 30 days, did you do any of the following because there wasn't enough money to make ends meet?

Yes; No

- a. Take out a new loan from friends or family
- b. Take out a new loan from a private company (e.g., payday, title, bank)?

39. Do you currently owe friends or family money?

- a) Yes
- b) No

IF YES

40. How much money do you owe friends or family? *Your best guess is fine.*

ASK ALL RESPONDENTS

41. Do you have any other loans (payday, title, bank, student)?

- a) Yes
- b) No

IF YES

42. Are you behind on your loan payments (payday, title, bank, student)?

- a) Yes
- b) No

IF YES TO BEHIND ON LOAN

43. How much money do you owe in back loan payments? *Your best guess is fine.*

ASK ALL RESPONDENTS

44. Do you have a credit card?

- a) Yes
- b) No

IF YES

45. Do you currently have any outstanding unpaid credit card debt (money that is back owed/or a balance you have been carrying)?

- a) Yes
- b) No

If YES

46. How much credit card debt do you have? *Your best guess is fine.*

Other financial wellbeing measures

ASK ALL RESPONDENTS

47. Do you own a computer, tablet or laptop?
a) Yes
b) No

[YCR, heavily modified] ASK
ALL RESPONDENTS

48. In the last 30 days (since late September), did you cut back on spending by doing without to make ends meet?
Yes
No
49. In the last 30 days, did you sell something you own in order to make ends meet? Yes
No

IF YES

50. In the last 30 days, how many times did you sell something you own in order to make ends meet?
_____ (prefill w/ zero; 0-30 range)

ASK IF Question above is >0

51. How much money did you make from selling things you own in the last 30 days? *Your best guess is fine.*

Child wellbeing

ASK ONLY IF LIVE WITH AT LEAST ONE CHILD.

[FFCWS heavily modified, based on Achenbach CBCL]

52. These questions are about your child(ren) and how they have been behaving over the past two weeks. How true are these statements?
Responses: very true or often true, sometimes or somewhat true, not true, prefer not to say
- a) They fight
 - b) They can't concentrate or can't pay attention for long
 - c) They are nice to others
 - d) They argue
 - e) They destroy or break things
 - f) They are helpful

- g) They are disobedient (or don't follow rules)
- h) They are confused or seem to be in a fog
- i) They are unhappy, sad or depressed
- j) They are withdrawn or unusually quiet
- k) They are having trouble sleeping

Parenting

ASK ONLY IF LIVE WITH AT LEAST ONE CHILD.

[FFCWS, modified]

53. Sometimes children behave pretty well and sometimes they don't. In the past two weeks, how often have the following things happened?

Responses: Not at all, several days, more than half the days, nearly every day

- a) I lost my temper, got angry or yelled at my child(ren)
- b) I hit or spanked my children
- c) I spent time teaching my children or helping my children with schoolwork
- d) I spent time with my children talking, reading, listening to music, playing a game, watching TV, or doing another activity (like art or cooking).

Relationship with Partner

ASK IF RESPOND THAT PARTNER OR SPOUSE LIVES IN THE HOUSEHOLD

[FFCWS modified from Conflict Tactics Scale CTS]

Now we would like to ask you some questions about your relationship with your spouse or partner, such as a boyfriend or girlfriend.

54. How often does your spouse/partner behave in the following ways (often, sometimes, never)?

Responses: Often, sometimes, never

- a. Insults or criticizes you or your ideas.
- b. Encourages you or helps you do things that are important to you
- c. Tries to keep you from talking with your friends or family
- d. Listens to you when you need someone to talk to
- e. Hits, slaps, kicks or otherwise physically hurts you
- f. Forces you to have sexual activities

Other Demographics

To finish up, we just have a few more questions about you and your family. [Census pulse survey, modified]

55. How many adults, including yourself, currently live in your household? (Please enter a number: range 1-20)

56. How many children (under 18) are currently living in your household?

- a) Under the age of 6 (range 0-7)?
- b) Between ages 6-12 (range 0-7)?
- c) Between ages 13-18 (range 0-7)?

