UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ESSAYS IN DEVELOPMENT ECONOMICS

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Naresh Kumar

March 2023

The Dissertation of Naresh Kumar
is approved:

Professor Jonathan Robinson, Chair

Professor Alan Spearot

Professor Ajay Shenoy

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Essays in Development Economics

by

Naresh Kumar

This dissertation contains three essays broadly related to program evaluation using randomized control trial in developing countries.

In Chapter 1, joint work with David Sungho Park, we evaluate the impact of a multifaceted female empowerment program on reducing intimate partner violence (IPV) in urban Liberia. We ran a randomized controlled trial (RCT) in partnership with the Liberian Red Cross. The program intervention includes intensive psychosocial therapy and vocational skills training throughout a full year. About 12 months after program completion, we find the program significantly reduced the proportion of women who experienced emotional, physical, and sexual IPV by 10-26 percentage points (from control bases of 24-62 percent). While there are multiple pathways through which IPV could be impacted, one channel is that the business training was highly effective: labor supply increased by 37 percent and expenditure by 49 percent. One focus of the program is psychological empowerment, and we find positive but statistically insignificant effects on distress and happiness indices. We also find improvements in social norms around IPV: perceived justifiability of IPV reduced by 0.3 standard deviations.

In Chapter 2, joint work with Shilpa Aggarwal, Dahyeon Jeong, David Sungho Park, Jonathan Robinson and Alan Spearot, we study the dynamic effects of large, unconditional cash transfers in rural Liberia and Malawi using bi-monthly surveys. We document improvements in food security until the end of surveying (about a year in Liberia and two in Malawi), but find a short-lived effect on food expenditures and no effect on non-agricultural income at any point. Increased productive

investments appear to drive increased food security. After 18-25 months, we also document improvements in IPV, psychological well-being, and resilience, as well as investment and agricultural output. We find no evidence of effects on local prices or of spillovers to untreated households.

In Chapter 3, joint work with Shilpa Aggarwal, Dahyeon Jeong, David Sungho Park, Jonathan Robinson and Alan Spearot, we quantify effects of survey fatigue by randomizing the order of questions in 2–3 hour-long in-person surveys. An additional hour of survey time increases the probability that a respondent skips a question by 10%–64%. Because skips are more common, the total monetary value of aggregated categories such as assets or expenditures declines as the survey goes on, and this effect is sizeable for some categories: for example, an extra hour of survey time lowers food expenditures by 25%. We find similar effect sizes within phone surveys in which respondents were already familiar with questions, suggesting that cognitive burden may be a key driver of survey fatigue.

To my mother, in loving memory.

Acknowledgments

This dissertation would not have been possible without the support and guidance of many people. First, I would like to express my sincere gratitiude to my advisor, Jonathan Robinson, for his continuous support throughout the PhD study and research. I consider myself fortunate to have him as my soource of inspiration in my academic life. I would not be here today without his guidance. Jon's passion and commitment to developmental research was intrumental in shaping my decision to pursue development economics. His experience, knowledge, and practical approach have taught me invaluable lessons in becoming a researcher. Discussions with him always challanged me to push my intellectual boundaries and shaped my critical thinking. I would also like to thank my dissertation committee members, Alan Spearot, and Ajay Shenoy for their continuous support and invaluable guidance. Talking to them helped me to get clarity on my resaerch. Their consistent guidance helped me in refining research skills. I am thankful to my coauthors: Shilpa Aggarwal, Jenny Aker, Dahyeong Jeong, and David Sungho Park.

I would also like to thank economics faculty members at UCSC who attended my seminars and offered constructive feedback on my research. I am grateful to many profersor from the UCSC Economics Department: Brenda Samaniego de la Parra,, Carlos Dobkin, George Bulman, Jeremy West, Kenneth Kletzer, Laura Guiliano, Natalia Lazzati, and Nirvikar Singh for their support and encouragement in research and teaching. Beside faculty, I would also like to thank Susan Leach, Sandra Reebie, Meenoo Kohli, and other Economics Department Staff for all administrative support throughout the PhD. I want to thank staff at Red Cross Liberia, IPA staff in Liberia and Malawi, and enumerators who worked extremely hard, collecting data in remote areas.

My academic journey has been shaped by many people who inspired and

touched my life: my friends, teachers at school, professors at college. I would like to thank them all for their support and encouragement. I also thank my UCSC friends: Prateek Arora, Dongwan Choo, Mai Hakamada, Luka Kocic, Zhaoqi Wang, Liwei Liu, Yanshuo Chen, Ke Ma ("Amy"), Gregory Klevans, Anirban Sanyal, Bhavyaa Sharma, Rolly Kapoor, Monica Shandal, and Guanghong Xu. I also thank Aditya Bhattacharya, Avirudh Kaushik, Daniel Paul Peña, Shally Saini, Balaji Sundararaman, and Subash Saladi for treating me like family and I could always ask for help. I also like to thank my colleagues and mentors from International Monetary Fund, Rahul Anand, Paul Cashin, Sanjaya Panth, Sonali Das, and Sudip Mohapatra for their continuous support. Finally, my deep gratitude and love to my famly: My father, Hari Kishan, my brothers Pawan and Rajesh, and my sister Neelam.

I dedicate this dissertation to the loving memory of my late mother Dhanno Devi, and my late brother Pawan Sharma.

Chapter 1

Reducing Intimate Partner Violence: Evidence from a Multifaceted Female Empowerment Program in Urban Liberia

1.1 Introduction

Intimate partner violence (IPV) is a serious public health problem which affects hundreds of millions of women globally. Worldwide, one in three women has experienced some form of physical or sexual IPV in their lifetime (WHO 2021; Devries et al. 2013b). IPV is associated with many negative physical (Smith et al. 2017) and mental (L. J. Bacchus et al. 2018) health outcomes.[1] Moreover, IPV inflicts considerable economic costs on both survivors and society (C. Peterson et al. 2018).

There have been many policy discussions around how to effectively prevent or respond to IPV, and public health professionals recommend that a problem like IPV be targeted in multiple directions at the same time (Ranganathan et al. 2021). This is because IPV is a complex problem caused by a variety of psychological, social, and economic factors. The public health literature on IPV has been centered around the "ecological" framework (Heise 1998), where violence is conceptualized by an interaction of individual, interpersonal, and sociocultural factors. There is no single cause of violence, thus both IPV prevention and response require an intervention that addresses multiple underlying drivers.[2]

To study the effectiveness of a holistic approach to reducing IPV, we partner with the Liberian Red Cross to conduct a randomized controlled trial of a multifaceted female empowerment program in Monrovia, Liberia. The baseline prevalence of IPV is very high in Liberia. In the most recent Liberia Demographic

[1] According to the U.S. Centers for Disease Control and Prevention (CDC), about 35% of female IPV survivors experience some form of physical injury related to IPV (Smith et al. 2017). In our study sample, about 25% of physical/sexual IPV survivors report a physical injury as a direct effect of the male partner's action of IPV.

[2] A "prevention" intervention is both to prevent violence for individuals who experienced violence earlier and to reduce the reoccurence of violence for those who already have. Note the difference from a "response" intervention, which targets at reducing revictimization of a survivor or recidivism of a perpetrator (Mary Ellsberg et al. 2015).

and Health Survey (DHS) 2019-2020, 35 percent of partnered women of age 15-49 reported to have experienced physical or sexual IPV in the 12 past months. This is particularly high even compared to other African countries, a geographic region which itself is notorious for high prevalence of IPV (about 26% on average from countries where DHS data is available). There could be many explanations why IPV is highly prevalent in today's Liberia, including poverty (being one of the poorest countries in the world[3]). Yet one possible factor is the civil war that took place in 1989-2003, during which violence against civilian women and girls was weaponized (Omanyondo 2005). Research suggests that one of the hidden costs of such brutal civil war may be a persisting, permissive environment of violence in everyday lives (Steenkamp 2005).[4]

Since 2009, the Liberia National Red Cross Society (LNRCS) has run a female empowerment program targeted at marginalized women in informal settlements of Monrovia, where most of the internally displaced population fled for safety during the civil war. The program goal is to empower women economically and psychosocially so that they can self-sustain their lives and protect themselves from abuse. The program has two major components. The first is aimed at psychosocial empowerment, and includes daily group counseling sessions and cognitive behavior therapy focused on relationships with their spouses and other family member or community members. The second is to improve economic livelihoods through vocational skills and business training centered around helping beneficiaries set up and manage a small business. The program is very intensive: participants attend meetings 4-5 hours every day during the 12-month period. The total number of hours in the program is about 1,200, far more than most other programs.

[3]CIA World Factbook.

[4]Sub-Saharan African countries with histories of internal conflict have 11%p ($p < 0.01$) higher physical or sexual IPV prevalence than countries with those (base=21%), based on authors' country-level analysis with data from Devries et al. (2013b).

Access to the program was randomized, and treatment was stratified by baseline characteristics, including whether having experienced physical or sexual IPV past year. After conducting a baseline survey and randomizing the sample into treatment and control, the treatment group was invited to the program. While the original study design was to pool three cohorts (each including 400 women), due to COVID disruptions and related funding problems, our implementing partner Red Cross has been able to enroll only one cohort. This paper includes only one cohort of the sample with about 400 women.

The primary outcome of our study is the prevalence of IPV. To measure IPV, we administered the WHO's Violence Against Women module, which is a standardized questionnaire that has been extensively used and vetted by large-scale, multi-country surveys like the DHS. The module consists of 20 questions, each describing a specific IPV incidence (e.g., "Did your man ever slap you or throw something at you that could hurt you in the past 12 months?").[5] To construct our primary outcomes, responses to each yes/no question are indexed into a binary variable for each of the four categories: controlling behavior, emotional IPV, physical IPV, and sexual IPV.[6] In addition, for each IPV question, conditional on an affirmative response, a followup question is asked about how frequent such episode happened: (a) one or two times; (b) three to five times; or (c) more than five times. For each IPV category, we construct a summary index incorporating responses to these frequency questions.[7]

We have three main findings. First, we find that the intervention has sizable effects on IPV. Twelve months after program completion, it significantly reduces

[5]See Appendix C for full description of the IPV questionnaire.

[6]For example, Controlling Behavior Index equals to one if the respondent said yes to at least one question under the category.

[7]For each IPV categories, responses to frequency questions are standardized into a z-score using inverse covariance weighting (Michael L. Anderson 2008b).

past-year emotional IPV by 23 percentage points (from a control base of 62 percent) and physical IPV by 26 percentage points (from 45 percent in the control). The effects on sexual IPV is 10 percentage points reduction (but insignificant). The effect sizes we find are very large compared to previous findings. For example, the cash transfer literature find effect sizes of 5-11 percentage point reductions in physical IPV (Buller et al. 2018). We also asked a set of questions for norms around IPV (e.g. "Is a husband justified in hitting or beating his wife if she burns the food?") and find that the program reduced justifiability of physical or sexual IPV by 0.3 standard deviations. This provides suggestive evidence for the change in social norms as one of the explanations for IPV reduction.

Second, we find significant improvements in economic livelihoods. Monthly expenditure increased by about $12 US from a control base of $25 (or about 49 percent). While we find no significant increase in our measure of monthly income, our survey module on expenditure is more comprehensive and contains a more exhaustive list of items, so that it could be a better measure of economic welfare (Deaton 1997). We also find the program increased labor supply on self employment by about 22 hours a month from a control base of 38 hours (or about 57 percent). This is not surprising given that the focus of the business training component of the program is on self-owned business. We find modest evidence for crowding out of labor hours from other sources, and the total labor supply hours increases by 19 hours a month (insignificant) from 51 hours in control.

Third, we find positive but statistically insignificant improvements in psychological distress and happiness. To measure distress, we use the Hopkins Symptom Checklist 10-questionnaire (HSCL-10) and construct a 1-4 scale. We find the program reduced the HSCL-10 distress index by 0.02 points (insignificant) on a control base of 2.01. For happiness we construct a summary index from responses

to the Happiness and Well-being questions in the World Values Survey,[8] and we find an effect of 0.07 standard deviation (insignificant). These results are surprising, considering that one of the major components of the program intervention is psychosocial therapy.

Recently there have been a lot of impact evaluations where IPV is an outcome. The majority of these are about cash transfers, which have increased in popularity for poverty alleviation programs. The empirical evidence shows that transfers targeted to female lead to reduction in IPV (Angelucci 2008; Hidrobo and Fernald 2013; Bobonis et al. 2013; Hidrobo et al. 2016; Haushofer et al. 2019; Roy et al. 2019),[9] and these tend to show real but modest effects in the order of about 5-11 percentage points for physical IPV (Buller et al. 2018). In a companion project in rural Liberia and Malawi (Aggarwal et al. 2020a), preliminary results show unconditional cash transfers reduced proportion of women experiencing physical IPV by 2-5 percentage points (but significant only when samples are pooled).