[Census pulse survey]

57. Are you...

- a) Male
- b) Female
- c) Other

[Census pulse survey]

58. What year were you born? (Select year: range 1930-2002) [Census pulse survey, modified]

59. Are you of Hispanic or Latino(a) origin or descent?

- a) Yes
- b) No

[Census pulse survey, modified]

60. What is your race? (Check all that apply)

- a) White
- b) Black or African American
- c) American Indian or Alaska Native
- d) Asian or Pacific Islander
- e) Other

[Census pulse survey]

61. What is the highest degree or level of school you have completed?

- a) Less than high school
- b) Some high school
- c) High school graduate or equivalent (for example GED)
- d) Some college, but degree not received or is in progress
- e) Associate's degree (AA, AS)
- f) Bachelor's degree (BA, BS, AB)
- g) Graduate degree (Masters, professional, doctorate)

ASK ONLY IF THEY GOT THE MONEY FROM GIVE DIRECTLY [TREAT ==1]

[FRESH EBT]

62. Please tell us how you used the \$1,000 from Give Directly/Hyperwallet/FreshEBT. *Check all that apply.*
- a. Pay bills
 - b. Pay rent
 - c. Buy food
 - d. Home repairs or items
 - e. Buy children's clothing
 - f. Buy other things for children (toys, books)
 - g. Put money into a savings account
 - h. Pay for health care
 - i. Pay loans
 - j. Buy other necessities
 - k. Pay for entertainment
 - l. Other _____

Your response has been recorded. Thank you for your participation!

Check the box below to receive your compensation for participating in the survey.

The University of Michigan research team intends to conduct additional surveys in the future. If you do not wish to be contacted about future surveys, please let us know by calling XXX or sending an email to XXX with "opt out of surveys" in the subject line.

If you would like to speak with someone about any of the sensitive topics discussed in this survey, please contact the following resources:

National Child Abuse Hotline

1-800-4-A-CHILD (1-800-422-4453)
<https://www.childhelp.org/childhelp-hotline/>

The National Domestic Violence Hotline

1-800-799-7233 (SAFE)
www.ndvh.org

National Suicide Prevention Lifeline

1-800-273-8255 (TALK)
www.suicidepreventionlifeline.org

Appendix 3: Crosswalk of Material Hardship Index Between RCT1 and RCT2

	RCT1	RCT2	Notes – to make RCT2 mimic RCT1
<i>General notes</i>	<i>Not applicable is not an option but in 3mo we do ask if they pay that bill.</i>	<i>Not applicable/don't pay this bill is an option throughout. I think they were set to missing.</i>	
Rent	Did you not pay the full amount of rent or mortgage payments?	Did you pay the full amount of your rent or mortgage payment?	Must be reverse coded
Bills/utilities	Did you not pay the full amount of any utilities bill (e.g., water, gas, oil, or electricity)?	Did you pay the full amount of your gas, oil, or electric bill?	Must be reverse coded <ul style="list-style-type: none"> Water not included in RCT2 as often included in rent
	Did you not pay the full amount of my phone, internet and/or data plan?	Did you pay the full amount of your phone or internet bill?	Must be reverse coded <ul style="list-style-type: none"> Data plan not included in RCT2 per advice from experts
	Was your water, gas, oil, electricity, internet or phone service cut off?	In the last 30 days (since late September) were any of the following services cut off? <ol style="list-style-type: none"> Your water Your gas/oil or electricity Your phone or internet bill 	Can be combined into a single question. Yes on any.
Food	I worried whether our food would run out before we got money to buy more. Yes/No	I worried whether our food would run out before we got money to buy more. Often, sometimes, never true	Recode often/sometimes into "yes". Not exactly analogous but should be close
	The food we bought just didn't last and we didn't have money to get more. Yes/No	The food that we bought just didn't last, and we didn't have money to get more. Often, sometimes, never true	Recode often/sometimes into "yes". Not exactly analogous but should be close
Medical	Was there anyone in your household who needed to see a doctor, go to the hospital or get a prescription filled but couldn't because of the cost?	In the last 30 days (since late September), did any of the following things happen to you or someone in your household? <ol style="list-style-type: none"> Someone needed to see a doctor or go to the hospital but could not because of the cost? 	Can be combined into a single question. <i>Note: in RCT2 we also asked a separate question re: dentist not asked in RCT1.</i>

		2. Someone needed to get a prescription filled but could not because of the cost?	
Housing	Did you move in with others, stay in a shelter or get evicted because of financial problems?	In the last 30 days (since late September), have any of the following things happened to you, even for one night: <ol style="list-style-type: none"> 1. You moved in with other people because of financial problems? 2. You stayed in a shelter? 3. You were evicted or your landlord forced you to leave your home or apartment for not paying the rent or mortgage? 	Can be combined into a single question. <i>Note: In RCT2 also asked: You stayed in another place not meant for regular housing like an abandoned building or an automobile? This wasn't in RCT1 so should be excluded but is commonly included in hardship batteries</i>
Transportation	Were you forced to miss an appointment, skip going somewhere, or miss work because you did not have a way to get there?	Were you forced to miss an appointment, skip going somewhere, or miss work because you did not have a way to get there?	Same question

Other differences:

In RCT1 we asked all the hardship question in the same battery of questions, whereas in RCT2 we asked more follow up questions pertaining to particular topics (like back rent due) so the hardship questions, although they were sequentially asked were not all in one battery.

Appendix 4: Data Cleaning

To clean and construct variables for analysis, we first combine raw survey response data with Fresh EBT application data provided by Propel. We define our analysis sample as all users who consented to participate in research, provided contact information, and responded to enough of our survey to have non-missing primary outcomes. Next, we clean and construct variables used in analysis.

1 Control Variables

We create variables equal to a respondent's number of children that are under six years old, ages 6 to 12, and ages 13 to 18. To capture household composition, we create indicators if respondents said they lived with a spouse, a cohabiting partner, their own children, the children of others, relatives, or unrelated adults. In addition, we create variables equal to the total number of children, total household size, and missing response indicators for all household composition variables. We top-code total household size at 15 members. To measure racial and ethnic identity of respondents, we construct a set of mutually exclusive indicators for the following subgroups: White non-Hispanic, Black non-Hispanic, Hispanic, other racial groups, and missing race/ethnicity response. We also create an indicator equal to one if respondents identify as men, women or other genders, as well as a missing response indicator. To control for a respondent's age, we combine information from the Fresh EBT application and our survey – we set a respondent's age equal to the age reported on our survey, and replace it with the application data if a user did not respond to the survey question. In addition, we replace a respondent's age with the application data if the survey-reported age differs by more than five years, and the user is younger than 21 or older than 55. Finally, to control for educational status, we create five mutually exclusive categories: less than a high school diploma, high school diploma, some college, associates degree or higher, and missing education status.

Before running regression models, we set all variables to zero when their corresponding missing response indicator is equal to one, i.e. for age, gender, race, education, and household size/composition. All variables included in regressions thus have balanced sample sizes. The following table compares variables used as controls in RCT-1, controls in RCT-2, and to randomize RCT-2 (used for re-randomization inference).

2 Other Characteristics

Because our data about employment, income and benefit receipt was collected four to six weeks after our intervention, we construct the following measures to be used in exploratory analysis. For employment, we create a set of mutually exclusive indicators. Respondents could mark multiple employment statuses on the survey, so we impose a hierarchy such that each respondent has one primary employment status. Any indication of full time employment will override other statuses, followed by part time employment. If a respondent indicates they are retired, all other employment statuses are set to zero. Finally, if they indicate employment statuses besides being on leave or other-employment, we set these two statuses to zero. If respondents did not indicate any employment status, we create an indicator equal to one for missing employment. In our final data sample, all respondents should have one, and only one, employment status, or have a non-zero missing status indicator.