Some studies evaluate the effect of business training programs coupled with cash transfers (Green et al. 2015; Blattman et al. 2016), but find insignificant effects on IPV.[10] While these studies are similar to ours in that they work with a marginalized population and the intervention includes business training, the intervention in our study is much more intensive. For example, about 400 hours throughout the program are spent solely on vocational skills and business training,

[8] Similarly to our frequency-integrated IPV indices, responses to each question are standardized into a z-score using inverse covariance weighting (Michael L. Anderson 2008b).

[9] Haushofer et al. (2019) find that IPV against women is reduced both when the cash transfers are targeted to the husband and the wife. Also, some studies find that the transfers to women lead to higher IPV for subgroups who face stronger social norms for gender roles (Angelucci 2008) or where women have the same as or higher education level than the men (Hidrobo and Fernald 2013), but overall there is less evidence that cash transfer programs increase IPV.

[10] Blattman et al. (2016) work with marginalized, war-affected women in Northern Uganda, and Green et al. (2015) extend the experiment by involving male partners, but either find no significant effects on IPV.

whereas in the other two studies program hours add up to about 100 hours.[11] More importantly, our intervention also includes psychosocial therapy.

In this vein, a closer study to ours is by Bandiera et al. (2020), who find that a multifaceted vocational and life skills training program to adolescent girls in Uganda decreased sex against their will, which is one form of sexual IPV. In addition to the similarities in aiming at economic empowerment, the life skills training component is similar to the psychosocial therapy in our study in that it addresses topics like conflict resolution and violence against women. However, the focus is more on sexual and reproductive health, whereas our intervention involves more intensive group counseling and cognitive behavioral therapy. The therapy sessions in our study also involve the female participants' partners and children.

This paper is also related to a growing literature on studying the effects of cognitive behavioral therapy (CBT) in developing countries. Blattman et al. (2017) find CBT coupled with $200 cash grant reduces violence committed by young men who were criminally engaged at baseline in Monrovia. Yet they find no effects on perpetrating IPV in particular. Another study in rural Kenya (Haushofer et al. 2020) finds that psychotherapy and $1,000 cash combined improve psychological wellbeing as well as economic outcomes like consumption. Instead of cash transfers, our intervention combines business training with CBT program, and we find strong evidence for improved economic livelihoods but modest effects on psychological wellbeing. This is surprising also in that the intensity of our CBT is stronger than the two other studies. The program in Blattman et al. (2017) consisted of 3 weekly sessions over 8 weeks and that in Haushofer et al. (2020)

[11] In the WINGS program evaluated by Blattman et al. (2016) and Green et al. (2015), the study sample received 4 days of training, 4-5 follow-up visits, and 3 days of self-group training (i.e., up to 96 hours total). Our intervention is unusually intensive even compared to the numerous business training programs or "graduation" programs that have been extensively tested in development economics. For example, the ILO's SIYB program (Mel et al. 2014) included training for 7 or 9 days for 7 hours a day (i.e., 49 or 63 hours total).

1 weekly session over 5 weeks, whereas our program involved 4-5 weekly sessions over 6 months. A recent paper by Barker et al. (2021) studies the standalone effect of CBT and finds significant improvements in mental health as well as downstream economic outcomes 3 months after the intervention.

The paper proceeds as follows. Section 1.2 describes the context and experiment and data collection. Section 3.3 presents the main results. Section 1.4 discusses possible threats to validity. Section 3.4 concludes.

1.2 Setting, Study Design, and Data

1.2.1 Context and Setting

This study was conducted in the capital city of Monrovia in Liberia, where IPV is highly prevalent. In the Liberia Demographic and Health Survey (DHS) 2019-2020, 35% of ever-partnered women of age 15-49 reported to have experienced physical or sexual IPV in the past 12 months, whereas the corresponding averages for Asian, Latin American and other African countries where DHS data is available are respectively 16%, 12%, and 26%. The study population targeted by the Red Cross reports much higher levels of IPV: in our baseline, we find that 51% of women report physical or sexual IPV in the past year.

There are numerous explanations for the high IPV prevalence in today's Liberia, including poverty.[12] Yet another contributing factor likely are the civil wars that took place in Liberia between 1989-1996 and 1999-2003 and killed around 250,000 people, amounting to approximately 10% of the population of the country then,

[12]Liberia is one of the poorest countries in the world (CIA World Factbook) with weak institutions, and many lack access to formal education and sustainable economic activities. For example, per one of the UN's Millennium Development Goals, the net primary education enrollment in Liberia was 37% in 2016, while the average of Sub-Saharan African countries was 78% (UNESCO Institute for Statistics).

and displaced more than another million. During the war, violence against civilians, especially women and girls, was systematically mobilized as a "weapon of war" to terrify and subdue communities. A WHO report documents that 2 in 3 Liberian women experienced sexual violence during the civil war (Omanyondo 2005).[13] Research suggests that these attitudes towards violence, once entrenched, may persist (Steenkamp 2005).[14]

1.2.2 Women Training and Integration (WIN) Program

The core intervention of this paper is a multifaceted female empowerment program called the Women Training and Integration (WIN) Program, which has been administered by the Liberian Red Cross since 2009. The program targets vulnerable women in informal settlements of Monrovia. Table 1.A2 lists the selection criteria for the WIN program. To qualify, an applicant must belong to a minimum of three groups. LNRCS has a thorough process of selecting beneficiaries. They review the application packets carefully, pay visits to the communities, and interview friends or neighbors to verify the reported information in the applications.

The program's main objective is to improve the participants' livelihoods in multiple dimensions. Specifically, the program aims at the following: 1. To economically empower women so that they can self-sustain themselves and their families; 2. To psychologically empower women so that they can better protect themselves from abuse; 3. To help establish and maintain positive relations with their families and communities; 4. To improve knowledge about and thus access to health care and psychological services.

[13] Also see Domingo et al. (2013), Jones et al. (2014), and Women (2013).

[14] Steenkamp (2005) suggests that a prolonged exposure to violence can give rise to a "culture of violence," which can be defined as "the system of norms, values, or attitudes which allow, make possible or even stimulate the use of violence to resolve any conflict or relation with another person" (Moser and Winton 2002).

The WIN program is very intensive and requires a 12-month commitment from participants, who need to be present at the WIN program center for 4-5 hours a day (either in a morning or afternoon session) for 5 days a week during the 12-month period.

The program has two major components. The first is psychosocial therapy, which includes one-to-one and group counseling sessions, thematic group discussions, cognitive behavioral therapy sessions, stress management, family/couple therapy, mediation, and conflict resolution. These aim to heal war-related trauma, reduce traumatic stress disorder, mediate family conflict situations, support coping mechanisms, build self-confidence, and promote social interaction and peaceful coexistence within their familes as well as communities.

The second is the vocational skills and business training. LNRCS offers three options for vocational skills: baking/catering, hairdressing/cosmetology, and tailoring. A participant can choose only one skill, and for those who do not have any preference, LNRCS assigns them one based on capacity constraints. The business training module provides training on handling day-to-day aspects of business, such as client interactions, sales-purchase bookkeeping, and inventory management. At the end of the program, the beneficiaries also receive business startup kits and cash grants to assist setting up their own businesses. However, due to financial constraints and COVID-related disruptions, LNRCS was not able to provide the business capital grants and cash grants for the cohort included in this paper.

The WIN program also includes several other components. The program provides routine health care check-ups and HIV/AIDS awareness and testing sessions in LNRCS's in-house clinic. Child care services are also provided when the beneficiary is at the program center. The adult literacy module targets unschooled participants and trains them in basic arithmetic, and English reading and writing

skills. The curriculum is aligned with the Ministry of Education's Alternative Learning Curriculum.

Experimental Design

The sampling frame is the pool of women who voluntarily applied to the program but selected by LNRCS through its need-based screening process. That is, our sample can be characterized by women who are disadvantaged enough for LNRCS to consider them as eligible for the program but at the same time are willing to improve their lives and have high enough agency to apply to such a program.

Several months before program start for every cohort, LNRCS advertises the program in target communities to encourage eligible women to apply. In February 2019 (for the first cohort of this study), LNRCS received about 600-700 applications in total, and after background checks and verification of the applicants' information, it shared with us a list of 450 eligible applicants divided into the "main" list of 400 and a "backup" list of 50 ranked in the order of eligibility status determined by LNRCS. In conducting the baseline survey, for those we couldn't reach after numerous attempts, we drew from the backup list in order. At the end, we enrolled 395 respondents for the study and conducted baseline in April 2019,[15] and randomly assigned 198 to treatment and 197 to control.

Treatment is stratified at two background characteristics collected in the baseline survey: (a) whether having experienced physical or sexual IPV in the past 12 months, and (b) having been affected by the civil war or having family members who have.[16]

[15] We had completed full interviews with 400 women, but LNRCS later decided to drop anyone under 17 from the sample due to potential conflict with school enrollment.

[16] Instances include: relocation, becoming disabled/amputated, family members being killed/dead.

Every woman in the treatment group was invited to the program, but some couldn't be reached or couldn't participate in the program for other reasons, and 152 women ultimately enrolled. Moreover, due to an administrative error, 2 people from the control group were invited and joined the program. For analysis, we report both intent-to-treat (ITT) and treatment-on-treated (TOT) estimates.

Our study has been significantly affected by COVID-19 disruptions. The full design was to conduct the experiment over three cohorts for about 1,200 women, each cohort including 400.[17] The first cohort of the study was enrolled in April 2019 and the program implementation ended in March 2020, right before the government lockdowns in Liberia. However, in compliance with government restrictions on in-person activities, our partner LNRCS suspended enrollment for the second cohort. While the government restrictions have been lifted since late 2020, due to financial difficulties, as of this writing, LNRCS hasn't yet been able to resume the program, and thus this paper includes only one cohort of the sample.

1.2.3 Data Collection

We conducted the baseline survey in April 2019, and the endline in April 2021, which was about 12 months after program completion. Our primary outcome is IPV but the survey also included questions on labor supply, income, expenditure, psychological well-being, social norms around IPV, transfers, and savings.

We used the WHO's Violence Against Women module[18] to measure IPV outcomes. The module consists of a group of questions each describing an IPV-related episode, providing the respondents with multiple opportunities to report violence. These binary questions are later grouped into: controlling behavior, emotional,

[17] Such pooled design was due to LNRCS's operational constraints which allow serving up to 200 beneficiaries at a time.

[18] https://www.who.int/gender/violence/who_multicountry_study/Annex3-Annex4.pdf.

physical or sexual IPV. For all questions, we restrict the recall period to the past 12 months prior to the survey date. Appendix C provides a more comprehensive description of the questionnaire.

1.2.4 IPV Measurement and Safety Protocols

We instituted WHO's ethics protocol for IPV research (WHO 2016). Study protocols have been reviewed and approved by the institutional review boards (IRBs) of the University of California, Santa Cruz and the University of Liberia, which is the relevant entity in Liberia. Second, we used the WHO's Violence Against Women module, which has been employed in multiple contexts and become a "gold standard" for IPV measurement. Third, we hired only female enumerators and provided special training both to safely conduct the interviews and to be prepared emotionally for the work. Fourth, as for the full survey itself, the survey was conducted privately without presence of anyone else than the enumerator and the respondent. Particularly for the IPV module, enumerators were trained to change questions to non-sensitive subjects in the event the survey is interrupted or eavesdropped by a third party. Fifth, while at the beginning of the whole survey respondents went through an informed consent procedure including information for the IPV, we reiterated informed consent right before the IPV module. Sixth, we prepared an information sheet that lists the services available for women experiencing IPV, including contact information for organizations where they can get help. This list was provided to every respondent who went through the IPV questionnaire, regardless of whether they reported any IPV experience.

1.2.5 Baseline Summary Statistics

Table 1.1 presents baseline summary statistics. The average age of women in the control group is about 29 years. They completed 7 years of education, on average, and about two-third of our sample have completed only primary school, while only 25% women have completed secondary school.

For the IPV questions, we restrict the sample to those who are currently partnered or have had an intimate partner 12 months prior to the survey, and the mean for this indicator at baseline was 92%.[19]

In Panel B we find that our sample had minimal access to her own income source or labor force participation. Only 11% report to have any job, and 25% are self-employed. The average income is a mere $8 dollars per month, with many zeros in the extensive margin. The mean for spouse's income is twice as large ($19). While our measures of income might not be exhaustive itself, the mean differences suggest that the women in our sample were not financially independent at baseline.