For our measure of monthly earnings, we asked respondents to indicate an interval for their earnings, and report a level of earnings if desired. We midpoint code the earnings bins, e.g. reporting earnings between 200-500 dollars is coded at 349.50. For those that reported a level, we create an alternative earnings measure that replaces bins with non-zero dollar amounts. We create an indicator equal to one for all users that reported monthly income less than 500 dollars. In addition, we compute per-capita earnings by dividing reported monthly earnings by total household size.

Table 1: Variables used for randomization and controls across experiments

	RCT2 Randomization (1)	RCT2 Controls (2)	RCT1 Controls (3)
Geographic Aggregates			
Zipcode PCT BA or More	*	*	
Zipcode Median HH Income	*	*	
Zipcode Median Rent	*	*	
Zipcode PCT Below FPL	*	*	
Zipcode Pop Density	*	*	
Zipcode PCT Black	*	*	
Zipcode PCT Asian	*	*	
Zipcode PCT Hispanic	*	*	
Metro Area	*	*	
Micropolitan Area	*	*	
Commuting Zone	*	*	
Small Rural Area	*	*	
Missing RUCA	*	*	
Per capita COVID deaths	*	*	
Fresh EBT Administrative Data			
Number of Days Using App	*	*	*
Had Account Before 2/2020	*	*	*
Made Account Day of RCT1			*
Number of App Uses	*	*	*
FEBT SNAP Benefits	*	*	*
Missing FEBT SNAP Benefits	*	*	*
Spanish Language	*	*	*
State Fixed Effects	*	*	*
Treatment Invariant Demographics			
Hispanic		*	*
White Non-Hispanic		*	*
Black Non-Hispanic		*	*
Other Race		*	*
Female		*	*
Missing Sex		*	*
Less than HS Diploma		*	*
HS Diploma		*	*
Some College		*	*
Associates Degree or More		*	*
Age		*	*
Household Size		*	*
Missing Household Size		*	*
Number of Kids Under 6		*	*
Number of Kids 6-12		*	*
Number of Kids 13-18		*	*
Total Number of Kids		*	*
Lives with Spouse		*	*
Lives with Unmarried Partner		*	*
Lives with Own Children		*	*
Lives with Children of Others		*	*
Lives with Related Adults		*	*
Lives with Unrelated Adults		*	*
Economic Characteristics Collected in RCT1 Baseline Survey			
Indicator if Finished Baseline Survey			*
COVID Hardship Screener Questions			*
Full time			*
Part time			*
Temporarily Laid Off			*
Unemployed			*
Other Employment			*
Use WIC			*
Use Housing Benefits			*
Use UI			*
Use TANF			*
Missing Benefits info at baseline			*

We also prepare measures using administrative data from Give Directly about when cash transfers were sent to Hyperwallet accounts, and when users cashed out their transfers if eligible. We calculate the number of days between cashing out a transfer and taking our survey, and set the number of days equal to zero if a user took our survey before cashing out their transfer. We also set our instrumental variable for cashing out a transfer to zero in these cases.

3 Benefit and Spending Variables

In our survey, we ask for detailed information about dollar amounts of public benefits received, spending, and debt in order to approximate total monthly resources and liquidity. To clean these variables, we apply a similar approach. We first replace all negative dollar amounts with zero, then top-code each raw dollar amount at the 99th percentile of the in-sample distribution. If a respondent said they did not receive income, spend money, or have debt in a category, we replace the dollar amount with zero. For all categories, a portion of respondents indicated that they spent or earned, but did not list a dollar amount. In these cases, we predict dollar amounts based on observable characteristics of non-missing dollar amount respondents. We restrict prediction samples to respondents living in zip-codes with fully populated 2018 ACS data. We replace any negative predictions with zero, and top code predictions at the maximum value of reported dollar amount data. For each category, we create flag variables to indicate which amounts were imputed.

To approximate total monthly resources, we create measures that combine reported earnings and benefit dollar amounts with the imputed amounts. We then sum earnings, benefits for unemployment insurance, TANF, SNAP, Social Security, Disability Insurance, and regular financial assistance such as child support. We construct per-capita monthly resources by dividing by total household size.

In order to measure changes in consumption, we asked respondents to report spending on items in 18 different categories that were based on questions asked in the Survey of Household Economics and Decision-making (SHED) as well as prior research. These categories include rent, utilities, transportation, childcare, food and education. See full survey in Appendix 2 for more details. Unfortunately, the low quality of the responses we obtained preclude us from saying much about the impact of the intervention on consumption. Nearly 20 % of respondents indicated spending nothing at all in the prior month, and 30 percent of respondents indicate that they spent money in three or fewer categories in the prior month, which seems implausible given the 18 potential categories, many of which involve expenditures on basic necessities. Focusing on the 33 percent of our sample that reported spending on at least four categories and provided an estimate of expenditures, the mean (standard deviation) of all expenditures and savings is \$1389 (\$1399). However, we find no significant difference between the treatment and control groups in terms of reported spending, which we define to include savings as well as expenditures. This leads us to conclude that the expenditure responses do not provide reliable information.¹

¹Assuming we believe that the randomization was valid and that the treatment individuals actually received the \$1000 cash transfer, it must be the case that these spending measures are not reliable. When we regress our spending measure on either monthly income or total monthly financial resources (only among the control group with valid spending data), we obtain a coefficient of roughly 0.2. If our spending measures truly capture all expenditures and savings, the coefficient should be 1 by construction.

Appendix 5: Recruitment and Enrollment Processes in RCT-1 and RCT-2 Processes

The procedures used to recruit, enroll and survey individuals differed across the two studies and may have resulted in some hard-to-observe differences between the composition of the studies. The following table provides a summary of the differences in recruitment and survey strategies across the two studies, which we describe in detail below.

Box 2: Recruitment and Survey Strategies for RCT-1 and RCT-2		
	<i>RCT-1</i>	<i>RCT-2</i>
<i>Recruitment approach</i>	<p>App segments (each segment looks identical) were assigned to treatment and control.</p> <p>In app banner opened to both treatment and control groups inviting them to take a short survey about COVID-19.</p> <p>At the end of the brief survey both groups were invited to participate in a study by the University of Michigan (U of M) and they provided an email address and phone number.</p> <p>Control group was thanked and the treatment group was sent to sign up for the \$1000.</p> <p>Treatment group received the cash regardless of their willingness to participate in the U of M study.</p> <p>All treatment and control members were recruited in a few hours.</p>	<p>The U of M team took a sample of Fresh EBT users and randomly assigned them to treatment and control users. Stratified by whether the individual used the app prior to February 2020 (before the start of COVID) and a 3-category measure of social welfare generosity measured by the maximum TANF benefit for a three-person household in a state).</p> <p>GiveDirectly invited a cohort of treatment users through a banner placed in the Fresh EBT app with an offer to get \$1000.</p> <p>Treatment members were enrolled over a period of 10 days (9/14/2020-9/24/2020).</p> <p>Notifications were sent to encourage treatment users to claim the cash.</p>
<i>Survey</i>	<p>Treatment and control groups were sent a baseline survey the evening they agreed to participate in the study.</p> <p>First follow-up survey was administered 4 weeks after the cash transfer and was open for 9 days (response rate 60%).</p> <p>The second follow-up survey was sent 3 months post cash transfer and was open 14 days (response rate 57%).</p>	<p>A banner in the Fresh EBT app was opened to both treatment and control groups on October 28, 2020, approximately 6 weeks after the cash transfer.</p> <p>Users who clicked on the banner were directed to a consent form, first providing contact information and then taking the survey.</p> <p>The survey was open for 35 days (response rate approximately 30% of all those who logged in during the intervention period - Sept 14-24).</p>
<i>Timing</i>	May 2020	September/October 2020
<i>Sample sizes</i>	9433 one-month follow up (5365 Treatment, 4068 Control)	13,755 (7035 Treatment, 6720 Control)