The baseline prevalence of IPV is very high. About 59% women reported having experienced emotional IPV, while the figure for the more severe form of IPV (physical or sexual) is slightly smaller (51%). This rate much higher than the national average reported in the Liberia DHS surveys, where the corresponding figures are 35% and 35% respectively in the 2019-2020 report. There could be two possible explanations. One is that our sample was selected by Red Cross in a way to be characterized as vulnerable, and one eligibility criterion was having experienced domestic abuse (Table 1.A2). Another is that the different survey tool between our baseline and Liberia DHS 2019-2020. While our study uses the

[19]We later show in Table 1.A3 that this indicator is slightly unbalanced between treatment and control at endline (statistically insignificant), and also report the Lee (2009) bounds results in Table 1.A5.

identical questionnaire to the DHS's Domestic Violence Module, at our baseline IPV was measured solely in audio computer-assisted self interviewing (ACASI), and DHS data are measured via traditional face-to-face interviewing (FTFI). In light of the findings in Section 1.4 and from our sister project in rural Liberia and Malawi (Park et al. 2021), the reported differences could be due to differing measurement modality, either through enhanced confidentiality or increased measurement error. Yet, the control group's IPV rates at our endline measured in FTFI only are still high–62% for emotional IPV, 45% for physical IPV, and 23% for sexual IPV.

Table 1.1: Baseline Summary Statistics and Randomization Check

	(1) Control Mean [SD]	(2) Treatment - Control
Panel A: Demographics		
Age	28.98	1.36*
	[7.29]	(0.73)
Years of education	7.27	0.45
	[4.11]	(0.40)
=1 if completed primary school	0.66	0.06
		(0.05)
=1 if completed secondary school	0.25	0.01
		(0.04)
=1 if currently partnered or had partner past year	0.92	-0.00
		(0.03)
Panel B: Self income and labor supply		
=1 if has own income source	0.34	0.06
		(0.05)
=1 if operated own business	0.25	0.04
		(0.04)
=1 if had any other temporary/permanent job	0.11	0.01
		(0.03)
Total income (USD)	8.38	3.36
	[27.57]	(3.09)
Panel C: Household economic well being		
Spouse's income (USD)	19.06	2.11
	[39.56]	(4.05)
Per capita expenditure (monthly, USD)	26.76	1.65
	[25.54]	(2.63)
Net value of physical assets (USD)	316.32	80.88
	[1,282.83]	(133.55)
Panel D: Intimate partner violence		
=1 if experienced the following (past 12 months):		
Controlling behavior	0.83	0.03
		(0.04)
Emotional IPV	0.59	0.00
		(0.05)
Physical IPV	0.50	-0.01
		(0.05)
Sexual IPV	0.16	0.03
		(0.04)
Physical or sexual IPV	0.51	-0.01
		(0.05)
Emotional, Physical or Sexual IPV	0.67	-0.02
		(0.05)

Note: Observations = 395.

1.2.6 Attrition Balance

In Table 1.A3, we check balance for two compliance measures: column (1) shows whether we were able to reach the respondent and complete the endline survey itself, and column (2) refers to whether she was eligible for the IPV section at endline. Given our IPV questions have a recall period of 12 months, we administered the IPV module only to those who are currently partnered or have been so in the past 12 months. Since the IPV analysis is indeed constrained to only those who went through the IPV questionnaire at all, it is necessary to check for any differential attrition in partner status. In addition, given that often in developed countries, IPV survivors are encouraged to leave the violent partner, this is also a meaningful outcome that shows how women in our study select in or out of a relationship.

For the endline survey, we were able to successfully track 359 women (91% of the baseline sample), and the attrition rate is balanced between treatment and control. We use IPV questions with a recall period of 12 months, thus we administer the IPV module to those who currently has an intimate partner or had one within 12 months prior to the survey date. Among the 359 we tracked for endline, 314 were eligible for the IPV module, and as in column (2) of Table 1.A3, we find a 2 percentage point difference between treatment and control in this partner status. While this difference is not statistically significant, we also report the Lee (2009) bound estimates for the effects on IPV outcomes in Table 1.A5.

1.3 Results

1.3.1 Effects on IPV

In this section, we examine the WIN program effects on our primary IPV outcomes. We run the following regression:

$$Y_i = \beta WIN_i + \gamma Y_{0i} + \boldsymbol{X}_{ic}'\boldsymbol{\theta} + \phi_s + \varepsilon_i, \tag{1.1}$$

where Y_i is the outcome of interest for individual i, WIN_i treatment status instrumented with original assignment, Y_{0i} baseline measurment of the outcome, $\boldsymbol{X}_i'$ a vector of individual characteristics chosen by post-double selection LASSO, and ϕ_s strata fixed effects. The coefficient of interest is β, which is the treatment-on-treated (TOT) estimates for the effects of the female empowerment program. We also report the reduced-form effects of the randomized treatment assignment. Due to problems we discuss further in Section 1.4, we exclude the random subsample for whom IPV was measured in self-interviewing modules.

The results for IPV are presented in Table 1.2. Emotional violence decreased by 23 percentage points and physical violence by 26 points from control bases of 62 percent and 45 percent, respectively.[20] The effect sizes we find are very large in comparison to the previous literature. Lighter-touch though similar interventions have shown to have null to modest effects on IPV (Green et al. 2015; Blattman et al. 2016; Bandiera et al. 2020). The cash transfer literature finds that physical violence reduces by 0-11 percentage points during the period the female receives the transfers (Buller et al. 2018).

[20]In Table 1.A5, we show the Lee (2009) bounds results based on the difference in partner status found in Table 1.A3. For emotional IPV, the lower bound becomes statistically insignificant, but the magnitude remains fairly large with the t-statistic well greater than 1. For physical IPV, the lower bound shows a slightly smaller magnitude but remains to be statistically significant.

Table 1.2: Program Effects on IPV Indices

	(1) Controlling Behavior	(2) Emotional Violence	(3) Physical Violence	(4) Sexual Violence
Panel A. ITT				
WIN treatment	-0.02	-0.17**	-0.19***	-0.07
	(0.06)	(0.07)	(0.07)	(0.06)
Control mean	0.80	0.62	0.45	0.24
Observations	169	169	169	169
Panel B. TOT				
WIN treatment	-0.03	-0.23**	-0.26***	-0.10
	(0.09)	(0.10)	(0.10)	(0.09)
Control mean	0.80	0.62	0.45	0.24
Observations	169	169	169	169

Note: Recall period is past 12 months prior to the survey. In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include baseline measurement of outcome, strata fixed effects, and control for ACASI vs. FTFI measurement of IPV. Standard errors in parentheses.

We next look into social norms around physical and sexual IPV. Social norms related to the acceptability of IPV has been one of the widely targeted pathways in the public health literature (Ranganathan et al. 2021). In the "social ecology" framework (Heise 1998), the dynamics between a couple are embedded in many other interpersonal relationships and the community, thus social norms around IPV is a crucial driver of IPV.

To measure social norms related to IPV acceptability, we asked relevant survey questions such as: "In your opinion, is a husband justified in hitting or beating his wife if she argues with him?" We had seven such questions and asked again each referring to what the respondent believes about the community: e.g. "In

your community, is it usual for husbands to hit or beat the wife if she argues with him?" We summarize the responses to these binary questions into a z-score per Michael L. Anderson (2008b).

Table 1.3 presents our findings on social norms around IPV. When the responses to each question are indexed, we find that justifiability of physical or sexual IPV decreases by 0.3 standard deviations. This suggests that the program did reduce the acceptability of physical or sexual IPV among the program beneficiaries and that this might have been a pathway to the reduction in actual IPV experience.

However, it's also noteworthy that most women in the control group as well report that violence is not justified in any of the given situations. Neglecting the children is where the most women said violence is justifiable in the control group (12%). Also arguing with the husband and going out without telling the husband have relatively high rates of acceptability (8% and 7% respectively). Yet, the program closes this gap, to make those cases not acceptable as excuses for violence.

In Table 1.A6, we report how women responded to similar questions but referring to what she thinks of others in her community. We find that the control means are evidently higher. One explanation is that providing affirmative responses to such questions might involve stigma or embarrassment so that when the question is directed towards others instead of the respondent herself, she might be more likely to truthfully report her belief.

Table 1.3: Program Effects on Perceived Justifiability of Physical/Sexual IPV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	=1 if husband is justified to beat/hit wife when she:						=1 if husband is justified to force sex	Z-score
	Argues w/ husband	Goes out w/o telling	Doesn't care children	Burns food	Financial pressure	Refuses sex		
Panel A. ITT								
WIN treatment	-0.05*	-0.03	-0.08***	-0.01	-0.02	-0.01	0.01	-0.20**
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.09)
Control mean	0.08	0.07	0.12	0.03	0.03	0.02	0.02	-0.03
Observations	359	359	359	359	359	359	359	359
Panel B. TOT								
WIN treatment	-0.06*	-0.05	-0.10***	-0.01	-0.03	-0.01	0.01	-0.26**
	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.12)
Control mean	0.08	0.07	0.12	0.03	0.03	0.02	0.02	-0.03
Observations	359	359	359	359	359	359	359	359

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment. and include strata fixed effects. Standard errors in parentheses.

1.3.2 Effects on Economic Livelihoods

Improving women's economic opportunities have been long argued as a key strategy to reducing IPV. For example, in a household bargaining model from the economics literature, increasing the wife's economic opportunities outside of the household could heighten her "threat point" and thus the husband would less likely to perpetrate violence in order to keep her in the relationship. On the other hand, if the husband's motivations are "intrumental" (e.g. to extract resources from the wife) or "backlash" (e.g. to re-assert dominance), then economically empowering the wife could lead to more IPV.[21]

In Table 1.4, we look at labor supply outcomes. We find that the program increases labor hours for self employment by 22 hours a month (or 57 percent),

[21] See Buller et al. (2018) for discussion of the pathways and review of related cash transfer studies.

while the extensive margin is not statistically distinguishable from zero. Considering the economic empowerment component of the WIN program focuses on vocational skills and business training for small businesses, this finding is not surprising. The null effect of the extensive margin is also consistent with the fact that, for the cohort we're evaluating, Red Cross was not able to provide business capital grants at the end of the program.

We check whether there was any crowding out from other sources, but we find no significant effects on either casual labor or other income sources. While it's marginally insignificant, we also find a sizeable increase in total labor hours.

In addition to the pathways discussed above, labor supply could have incapitation effects. That is, spending more time on her own business or occupation, which is likely outside of the household or intimate relationship, leads to less time spent with her partner and thus leads to a mechanical reduction in IPV.

Table 1.4: Program Effects on Labor Supply

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Self employment		Casual labor		Other income		Total	
	=1 if any	hours	=1 if any	hours	=1 if any	hours	=1 if any	hours
Panel A. ITT								
WIN treatment	0.04	16.50*	-0.03	1.33	-0.05	-3.51	-0.03	14.32
	(0.05)	(9.72)	(0.03)	(1.67)	(0.03)	(4.56)	(0.05)	(10.30)
Control mean	0.46	38.38	0.08	1.34	0.12	11.36	0.63	51.08
Observations	359	359	359	359	359	359	359	359
Panel B. TOT								
WIN treatment	0.06	21.88*	-0.04	1.77	-0.06	-4.65	-0.04	19.00
	(0.07)	(12.87)	(0.03)	(2.20)	(0.04)	(6.02)	(0.07)	(13.60)
Control mean	0.46	38.38	0.08	1.34	0.12	11.36	0.63	51.08
Observations	359	359	359	359	359	359	359	359

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include strata fixed effects. Standard errors in parentheses.

In Table 1.5, we examine how the program affected other economic outcomes. Results show that the program increased expenditure by 49 percent. The effect sizes are surprisingly large. In Table 1.A7, we show effects by expenditure categories, and we see that the effects are mostly driven by expenses on food items and nondurables. While we find no significant effects on income, our survey questions for income might not be as exhaustive as in the expenditure section to capture many income sources. Thus expenditure is our preferred measure for economic welfare.

Table 1.5: Program Effects on Economic Outcomes

	(1) Expenditure	(2) Income	(3) Food Security	(4) Net Wealth
Panel A. ITT				
WIN treatment	9.10***	-1.17	0.06	80.25
	(2.79)	(4.11)	(0.11)	(101.98)
Control mean	24.81	21.71	-0.00	453.37
Observations	359	359	359	359
Panel B. TOT				
WIN treatment	12.07***	-1.55	0.08	106.44
	(3.78)	(5.41)	(0.14)	(134.46)
Control mean	24.81	21.71	-0.00	453.37
Observations	359	359	359	359

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include strata fixed effects. Standard errors in parentheses.

1.3.3 Effects on Psychological Wellbeing

Psychological wellbeing is also a primary outcome of the program, given that counseling is one of the key "response" interventions recommended by public health experts (Ghandour et al. 2015), suggesting that IPV victimization is correlated with mental health disorders (Devries et al. 2013a; Fulu et al. 2013; Machisa et al. 2017; Trevillion et al. 2012).