1 RCT-1

In RCT-1, a very large (approximately 400,000) randomly selected set of (treatment group) users were shown an in-app banner inviting them to take a survey about the impact of the coronavirus on their family. Another large set of

randomly selected users (control group) were shown an identical banner at the same time. While nothing in the banner mentioned the cash award explicitly, the GD program had received considerable media attention and at least some app users were actively looking for such banners with the expectation that they would lead to cash awards.

Both treatment and control users who clicked on the banner were asked seven identical screener questions, including one that allowed us to identify families with children. If users indicated that they had children and no SSI/SSDI benefits, individuals were asked whether they would like to participate in a research study about how the coronavirus pandemic is affecting families. After indicating their interest and providing contact information, individuals in the control group were simply thanked while individuals in the treatment group were given the opportunity to receive a \$1,000 cash transfer from GD.¹ One month later all study participants were invited to complete a survey whose purpose was described as assessing the social and economic well-being of families during the pandemic.

Given this approach to recruitment and enrollment, individuals in RCT-1 were not only those most likely to be actively viewing the app, but also those likely to click on the banner. The banners first appeared on user screens at 7am EST on May 21st. After only 5 hours, enough treatment group users had clicked the banner and enrolled in the GD program that all of the funding was spent.² Participants in RCT-1 might be described as “eager beavers” by which we mean either eager to receive the cash transfer (if they were aware of the GD program) or eager to complete a research survey. The take-up rate for treatment group individuals was very high by construction. Only 2.5% of treatment group individuals who made it through the screener questions and consented to participate in the research study did not end up receiving the cash transfer because they were unwilling or unable to complete the GD enrollment forms. Moreover, roughly 60% of those who agreed to participate in the study responded to the one-month follow-up and were thus included in the RCT-1 analysis sample.

2 RCT-2

In contrast, we randomly assigned a much smaller group of individuals to the treatment and control groups in RCT-2 (recall the study sample was roughly 46,000), with the goal of including as many users in the study as possible and thus increasing the external validity of the analysis. As described earlier, enrollment for the intervention proceeded much more slowly in RCT-2. It took 10 days and a variety of email reminders and push notifications to recruit a sufficient number of households to fully spend the funds allocated for intervention. Note that there was no mention of the research study during this enrollment period for the treatment individuals. Users we randomly assigned to the control group did not see any banner during this period.

Relative to RCT-1, we suspect that the enrollment process in RCT-2 resulted in less eager participants. Only 41.2% of those assigned to the treatment group in RCT-2 claimed the cash award, as opposed to 97.5% of those assigned to the treatment group in RCT-1. Some treatment individuals never clicked on the banner to read about the opportunity, and others did click the banner but did not follow through to obtain the transfer. It appears that most individuals in this latter group were dissuaded by the waivers and other documents that they needed to sign in order to claim the award. This type of individual probably does not appear in the RCT-1 sample since individuals had to engage sufficiently with the app to consent to participate in the research study and provide their contact information. Consistent with these sample differences, only 30% of study participants responded to our survey in the second experiment compared with almost 60% in the first experiment.³

¹Neither eligibility for the study nor willingness to share contact information with researchers affected receipt of the cash transfer. As soon as individuals completed the enrollment form and agreed to GiveDirectly’s terms, the money was transferred in as little as 24 hours to a Hyperwallet account. Recipients used the Hyperwallet platform to choose how to receive their payments; options included PayPal, cash pickup, virtual or physical prepaid cards, ACH bank account transfer, or paper check.

²At this point, the “recruitment banners” were taken down for all remaining treatment and control users. Fresh EBT did not actively contact users in any way during this period.

³However, this may in part be due to the difference in design - in RCT-2 the cash transfer was not in any way connected to the research study unlike in RCT-1.