We use two main outcomes. First is the distress index from the 10-question Hopkins Symptom Checklist (HSCL-10). HSCL is generally used in clinical and epidemiological settings to measure psychological distress with a fairly straightfor-

ward set of 10 questions, such as "In the past 7 days, how often were you blaming yourself for things?" Respondents choose an option among "Not at all," "A little," "Quite a bit," and "Extremely," and we add up the responses by the assigned numeric codes. Second, we construct a happiness index using the Happiness and Well-being questions from the World Values Survey. An example question is: "In a 1 to 10 scale, how much freedom of choice and control you feel you have over the way your life turns out?" Responses to such five questions are standardized to a z-score per Michael L. Anderson (2008b).

In Table 1.6, we find rather modest effects. Both outcomes go in the expected direction, a reduction in distress and an increase in happiness, but the magnitudes are small and not statistically significant. These are indeed surprising, considering the program heavily focuses on psychological therapy sessions. Yet, the endline was 12 months after program completion, and it is possible that the effects quickly dissipated within the year. Blattman et al. (2017) and Haushofer et al. (2020) find similar results where the effect of psychotherapy sessions show significant improvement psychological wellbeing in the short term, but no effect after one year since the last therapy session.

Table 1.6: Program Effects on Psychosocial Wellbeing

	(1) Distress Index (HSCL-10)[a]	(2) Happiness Index (z-score)[b]
Panel A. ITT		
WIN treatment	-0.01	0.06
	(0.05)	(0.10)
Control mean	2.01	0.00
Observations	359	359
Panel B. TOT		
WIN treatment	-0.02	0.07
	(0.07)	(0.14)
Control mean	2.01	0.00
Observations	359	359

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include strata fixed effects. Standard errors in parentheses.
[a] 10-question Hopkins Symptom Checklist (HSCL-10).
[b] Happiness and Well-being questions from the World Values Survey, standardized per Michael L. Anderson (2008b).

1.4 Threats to Validity

1.4.1 IPV Measurement Error

A possible threat to validity of our analysis comes from the fact that our outcomes are measured by survey responses. In particular, the IPV outcomes are constructed from what women in our sample self report in our surveys, and this might lead to several concerns. In this section we address each of them.

Underreporting of IPV in surveys

It is widely concerned that IPV is underreported possibly due to factors like social taboos, feeling of shame, emotional pain, and fear of retribution (WHO 2012; Garcia-Moreno et al. 2013). However, in a professionally administered survey, these factors are likely mitigated owing to the fact that the respondent goes through an informed consent procedure where confidentiality of what she reports is assured and often the enumerator has no reason to interact with the respondent again. Yet even with underreported *levels* of IPV, these do not necessarily introduce bias to treatment effects in an impact evaluation setting, because the (nonclassical) measurement error is canceled out by taking the *differences* between treatment and control.

However, one might be concerned that the true levels of IPV become different between treatment and control (e.g. lower in the treatment if the intervention was effective), and even if the probability of IPV being underreported is contant, this could attenuate the treatment effect.[22] We cannot directly test this in this paper's setting, because the underreporting propensity is unlikely to be the same between treatment and control (discussed more in following points). Instead, in a companion project where we evaluate the effect of unconditional cash transfers in rural Liberia and Malawi (Park et al. 2021), we introduce an alternative survey tool that could alleviate social desirability bias (as we do in this paper too, and explained more below), and we find no differential cash effects on IPV between survey modes. This finding suggests that underreporting of IPV itself does not bias the treatment effects at least when the measurement error is not correlated with treatment (like unconditional cash transfers).

[22] Assume the true prevalence of IPV is $(y - \beta)$ in treatment and y in control, and that the proportion of people who truthfully report IPV is $p < 1$ (constant between treatment and control). Then the estimated treatment effect based on reported IPV rates are $-p \cdot \beta$, which is smaller in magnitude than the true treatment effect β.

Experimenter demand effects

Nonetheless, the analysis in this paper could be threatened by differing IPV reporting behavior between treatment and control. One possiblity is experimenter demand effects. Given the intervention involves psychotherapy for relationships with spouses or intimate partners, the respondents in the treatment group might believe that the researchers expect them to have a better marital relationship and experience less IPV, and thus feel pressure to underreport IPV. This would overestimate the treatment effects.

Research suggests that experimenter demand effects are modest in many settings even when the researchers made the research hypothesis salient to the study sample (Quidt et al. 2018; Dhar et al. 2018; Mummolo and E. Peterson 2019). Moreover, our endline survey was conducted by an independent survey firm that the respondents had no reason to associate with the program implementer. Also the timing of the endline was 12 months after the program had ended, so it is less likely that reporting behavior at endline was driven by the treatment.

However, to address this issue more rigorously, we cross-randomized an IPV measurement experiment at endline, where respondents answered IPV questions in either self interviewing (SI) or conventional face-to-face interviewing (FTFI). Whereas under FTFI the enumerator asks each question and the respondent responds verbally, in SI women listen to pre-recorded questions through earphones and make choices on a touchscreen by herself.[23] The main difference is that the SI module allows the respondents to report their responses anonymously to the human enumerator, which could minimize social desirability bias in IPV reporting and thus experimenter demand effects (i.e. the difference in social desirability bias

[23]We use one type of SI called audio computer-assisted self interviewing (ACASI) (Figure 1.A2). The ACASI module and the experimental design are almost identical to those of our sister project (Park et al. 2021), where we study effects of SI on IPV reporting in rural Liberia and rural Malawi.

between treatment and control).[24]

In Table 1.B5, we see the treatment effects are smaller when IPV was measured in SI, which would suggest that our main analysis based on FTFI might be driven by experimenter demand effects. However, the attenuation could be explained by measurement error introduced by the SI survey tool, which we extensively document in Park et al. (2021). If the respondent doesn't fully understand how to use the tool, she'd be making mistakes when choosing responses (classical measurement error). Since the mean of an individual yes/no IPV question is typically below 0.5, such measurement error would *increase* the rate (biased towards 0.5), and this could attenuate the treatment effect estimate in SI.[25]

In fact, a significant portion of our sample seems to be making mistakes under SI. In Table 1.B2, we find sizeable differences in how people report to a set of innocuous questions between FTFI and SI. For example, while everyone in the control under FTFI said "yes" to the questions "Did it rain in your community last year?" and "Did you sleep at all past week?", only 82% in the control group and 90% in the treatment group did so under SI. Overall, among seven questions, five of them indicate statistical significance when SI effects are pooled. Except for one question, we don't find evidence that either the treatment or control group is making less mistakes. Assuming that these questions are truly innocuous and respondents have no other reason to differentially report by FTFI and SI, the

[24]While the original intent of SI is to minimize underreporting by protecting the respondents from feeling shame or discomfort, it is also possible on the other hand that the respondent could feel more comfortable sharing unfortunate experiences with a human being. Conducting the IPV module is typically considered a conversation, and often respondents seek counseling from the human enumerator (M. Ellsberg et al. 2001).

[25]Suppose the reported IPV rates under FTFI are $(y - \beta)$ for treatment and y for control. Under SI, assume there are two types: p fully understand the module and respond in the same way she would have under FTFI, and $(1 - p)$ make mistakes under SI and randomly choose between yes and no. Then the reported rates under SI are $p \cdot (y - \beta) + (1 - p) \cdot 0.5$ for the treatment and $p \cdot y + (1 - p) \cdot 0.5$ for the control, and taking the difference, the estimated treatment effect under SI is $-p \cdot \beta$. This is smaller in magnitude than that under FTFI, $-\beta$, and the difference is determined by how many people don't understand the SI tool $(1 - p)$.

results altogether suggest that many are making mistakes in SI and the attenuation in shown Table 1.B5 is not necessarily explained by experimenter demand effects.

Enhanced sensitization of IPV

It's also possible that IPV reporting behavior is correlated with treatment in the other direction. While the treatment group becomes more sensitized of their IPV experience and more likely to truthfully report IPV, the control group might not be sensitized enough and remain underreporting IPV. This would *underestimate* the treatment effect. One could have such concern given that we find treatment effects in perceived justifiability of IPV in Table 1.3. However, it's noteworthy even among the control group, a vast majority thinks violence is not justified. One deviation is for the situation where the wife neglects the children; 13% reported that physical violence can be justified in this case, whereas the means for other questions are 3-9%. Yet, at least from what's reported, our study sample overall appears to be a context where already violence is not justified in most cases. However, even if IPV reporting behavior is significantly affected by this factor, the main results we find on IPV would be the lower bounds of the true effect.

Control group pretending to look worse

One might be concerned that the control group reports higher rates of IPV in order to look more disadvantaged. This might be plausible because our sampling frame were women who had voluntarily applied to the program for consideration. Even though this was more than two years prior to our endline, it's possible that they are still willing to be eligible for future program enrollments. However, as explained earlier, respondents had virtually no reason to link our enumeration

team to the program or Red Cross. In the informed consent form we administer at the beginning of every survey, we make it clear that no personal or identifiable information will be shared with any party, including the government or any non-government organizations. Therefore, it's unlikely that anyone in our study sample believes what she reports to us could affect her prospects for any program.

1.4.2 Incapacitation Effect

Another type of concern is that IPV experience might be reduced in the treatment group mechanically because they spend more time in the program. This could be especially concerning since the treatment group had to attend the program center 4-5 hours a day, which amounts to at least 20-25 hours a week physically away from the spouse.[26] However, our endline survey was conducted about 12 months after the program had ended, and we have no outcomes measured for more than 12 months prior to the survey. Therefore, the outcomes do not capture anything that happened while the program was running. Yet, *after* the program, as we find in Table 1.4, treatment group worked more outside of the household (and away from her partner), and it's possible this was one of the mechanisms through which IPV was reduced.

1.5 Conclusion

Our randomized evaluation of a multifaceted female empowerment program finds that it considerably reduces emotional and physical IPV experienced by women, restricting the analysis to IPV outcomes measured in a conventional setting. We also find sizeable effects on labor supply and expenditure. After 12

[26]While some of controlling behavior and emotional IPV can be perpetrated remotely (e.g. over the phone), physical and sexual IPV do require physical contact.

months since the program, we find small insignificant effects on psychological wellbeing.

These findings suggest that a holistic approach to IPV prevention is effective. This is consistent with the public health literature on IPV emphasizing that the multi-level factors of IPV are important in designing interventions. One caveat of this study is that we cannot quantify the marginal benefit of a single program component. We leave this to future research.

1.6 Appendix

Appendix A

Figure 1.A1: Study Timeline and COVID-19 Disruptions

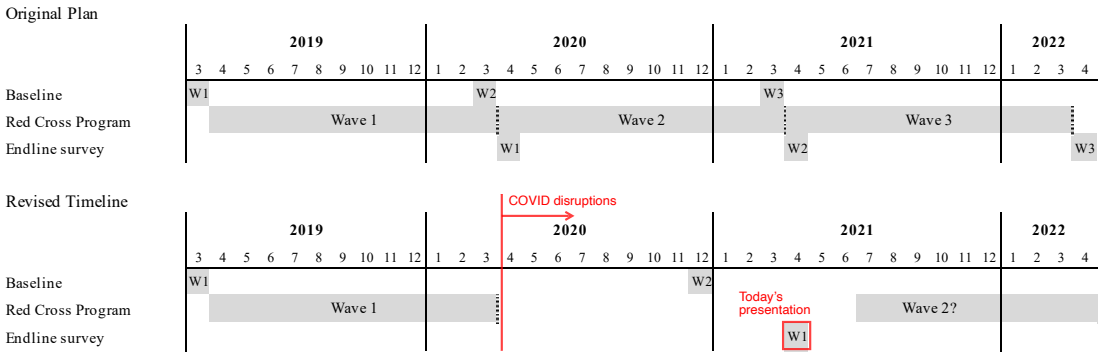


Figure 1.A2: Self Interviewing (SI) Survey Module

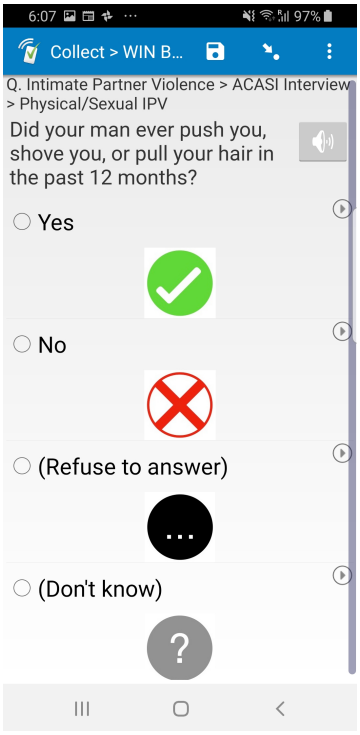


Table 1.A1: WIN Program Components

Program Component	Description
Psychological support	One-to-one and group counselling, stress management, family/couple therapy
Literacy classes	Reading and writing curriculum by Ministry of Education
Child care	During program participation
Medical checkups	Free primary medical check-ups at Red Cross clinic
Vocational skills training	Baking, cosmetology, and tailoring
Entrepreneurship training	Financial literacy, business planning/management, etc.
Business start-up capital	250 USD worth of capital along with 30 USD cash grant

Table 1.A2: Selection Criteria of WIN Program

1. Ex-combatant	5. Single mother/self-supported
2. Previous commercial sex worker	6. Illiterate
3. Victims of rape/domestic violence	7. Economically vulnerable
4. Witness of extreme violence	8. Drug user

Table 1.A3: Attrition Balance

	(1) =1 if completed endline survey	(2) =1 if completed IPV survey at endline[a]
WIN treatment	0.00	-0.02
	(0.03)	(0.04)
Control mean	0.91	0.81
Overall mean	0.91	0.79
Observations	395	395

Note: Regressions include strata fixed effects. Standard errors in parentheses.
[a] IPV questionnaire is administered to only those who are currently married or has an intimate partner, or have been so in the 12 months prior to the survey.

Table 1.A4: Program Effects on Frequency-integrated IPV Indices

	(1)	(2)	(3)	(4)
	Frequency-integrated Indices[a]			
	Emotional IPV	Physical IPV	Sexual IPV	Any IPV
Panel A. ITT				
WIN treatment	-0.15	-0.30***	-0.18	-0.35***
	(0.14)	(0.11)	(0.12)	(0.11)
Control mean	0.00	-0.00	-0.00	0.00
Observations	169	169	169	169
Panel B. TOT				
WIN treatment	-0.20	-0.42**	-0.25	-0.48***
	(0.19)	(0.16)	(0.17)	(0.16)
Control mean	0.00	-0.00	-0.00	0.00
Observations	169	169	169	169

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include baseline measurement of outcome, strata fixed effects, and control for ACASI vs. FTFI measurement of IPV. Standard errors in parentheses.

Table 1.A5: Program Effects on IPV Indices - Lee Bounds

	(1)	(2)	(3)	(4)	(5)	(6)
	Emotional IPV			Physical IPV		
	Baseline	Lower Bound	Upper Bound	Baseline	Lower Bound	Upper Bound
WIN treatment	-0.20**	-0.15	-0.26***	-0.22**	-0.16*	-0.25***
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Control mean	0.62	0.59	0.68	0.45	0.41	0.49
Observations	169	162	162	169	162	162
	Sexual IPV			Any IPV		
	Baseline	Lower Bound	Upper Bound	Baseline	Lower Bound	Upper Bound
WIN treatment	-0.10	0.00	-0.11	-0.18**	-0.14	-0.23**
	(0.08)	(0.08)	(0.09)	(0.09)	(0.10)	(0.09)
Control mean	0.24	0.17	0.26	0.66	0.63	0.72
Observations	169	162	162	169	162	162

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include baseline measurement of outcome, strata fixed effects, and control for ACASI vs. FTFI measurement of IPV. Standard errors in parentheses.

Table 1.A6: Program Effects on Perceived Others' Justifiability of Physical/Sexual IPV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	=1 if husband is justified to beat/hit wife when she:						=1 if husband is justified to force sex	Z-score
	Argues w/ husband	Goes out w/o telling	Doesn't care children	Burns food	Financial pressure	Refuses sex		
Panel A. ITT								
WIN treatment	-0.04	-0.07	-0.11***	-0.09***	-0.02	-0.07*	-0.07**	-0.22**
	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.09)
Control mean	0.30	0.30	0.27	0.17	0.13	0.16	0.14	-0.02
Observations	359	359	359	359	359	359	359	359
Panel B. TOT								
WIN treatment	-0.05	-0.09	-0.15***	-0.12***	-0.03	-0.09*	-0.09**	-0.29**
	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.13)
Control mean	0.30	0.30	0.27	0.17	0.13	0.16	0.14	-0.02
Observations	359	359	359	359	359	359	359	359

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment. and include strata fixed effects. Standard errors in parentheses.

Table 1.A7: Program Effects on Expenditure Items

	(1) Food	(2) Nondurables	(3) Clothes	(4) Education	(5) Health	(6) Religious contributions	(7) Family events	(8) Nonmedical emergency
Panel A. ITT								
WIN treatment	3.74**	4.67	1.17	0.69	0.42	0.33	-0.54	0.10
	(1.65)	(2.99)	(1.87)	(2.20)	(1.43)	(0.64)	(1.52)	(0.13)
Control mean	10.05	27.06	6.54	15.15	6.07	2.99	5.07	0.11
Observations	359	359	359	359	359	359	359	359
Panel B. TOT								
WIN treatment	4.96**	6.19	1.55	0.92	0.56	0.44	-0.72	0.14
	(2.21)	(4.00)	(2.47)	(2.90)	(1.88)	(0.85)	(2.01)	(0.16)
Control mean	10.05	27.06	6.54	15.15	6.07	2.99	5.07	0.11
Observations	359	359	359	359	359	359	359	359

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include baseline measurement of outcome, and strata fixed effects. Standard errors in parentheses.

Table 1.A8: Program Effects on Income

	(1)	(2)	(3)	(4)
	Respondent			Spouse's income
	Self employment	Casual labor	Other job	
Panel A. ITT				
WIN treatment	3.63	-1.25	-3.55	-0.99
	(3.63)	(0.80)	(2.23)	(5.79)
Control mean	12.40	1.91	7.40	33.44
Observations	359	359	359	359
Panel B. TOT				
WIN treatment	4.82	-1.66	-4.71	-1.32
	(4.79)	(1.06)	(2.95)	(7.63)
Control mean	12.40	1.91	7.40	33.44
Observations	359	359	359	359

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include baseline measurement of outcome, and strata fixed effects. Standard errors in parentheses.

Table 1.A9: Program Effects on Assets

	(1) Business capital	(2) Durables	(3) Livestock	(4) Savings	(5) Debt
Panel A. ITT					
WIN treatment	5.90	63.95	0.31	13.87	3.79
	(16.14)	(90.02)	(9.16)	(17.01)	(3.44)
Control mean	44.19	361.22	23.00	30.46	5.49
Observations	359	359	359	359	359
Panel B. TOT					
WIN treatment	7.83	84.82	0.42	18.40	5.02
	(21.24)	(118.73)	(12.07)	(22.38)	(4.53)
Control mean	44.19	361.22	23.00	30.46	5.49
Observations	359	359	359	359	359

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include baseline measurement of outcome, and strata fixed effects. Standard errors in parentheses.

Table 1.A10: Program Effects on Interpersonal Transfers

	(1)	(2)	(3)	(4)
	Transfers sent		Transfers received	
	Spouse	Non-spouse	Spouse	Non-spouse
Panel A. ITT				
WIN treatment	-0.22	-1.53	2.59	1.68
	(0.48)	(1.48)	(4.52)	(2.89)
Control mean	1.40	6.41	37.40	8.15
Observations	278	359	278	359
Panel B. TOT				
WIN treatment	-0.28	-2.03	3.33	2.23
	(0.61)	(1.95)	(5.77)	(3.80)
Control mean	1.40	6.41	37.40	8.15
Observations	278	359	278	359

Note: In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include baseline measurement of outcome, and strata fixed effects. Standard errors in parentheses.

Appendix B: Possible Threats to Validity

Table 1.B1: SI Screening

	(1) Mean (=1 if yes)
Are you a woman?	0.98
Do you live in [the county/district where the survey is being conducted]?	0.97
In the past week, did you sleep, during day or night?	0.97
In the past year, did it rain in your village one time or more?	0.96
=1 if yes to all questions	0.90
=1 if yes to woman and rain questions	0.98
Observations	303

Note: These four questions were asked in SI to everyone included in SI measurement experiment.

Table 1.B2: SI Effects on Placebo Questions, by WIN treatment status

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Questions for which answer should be yes:				Questions for which answer could be yes/no:				
			Index						
	Rain	Sleep	%(yes)	=1 if yes to all	Farm work	Market	Int'l travel	Rice	Meat
SI × WIN control (β)	-0.07**	-0.14***	-0.11***	-0.18***	0.11**	0.08	-0.04	-0.10***	-0.16*
	(0.03)	(0.04)	(0.03)	(0.05)	(0.05)	(0.05)	(0.03)	(0.04)	(0.08)
SI × WIN treatment (γ)	-0.04	-0.09**	-0.07**	-0.10**	-0.02	0.07	0.01	-0.13***	-0.13
	(0.03)	(0.04)	(0.03)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.08)
WIN	-0.01	-0.03	-0.02	-0.05*	0.03	0.02	-0.02	-0.03	0.01
	(0.02)	(0.02)	(0.01)	(0.03)	(0.04)	(0.06)	(0.04)	(0.02)	(0.08)
FTFI × WIN control mean	1.00	1.00	1.00	1.00	0.04	0.84	0.06	1.00	0.56
p-value ($\beta = \gamma$)	0.609	0.361	0.356	0.241	0.053	0.890	0.334	0.617	0.737
Observations	298	298	298	298	298	298	298	298	298
Post-estimation calculation									
Pooled SI effects	-0.06	-0.11	-0.09	-0.14	0.05	0.07	-0.01	-0.11	-0.14
p-value	0.010	0.000	0.000	0.000	0.139	0.049	0.659	0.000	0.016

Note: Regressions include individual controls (including all variables in Table 1.B6). "Screen Pass" is defined by selecting "yes" to all questions in Table 1.B1. Standard errors in parentheses.

Table 1.B3: SI Effects on IPV Questions, by WIN treatment status

	(1)	(2)	(3)	(4)	(5)
	=1 if responded yes to individual question in the following category:				All questions pooled
	Controlling Behavior	Emotional IPV	Physical IPV	Sexual IPV	
SI × WIN control (β)	0.01	-0.03	-0.06	0.06	-0.01
	(0.04)	(0.06)	(0.04)	(0.05)	(0.04)
SI × WIN treatment (γ)	0.11***	0.04	0.06	0.05	0.07**
	(0.04)	(0.06)	(0.04)	(0.05)	(0.04)
WIN	-0.11***	-0.08	-0.12***	-0.05	-0.10**
	(0.04)	(0.06)	(0.04)	(0.05)	(0.04)
FTFI × WIN control mean	0.37	0.38	0.22	0.16	0.29
p-value ($\beta = \gamma$)	0.097	0.409	0.057	0.947	0.142
Number of individuals	298	298	297	298	298
Observations	2,056	1,184	1,776	889	5,905
Post-estimation calculation					
Pooled SI effects	0.06	0.01	0.00	0.06	0.03
p-value	0.046	0.845	0.963	0.112	0.255

Note: Observations at respondent-question level. See Table 1.B4 for index-level results. Regressions include question-level fixed effects. Standard errors clustered at individual level in parentheses.

Table 1.B4: SI Effects on IPV Indices, by WIN treatment status

	(1)	(2)	(3)	(4)	(5)
	=1 if responded yes to *any* question in the following category:				Any IPV
	Controlling Behavior	Emotional IPV	Physical IPV	Sexual IPV	
SI × WIN control (β)	0.09	-0.04	-0.14*	0.09	-0.02
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
SI × WIN treatment (γ)	0.18***	0.09	0.01	0.12*	0.10
	(0.06)	(0.08)	(0.07)	(0.07)	(0.08)
WIN	-0.02	-0.16**	-0.21***	-0.08	-0.15*
	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)
FTFI × WIN control mean	0.77	0.63	0.47	0.24	0.67
p-value ($\beta = \gamma$)	0.290	0.271	0.164	0.762	0.301
Observations	298	298	298	298	298
Post-estimation calculation					
Pooled SI effects	0.14	0.03	-0.07	0.11	0.04
p-value	0.002	0.629	0.207	0.037	0.478

Note: See Table 1.B3 for question-level results.

Table 1.B5: Program Effects and SI Effects on IPV Indices - TOT - Screen Pass only

	(1)	(2)	(3)	(4)	(5)
	=1 if experienced any instance of the following category:				Any IPV
	Controlling Behavior	Emotional IPV	Physical IPV	Sexual IPV	
Panel A. ITT					
WIN × FTFI (γ)	-0.01	-0.15**	-0.20***	-0.07	-0.14*
	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)
WIN × SI (β)	0.07	-0.03	-0.06	-0.04	-0.04
	(0.05)	(0.08)	(0.07)	(0.08)	(0.08)
Non-WIN × FTFI mean	0.77	0.63	0.47	0.24	0.67
Non-WIN × SI mean	0.84	0.54	0.29	0.31	0.60
p-value ($\beta = \gamma$)	0.315	0.280	0.177	0.755	0.339
Observations	298	298	298	298	298
Post-estimation calculation					
Pooled program effects	0.03	-0.10	-0.13	-0.06	-0.09
p-value	0.508	0.083	0.014	0.248	0.108
Panel B. TOT					
WIN × FTFI (γ)	0.00	-0.25**	-0.28***	-0.10	-0.22**
	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)
WIN × SI (β)	0.12*	0.01	-0.03	-0.04	0.00
	(0.06)	(0.10)	(0.09)	(0.09)	(0.10)
Non-WIN × FTFI mean	0.77	0.63	0.47	0.24	0.67
Non-WIN × SI mean	0.84	0.54	0.29	0.31	0.60
p-value ($\beta = \gamma$)	0.296	0.074	0.060	0.584	0.111
Observations	298	298	298	298	298
Post-estimation calculation					
Pooled program effects	0.06	-0.13	-0.16	-0.07	-0.11
p-value	0.293	0.079	0.020	0.254	0.108

Note: Sample includes only those who passed screening, i.e. those who selected "yes" to all questions in Table 1.B1. In Panel B, regressions are TOT estimates, where the treatment indicator is instrumented with the original assignment to treatment, and include baseline measurement of outcome, and strata fixed effects. Standard errors in parentheses.

Table 1.B6: SI Randomization Check

	(1) Control Mean [SD]	(2) Treatment - Control
Panel A. Demographics		
=1 if currently married or has partner	0.88	0.02
		(0.04)
Age	30.44	1.74**
	[6.79]	(0.83)
Number of household members	5.06	0.72**
	[2.71]	(0.35)
Panel B. Education and digital literacy		
Years of education	8.28	-0.52
	[4.14]	(0.47)
=1 if able to write/read in English	0.84	0.03
		(0.04)
=1 if has access to mobile phone	0.89	-0.00
		(0.04)
Panel C. Household wealth		
Food security index (z-score)	0.00	-0.05
	[1.00]	(0.12)
Total expenditure (monthly)	124.08	3.06
	[83.00]	(10.21)
Net value of durables, livestock, and financial asset	421.43	121.86
	[828.44]	(108.39)
Non-agricultural income (monthly)	21.45	3.96
	[38.75]	(4.77)
Panel D. Empowerment-related outcomes		
=1 if has her own income source	0.60	0.04
		(0.06)
Number of children	2.35	0.47**
	[1.68]	(0.20)
Observations	303	

Table 1.B7: Post-SI Survey of Technical Difficulties Self-reported by Respondents

	(1) Mean (=1 if yes)
Was the audio loud enough to hear?	0.99
Was the audio speaking speed okay?	0.98
Was it easy for you to remember the meaning of pictures?	0.97
Was it easy for you to choose answers on the screen?	0.97
Was it easy for you to move between questions on the screen?	0.97
Observations	145

Note: Questions were asked only to those in the SI treatment group (i.e., the FTFI group did not get these questions).

Appendix C: Survey instrument

Controlling behavior

1. Did your man ever try to keep you from seeing your friends in the past 12 months?

2. Did your man ever try to stop you from meeting or speaking to your family of birth in the past 12 months?

3. Did your man ever need to know where you are all the time in the past 12 months?

4. Did your man ever stop talking to you or treat you with no interest in the past 12 months?

5. Did your man ever get angry if you speak with another man in the past 12 months?

6. Did your man often think that you are unfaithful in the past 12 months?

7. In the past 12 months, did your man ever expect you to ask for his approval before you go to a health clinic or hospital?

Emotional IPV[27]

1. Did your man ever insult you or make you feel bad about yourself in the past 12 months?

2. Did your man ever make you feel small in front of other people in the past 12 months?

3. Did your man ever mean to scare you (for example, by the way he looked at you, by yelling and bursting things) in the past 12 months?

4. Did your man ever threaten to hurt you or someone you care about in the past 12 months?

[27] For each IPV question, if the answer is "yes", a follow-up question about frequency appears, asking whether it happened (i) one or two times, (ii) three to five times, or (iii) more than five times.

Physical IPV[27]

1. Did your man ever slap you or throw something at you that could hurt you in the past 12 months?
2. Did your man ever push you, shove you, or pull your hair in the past 12 months?
3. Did your man ever hit you with his hand or with something else that could hurt you in the past 12 months?
4. Did your man ever kick you, drag you or beat you up in the past 12 months?
5. Did your man ever mean to choke or burn you in the past 12 months?
6. Did your man ever threaten to use or actually use a gun, knife or other weapon against you in the past 12 months?

Sexual IPV[27]

1. Did your man ever physically force you to do man and woman business when you did not want to in the past 12 months?
2. Did you ever do man and woman business when you did not want to because you were afraid of what your man might do in the past 12 months?
3. In the past 12 months, while doing man and woman business, did your man ever force you to do something that made you feel small or bad about yourself?

Non-sensitive placebo questions

1. Did it rain in your village one time or more in the past year?
2. Did you do any farm work in the past year?
3. Did you sleep in the past week, during day or night?
4. Did you go to the market in the past week?
5. Did you travel outside of Liberia in the past week?
6. Will you, or anyone in your household, eat any rice next week, one time or more?
7. Will you, or anyone in your household, eat any type of meat next week, one time or more?

Chapter 2

The Dynamic Effects of Cash Transfers: Evidence from Rural Liberia and Malawi

2.1 Introduction

A substantial literature has documented the beneficial effects of cash transfers on various measures of well-being. However, while cash will tautologically increase the immediate consumption of any normal good, evidence on whether and how these effects persist is limited. This is particularly relevant for unconditional cash transfers (UCTs) paid out in a lump sum, because realizing sustained impacts from one-time infusions of cash requires productive investment, and in many contexts it is not obvious whether such productive investment opportunities exist for the average household. In their absence, households tend to spend money on financing immediate consumption and program effects tend to be temporary.

This paper measures the dynamic effects of a randomized cash transfer program in 300 villages each in Liberia and Malawi, two of the poorest countries in the world.[1] In the experiment, all households within half of the sampled villages received cash transfers, and those in the other half did not. The value of the transfer was large, averaging $500, equivalent to 86% of estimated *annual* household average expenditure in Liberia and 126% in Malawi. We measure outcomes via bi-monthly panel phone surveys with 20% of the household sample (spanning all sampled villages). These surveys continued for about 1 year after disbursement in Liberia and 2 years in Malawi. We supplement this information with an in-person endline with the entire sample about 18-25 months after the transfers were disbursed.

We calculate dynamic treatment effects for 4 pre-specified outcomes: food security, expenditures, income, and inter-personal transfers. These show meaningful dynamics. In both countries, we find lasting effects on food security until the end

[1] In 2016, the two countries were ranked 218th and 222nd (out of 226) in an IMF ranking of GDP per capita.

of data collection. This is despite the fact that effects on food expenditures are short-lived, and we observe no measurable effect on non-agricultural income in either country. Dynamics for non-food expenditures differ somewhat in the 2 countries: we find lasting effects in Liberia but not in Malawi. We find no effect on inter-personal transfers in either country, implying that the most of the cash was spent within the beneficiary household.

We also use the endline to examine effects on other pre-specified outcomes, specifically intimate partner violence (IPV), psychological well-being, and (self-reported) resilience. We find an 8 percentage point decline in IPV incidence over the past year in Liberia but no effect in Malawi (likely due in part to much higher baseline prevalence in Liberia). We also find a meaningful increase in psychological well-being (0.34 standard deviations in Liberia and 0.10 in Malawi) and resilience (0.09 standard deviations in Liberia and 0.12 in Malawi) in both countries. We also examine other secondary outcomes, and find a decline in casual labor income in both countries, an increase in school enrollment and education spending in Liberia (where school enrollment of primary-aged children is only 52% in our data), and increases in health investment and in spending in response to health shocks in Liberia.

Our results open questions as to the pathways by which food security was lastingly affected, given that neither income nor expenditures rose. We find suggestive evidence of increased investment in several forms of productive activities which may have led to greater home production, including greater animal ownership and agricultural output, as well as an increase in the number of crops planted, despite nil (Liberia) to modest (Malawi) impacts on measured inputs. We also find increased investment in self-enterprise, suggesting that perhaps small gains in income were realized (although the measured effect on profits is modest). In-

creased financial assets do not appear to be a pathway: most recipients withdraw the money immediately, and cash savings at endline increase by only $6 (base $18) in Liberia, and $3 (base $6) in Malawi, a tiny fraction of the UCT amount.

Lastly, we examine possible spillover effects for non-recipients. As mentioned above, very little of the transfer was shared: in Liberia, we find modest evidence of increased inter-personal transfers post-disbursement, but point estimates are only a few dollars; in Malawi, we observe no effect on transfers sent at any point. We observe similar, though slightly larger, effects on transfers received. We also find no effect on labor demand by beneficiaries. In Malawi and the majority of the Liberian villages, we randomized the intensity of treatment across geographic clusters, and using this, we find minimal evidence of spillovers on all primary outcomes, though confidence intervals are wide. We non-experimentally estimate price spillovers, comparing treated and comparison markets, and find little effect (which is not surprising since the treatment affected only a small percentage of the population).

Our results are related to a vast literature on cash transfers.[2] The most closely related papers are studies about the effect of *large* UCTs, i.e., Blattman et al. (2014), Haushofer and Shapiro (2016), Egger et al. (2019), McIntosh and Zeitlin (2021) and McIntosh and Zeitlin (2022). Relative to this literature, our contribution is in estimating dynamic effects. In this sense, our research is related to studies which examine the longer-term effects of UCTs, particularly Blattman et al. (2020) and Haushofer and Shapiro (2018). Methodologically, our study is differentiated by tracking outcomes at a higher frequency. Our specific results also differ from these prior settings, particularly in that we observe no effect on in-

[2]This research has largely focused on conditional cash transfers. See Kabeer and Waddington (2015) and Millán et al. (2019) for reviews, and Parker and Todd (2017) for an *Oportunidades*-specific review.

come even in the short-term, whereas Haushofer and Shapiro (2016) do find large effects on income, as does Blattman et al. (2014) (in fact, the program of study in that case was specifically designed to support businesses). Our study also differs from Haushofer and Shapiro (2018) in that our targeting was universal within villages, allowing us to abstract away from the confounding effects of within-village spillovers (whereas in Haushofer and Shapiro 2018, several treatment effects are smaller across-village than within-, suggesting possible negative spillovers).[3]

2.2 Experimental Design

2.2.1 Experimental context and design

The NGO GiveDirectly (henceforth, GD) implemented the cash transfer program in Liberia and Malawi in 2019-2021, aiming to enroll 150 villages each in the treatment and control arms in each country. Targeted counties and districts were identified by GD and the funding partner, USAID, based on poverty levels, mobile phone coverage, and proximity to roads. Villages within each county/district were eligible if their population (measured via the most recent population census), was below a threshold size (100 households in Malawi and 125 households in Liberia).[4] All households in treatment villages were eligible for the transfer. Within each household, the transfer was made to a beneficiary chosen by the household. As the household beneficiary had to be home at the time of enrollment, the majority of them are women. Since beneficiary selection was endogenous, the household is

[3] Angelucci and De Giorgi (2009) recommend that transfers be targeted universally within a village to be able to disentangle treatment effects from inter-household spillovers, as these are likely bigger *within* a village.

[4] Since the transfers were universal, GD targeted smaller villages in order to cover enough villages while staying within its budget.

our unit of analysis.

Amongst the 600 villages, we randomized treatment, stratifying by country and district/traditional authority. Treatment villages were randomized into one of three amounts: $250, $500 or $750. Within each treatment village in Liberia, transfers were also randomized between being paid as lump sum or quarterly.[5] As village enrollment took several months, the start date of transfers varied across villages. There was also some variation in the roll-out of the transfers between countries. In Liberia, the project was implemented in two waves: a smaller "Wave 1" with 90 villages, in which transfers were disbursed from March 2019 to February 2020; and a larger "Wave 2" with 210 villages, in which transfers were disbursed from March 2020 to July 2021. The timing of transfers in Wave 2 was affected by COVID in 2020 because enrollment was paused during lockdowns. In Malawi, all 300 villages were enrolled in a single wave (though enrollment took several months) and transfers were disbursed from July 2019 to February 2020. Transfers were disbursed via mobile money; households could purchase a cell phone with a mobile-money-enabled SIM during enrollment from GD.

Households were identified in collaboration with GD, who visited every village in the sampling frame and recorded every habitation structure with a GPS pin. This enabled verification of the village population, as well as provided a sampling frame for the baseline survey, which was carried out independently of GD. For data collection, we sampled 10 pins from this list (with replacements) and attempted to interview those households (6,000 households across both countries). In some villages, we could enroll fewer households, leading to a final sample of 2,715 in

[5]For the lump sum sub-treatment, transfers were disbursed in 1-3 tranches. GD capped disbursements at $250 per tranche, making additional tranches in the following months. Thus, respondents receiving $250 received one transfer; those receiving $500 or $750 received 2 or 3 transfers over consecutive months.

Liberia and 2,944 in Malawi. Since IPV is a primary outcome, surveys targeted female heads of households.[6]

Two of the 10 households from every village were further sampled to answer a monthly phone survey designed to measure a pre-defined set of outcomes. Each phone survey respondent received a phone (worth $10-15). We called one household per village in even-numbered months, and the other in odd-numbered months. This results in a monthly village-level panel and a bi-monthly household-level panel. Because households were randomly selected, these respondents represent approximately 32,000 households (or about 150,000 people) in these 600 villages. Figure 3.A1 provides a timeline of activities.

2.2.2 Data

We use data from 4 primary sources. First, we conducted baseline surveys in November-December 2018 for Liberia Wave 1, November-December 2019 for Liberia Wave 2, and April-July 2019 for Malawi. The surveys took about 2-3 hours to administer, and included questions on demographics, agriculture, cash flows, food security, mobile money usage, shocks and resilience, and IPV (among other subjects).

Second, we conducted endline surveys in late 2020 for Liberia Wave 1 (18-20 months after disbursement), late 2021 for Liberia Wave 2 (18-22 months after disbursement) and April-July 2021 in Malawi (21-25 months after disbursement). These surveys were similar to the baselines in length and scope.[7]

Third, as discussed above, 20% of the sample was randomized into phone

[6]Male heads were interviewed only when the female was absent and unreachable.

[7]See Jeong et al. (2022) and Park et al. (2022a) for analyses of cross-randomized survey experiments on survey length and interview modality of IPV, respectively, in these baseline and endline surveys.

surveys, which included questions on food security, expenditures, income, labor supply, transfers, savings, and credit.[8] Each household was called every other month, with households within a village alternating months (so that each village has a data point for every month). The phone surveys were administered from July 2019 to August 2021 in Malawi, from February 2019 to September 2020 for Liberia Wave 1, and from January 2020 to October 2021 for Wave 2.

Fourth, we collected monthly prices from 80 markets in Liberia and 95 in Malawi over a two-year period, starting before the transfers began.[9] Figure 2.A2 shows the location of study villages and markets. Market prices were collected from vendors enrolled by us for the study: 1,220 vendors in Liberia (Wave 1: 333, Wave 2: 887), and 1,378 in Malawi.[10]

Attrition for the endline and phone surveys is shown in Table 2.A1 and Table 2.A2, respectively. In both countries, our endline attrition was low and balanced across treatments: 96% of all households completed the endline in Liberia, and 94% in Malawi, and there is no evidence of differential attrition (Columns 1 and 2). In the phone survey, attrition is relatively low and balanced by treatment in Malawi. Over 95% of the sample participated in early rounds; though this percentage fell over time, we still successfully interviewed 80% or more after 2 years. However, attrition is substantially higher in Liberia, largely due to the country's inferior phone network. Also, in Wave 1, we noticed that households in the treatment group were more likely to switch to the SIM card provided by GD, thus making it more difficult to reach these respondents. We, therefore, drop

[8]During COVID-19 lockdowns, we added questions aimed at measuring their impact. Aggarwal et al. (2020b) documents the impact of COVID in these two countries.

[9]Twenty-three of the 80 markets in Liberia and 10 of the 95 in Malawi were in cash transfer areas, while the remaining markets were in non-treatment areas. We selected 1 treatment market per traditional authority in Malawi, and all markets in treatment areas in Liberia.

[10]Vendors were enrolled if they had access to a mobile phone, and sold the items on our list. We tried to enroll at least 2 vendors per market.

Liberia Wave 1 from the phone survey analysis. In Wave 2, we took proactive steps to avoid this problem and managed to achieve balanced compliance, but is lower than Malawi: compliance peaks at 75% immediately after enrollment, but falls below 50% within 8 or 9 rounds (16-18 months).

2.2.3 Defining outcomes

Primary outcomes for this study were pre-defined in a pre-analysis plan (Aggarwal et al. 2021a). Our analysis of dynamic treatment effects focuses on 4 outcomes: (1) a food security index (FSI), comprised of standardized scores of the Household Dietary Diversity Score (HDDS), measured over the past 24 hours, the Food Consumption Score (FCS), measured over the past 7 days, and the Household Hunger Scale (HHS), measured over the past month;[11] (2) food expenditures (past month); (3) non-food expenditures (past month), and (4) non-agricultural income (past month). Using data from the endline, we analyze other primary outcomes including IPV, psychological well-being, and self-reported resilience, as well as outcomes to examine pathways.

2.2.4 Summary statistics and randomization check

Table 2.1 presents summary statistics and a randomization balance check. Columns 1 and 4 show the means and standard deviations of the control group in Liberia and Malawi, respectively. Columns 2 and 5 show the p-values for a test of equality between pooled treatment and control, and Columns 3 and 6 report

[11]The PAP also includes a fourth measure, the Food Insecurity Experience Scale (FIES), which we didn't measure in the phone surveys because it has a one-year recall. The 4 measures of food security have been shown to be correlated with food consumption in prior work and are validated for use in diverse settings. See FAO (2013), Ballard et al. (2011), Cafiero et al. (2018) and USAID (2019) for more details.

the p-values for an F-test of equality of means across the 3 sub-treatments ($250, $500, $750) and control. The underlying regressions control for strata fixed effects and cluster standard errors by village.

Because we targeted women for surveys, the sample skews female: 77% of the sample in Liberia and 94% in Malawi are women. Eighty-four percent are married in Liberia, and 67% in Malawi. The average age (about 40 years) is similar in the 2 countries. Education levels are low, averaging 2.9 years in Liberia and 4.8 years in Malawi. On average, households have 4.6-4.8 members. There are no significant differences across experimental arms.

Panel B shows primary outcomes as measured at baseline: food security,[12] food and non-food expenditures, assets, income, IPV, transfers, resilience to shocks, and agricultural input purchases. Overall, total monthly expenditures in the control group are $49 in Liberia and $33 in Malawi, with food expenditures being about 42% of the total. The total value of durables, livestock, and financial assets is $100 in Liberia and $90 in Malawi. Approximately 48% of women in Liberia and 32% in Malawi reported having experienced some form of IPV. Again, the sample is balanced in these attributes across treatment and control. Overall, randomization appears successful; in any case, all regressions were pre-specified as ANCOVA and we control for baseline measures throughout the paper.[13]

[12]The baseline and endline FSI also includes FIES. A z-score is calculated using inverse covariance weighting (Michael L. Anderson 2008b).

[13]As mentioned above, in Liberia we further randomized whether the cash was disbursed as a lump-sum or flow. This randomization was done "in the field" in collaboration with GD; IPA enumerators left chalk marks of different colors on the inside of respondents' doors. However, the match-rate of households between IPA and GD databases was only 87%. Since being matched is endogenous, we can perform the analysis only for the matched households. We check the balance between lump-sum and flow groups in Table 2.E1. Wave 1 is imbalanced on 1 variable (food expenditure), significant at 5%; Wave 2 is imbalanced on two variables (household size and transfers received), significant at 10%.

Table 2.1: Baseline Summary Statistics and Experimental Balance

	(1)	(2)	(3)	(4)	(5)	(6)
	Liberia			Malawi		
	Control Mean [SD]	p-value: pooled treatment = control	p-value: equality over 4 arms	Control Mean [SD]	p-value: pooled treatment = control	p-value: equality over 4 arms
Panel A. Demographics						
=1 if female	0.77	0.630	0.899	0.94	0.695	0.487
=1 if currently married or has partner	0.84	0.188	0.101	0.67	0.263	0.689
Age	39.14	0.998	0.995	40.45	0.607	0.879
	[13.92]			[15.08]		
Years of education	2.90	0.673	0.460	4.75	0.430	0.778
	[3.76]			[3.41]		
Number of household members	4.58	0.618	0.706	4.76	0.448	0.734
	[2.21]			[2.09]		
Panel B. Primary outcomes measured at baseline						
Food security index (z-score)	-0.00	0.992	0.002	0.00	0.445	0.833
	[1.00]			[1.00]		
Food expenditure (past month)	20.52	0.827	0.650	13.96	0.677	0.869
	[16.71]			[14.81]		
Non-food expenditure (past month)	28.04	0.307	0.730	19.18	0.800	0.569
	[28.56]			[21.66]		
Net value of durables, livestock, financial assets	101.94	0.796	0.936	89.74	0.334	0.796
	[207.58]			[187.75]		
Non-agricultural income (past month)	6.61	0.229	0.607	5.50	0.732	0.111
	[15.55]			[14.20]		
=1 if any IPV (past year)	0.48	0.929	0.224	0.32	0.921	0.726
Transfers received (USD, past month)	0.33	0.728	0.841	0.14	0.061	0.079
	[0.97]			[0.38]		
Transfers sent (USD, past month)	0.98	0.728	0.841	0.41	0.061	0.079
	[2.92]			[1.15]		
Resilience to shocks (z-score)	-0.00	0.135	0.414	0.00	0.431	0.767
	[1.00]			[1.00]		
Agricultural input purchase (USD, past year)	4.35	0.597	0.928	18.18	0.644	0.667
	[13.18]			[23.25]		
Observations	2,715			2,944		

Note: Columns 1 and 4 present the mean for the control groups; Columns 2 and 5 report the p-values for testing difference between the pooled cash treatment and control groups; Columns 3 and 6 report the p-values for testing difference across individual treatment arms by cash amounts (i.e. 250, 500, or 750 dollars) and the control group. Standard deviations are in square brackets in Columns 1 and 4 and standard error clustered at village level in parentheses in Columns 2,3,5 and 6. Monetary outcomes are in USD and winsorized at the 99th percentile. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

2.3 Results

2.3.1 Dynamic treatment effects

We estimate time-varying treatment effects using the phone survey data as follows:

$$Y_{ivst} = \sum_t \beta_t Cash_{vs} D_{tvs} + \gamma Y_{ivs0} + \phi_m + \lambda_s + \varepsilon_{ivst} \tag{2.1}$$

where Y_{ivst} is an outcome for individual i in village v and strata s at time t, which is defined as the number of months since cash transfers began (defined for each stratum, and thus taking on values for both treatment and control). $Cash_{vs}$ is a binary variable equal to 1 for villages assigned to any cash transfer, 0 otherwise; D_{tvs} is a binary variable indicating t number of months since transfers began; Y_{ivs0} is the baseline value of the outcome variable; and ϕ_m and λ_s are calendar month and strata fixed effects, respectively. We cluster standard errors by village, the level of randomization. We supplement results from this specification with a second set (shown in Appendix A) which pools surveys across all time periods.

Figure 2.1 plots the coefficient and confidence intervals from Equation 3.1 for the 4 main outcomes. Figures pool 2 months together so that the comparison across points in the graph is for the same set of respondents. In Malawi, we observe a spike in food security, food expenditures, and non-food expenditures immediately after disbursement, but no effect on non-agricultural income. For food security and expenditures, we observe clear evidence of time-varying treatment effects. For example, food security increased by over 0.5 standard deviations in the first six months (statistically significant at 1%) but then fell to approximately 0.2 standard deviations by the 8th month. These levels persist for the duration of the survey period (24-26 months after the initial transfer). For food

and non-food expenditures, treatment effects are substantial initially, but then fall to being indistinguishable from zero within 10-12 months. Non-agricultural income shows little effect throughout.

The picture is slightly different for Liberia. While there is also a clear increase in food security immediately post transfer, the magnitude is smaller, 0.25-0.4 standard deviations. This effect persists for the duration of surveying.[14] Food expenditures show little evidence of an increase, and effects on non-food expenditure are also fairly modest and indistinguishable from zero. There is no strong evidence of effects on non-agricultural income in Liberia either, though the results are noisier.[15]

2.3.2 Treatment effects at endline

To corroborate the findings from the phone surveys, we estimate the impact at endline using a similar specification:

$$Y_{ivs} = \beta Cash_{vs} + \gamma Y_{ivs0} + \lambda_s + \varepsilon_{ivs} \tag{2.2}$$

where Y_{ivs} is the value of the outcome at endline. Standard errors are clustered by village.

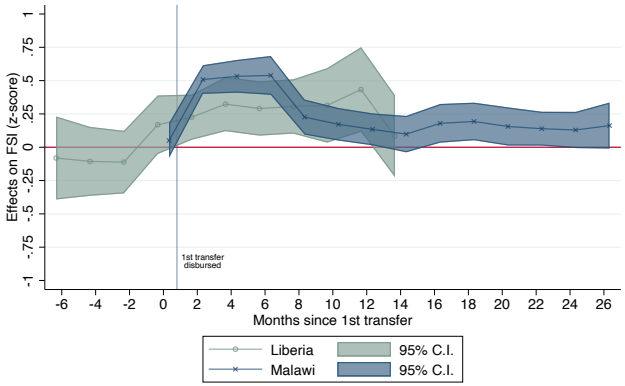
Table 2.2 shows effects on our main outcomes, with Panel A for Liberia and Panel B for Malawi. In the first row, we start by showing effects pooled across the different amounts. Consistent with the phone survey, we see significant im-

[14]Figure 2.1 suggests that food security outcomes in the treatment group improved in the round *before* the first payment. Although not statistically significant, it is possible that effects manifest before the recorded date because of measurement error in the date of the first GD transfer, or because treatment households could have started spending in anticipation.

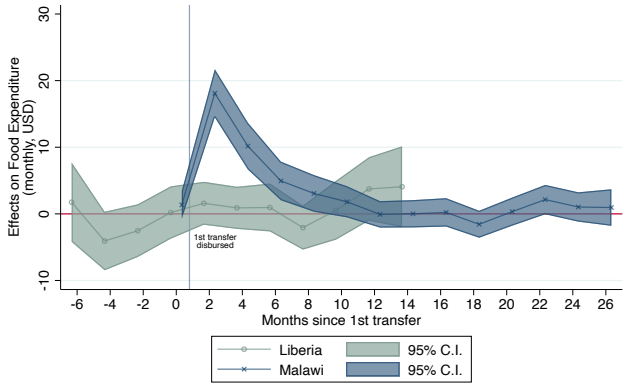
[15]Table 2.A3 shows pooled effects across rounds. This shows effects on food security, non-food expenditures, and transfers sent in both countries, and additionally, on non-food expenditures, and savings in Malawi.

Figure 2.1: Effects of Cash on Food Security, Expenditures and Income Over Time

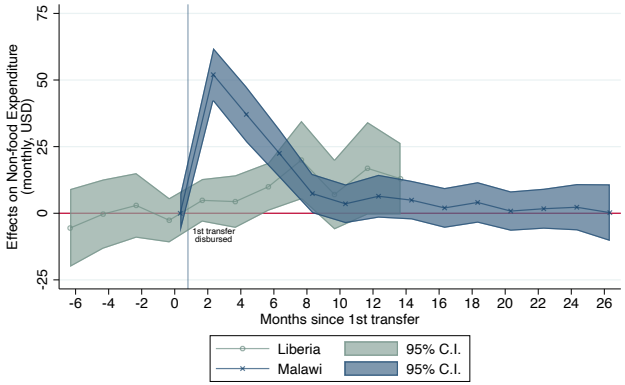
(a) Food Security Index[a]



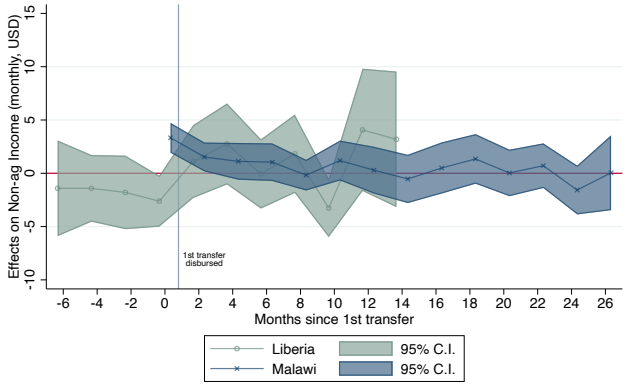
(b) Food Expenditures



(c) Non-food Expenditures



(d) Non-agricultural Income



Note: Regressions include baseline measurement of outcome and strata fixed effects. Standard errors clustered at village level. First transfer for each treatment household was made across July-October 2019 for Malawi and March-September 2020 for Liberia (Wave 2). Sample includes 596 households in Malawi and 358 in Liberia (Wave 2).

[a]Outcome variable is Food Security Index (FSI), a re-standardized z-score of HDDS, FCS, and HHS (negatively weighted) per Michael L. Anderson (2008b).

provements in food security. Households in treated villages had an FSI that was 0.31 standard deviations higher in Liberia and 0.12 standard deviations higher in Malawi, both statistically significant at 1 percent.[16] However, we see no evidence of an increase in food expenditures or non-agricultural income in either country (Columns 2-4). We do observe an increase in non-food expenditures in Liberia, and the effect is sizeable: a $5.9 increase on a $31.6 base. However, we find no such effect in Malawi. Overall, these results confirm the findings from the phone surveys, and validate them on a larger sample.[17]

Finally, while the average transfer amount was $500, villages were randomized into one of three cash transfer amounts. Table 2.2 also shows treatment effects by transfer size. Generally, for those outcomes which show pooled effects (Columns 1 and 3), effects tend (unsurprisingly) to be increasing in the transfer amount.

2.3.3 Effects on other outcomes

Table 2.A4 shows other primary pre-specified outcomes, specifically IPV, psychological well-being, and self-reported resilience. In Liberia, we find an 8 percentage point reduction in IPV (base of 38%), a 0.34 standard deviation increase in psychological well-being, and a 0.09 standard deviation increase in resilience. In Malawi, we find no effect on IPV, but a 0.10 standard deviation increase in

[16]Table 2.D1 shows clear evidence of improvements across index components: 3 of 4 measures (HDDS, HHS, and FIES) are statistically significant in Liberia, and 2 of 4 (HHS and FIES) in Malawi. The HHS and FIES focus on the quantity of food consumed (i.e. How many meals were skipped? How many days did the household go without food?), and cover a longer time period (a month and a year, respectively), whereas the HDDS and FCS measure the quality of consumption over the past day, and past week respectively. When pooled, both HDDS and FCS are in the expected direction.

[17]In Table 2.E3, we also show results for these main outcomes for the lump and flow randomization (see subsection 13 for details on this sub-treatment). We find no differences in outcomes by payment frequency.

Table 2.2: Treatment Effects at Endline (18-24 months post-disbursement)

	(1) Food Security Index[a] (past year)	(2) Food Expenditures (past month)	(3) Non-food Expenditures (past month)	(4) Non-agricultural Income[b] (past month)
Panel A. Liberia				
Pooled cash treatment:				
Cash	0.31***	0.68	5.91***	1.54
	(0.04)	(0.89)	(1.58)	(1.33)
Individual treatments by cash amount:				
Cash 250	0.19***	-1.13	2.64	2.57
	(0.06)	(1.17)	(2.67)	(2.55)
Cash 500	0.28***	1.87	7.20***	1.98
	(0.06)	(1.37)	(2.23)	(1.58)
Cash 750	0.47***	1.31	7.90***	0.07
	(0.06)	(1.11)	(1.99)	(1.89)
Control mean	0.00	26.91	31.63	8.15
Control SD	1.00	21.46	37.38	27.30
p-value (all three equal)	0.000	0.092	0.205	0.606
Observations	2,595	2,595	2,595	2,595
Panel B. Malawi				
Pooled cash treatment:				
Cash	0.12***	0.45	0.56	0.90
	(0.04)	(0.47)	(0.56)	(0.82)
Individual treatments by cash amount:				
Cash 250	0.06	0.40	0.09	2.39*
	(0.05)	(0.60)	(0.76)	(1.36)
Cash 500	0.12**	-0.09	0.53	-0.71
	(0.05)	(0.65)	(0.78)	(0.96)
Cash 750	0.17***	1.04	1.07	1.01
	(0.06)	(0.75)	(0.88)	(1.13)
Control mean	0.00	9.56	12.29	9.38
Control SD	1.00	10.81	14.63	20.43
p-value (all three equal)	0.268	0.429	0.642	0.082
Observations	2,784	2,784	2,784	2,784
Panel C. Pooled				
Pooled cash treatment:				
Cash	0.21***	0.56	3.17***	1.17
	(0.03)	(0.49)	(0.83)	(0.77)
Individual treatments by cash amount				
Cash 250	0.13***	-0.32	1.36	2.36*
	(0.04)	(0.65)	(1.35)	(1.42)
Cash 500	0.21***	0.85	3.79***	0.49
	(0.04)	(0.75)	(1.16)	(0.90)
Cash 750	0.30***	1.17*	4.39***	0.65
	(0.04)	(0.66)	(1.10)	(1.10)
Control mean	0.00	17.98	21.68	8.78
Control SD	1.00	18.94	29.69	24.02
p-value (all three equal)	0.002	0.134	0.144	0.469
Observations	5,379	5,379	5,379	5,379

Note: The endline was conducted about 18-22 months after first transfers were received in Liberia and 21-25 months in Malawi. Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Monetary outcomes are in USD and Winsorized at the 99th percentile. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.
[a] Food Security Index is standardized z-score of HDDS, FCS, HHS (negatively weighted), and FIES (negatively weighted), using inverse covariance weighting (Michael L. Anderson 2008b) relative to the control mean and SD in each country.

psychological well-being and a 0.12 standard deviation increase in resilience.[18]

In addition, we pre-specified several secondary outcomes, which we present exhaustively in Appendix B. We discuss the most striking ones here. Table 2.B1, shows a decrease in casual labor in both countries (a 10 percentage point increase on a 32% base in Liberia, and a 4 percentage point decrease on a 43% base in Malawi). While we are unable to track where this time went (we did not measure other types of labor), this result is similar to Fink et al. (2020), where credit reduces casual labor in favor of on-farm labor, and Aggarwal et al. (forthcoming), where savings accounts given to entrepreneurs reduce labor supply in the primary business and increase agricultural labor. Both results are consistent with casual or employment labor being used as a way to generate cash for day-to-day expenses, and for financial interventions to reduce the need for such cash.

Table 2.B2 shows effects on school enrollment and school investment. We find sizeable effects in Liberia, where enrollment increased by 10 percentage points, education expenditure by about 38%; and school attendance improved. In Malawi, we find no such effects, presumably because school enrollment is dramatically higher than Liberia (only 52% of school-aged children in Liberia were enrolled at the endline, compared to 94% in Malawi).

There is some evidence of increased health investment in Liberia, specifically the proportion of people sleeping under a bednet increased (Table 2.B3). Households in Liberia take more days off per illness (perhaps because they can afford to forego the income), and spend more to treat it (Table 2.B4). Finally, child anthropometrics (Table 2.B5, measured in Malawi only), social capital (Table 2.B6)

[18]For IPV, we cross-randomized audio computer-assisted self-interviewing (ACASI) and conventional, face-to-face interviewing (FTFI). Our evidence suggests that a significant portion of the sample is making mistakes in the ACASI module (Park et al. 2022a), thus this analysis is restricted to the FTFI sample.

and public goods investment (Table 2.B7) remain unimpacted.

2.3.4 Pathways

We find that improvements in food security persist for up to 2 years post-disbursement, although expenditures converge to the control group within 6 months and we observe no evidence of an increase in income at any point. These results imply that households must increase productive investment in some fashion. In this section, we investigate what those channels are. The main results are shown in Table 2.3.

We find increases in financial and non-financial assets, with larger effects on non-financial assets. Cash savings increase by about $6 in Liberia and $3 in Malawi, which are large effects relative to the control mean but still are small absolute sums. By contrast, we see that durable goods increase by about $41 in Liberia and $18 in Malawi, equivalent to about 76% and 29% increases compared to the control group. We also find an increase in animal assets, though only significant in Malawi (an increase of $11 on a base of $49, i.e. 22%).[19]

Another potential channel is an increase in agricultural output, which we show in Columns 5-7. In both countries, we see large increases in the value of harvest (Column 7). We see no effect on intermediate inputs in Liberia (Column 5), but we see a $2.7 increase in Malawi ($16.7 base). We also observe increases in farm tools in both countries. However, the effects on output seem large relative to these relatively modest effects on investment. To explore this further, we turn to Table 2.A6, which shows evidence that households switched crops, particularly to grow more legumes, and were slightly more likely to plant a staple crop. Table 2.A7

[19]In Table 2.A5, we show disaggregated results for the number and value of animals (by type), and find effects for goats and chickens in both countries.

Table 2.3: Pathways

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Asset values				Agriculture (past season)			Self-own Business (past month)	
	Savings	Debt	Durables	Livestock	Input purchase	Value of farm tools	Harvest value of all crops	Value of capital	Revenue
Panel A. Liberia									
Pooled cash treatment:									
Cash	6.29**	-0.20	40.78***	3.93	-0.17	2.34***	83.86***	4.07***	2.32
	(2.72)	(0.87)	(7.92)	(3.44)	(1.21)	(0.71)	(28.63)	(1.38)	(1.42)
Individual treatments by cash amount:									
Cash 250	4.52	-0.67	31.33***	3.65	-2.66**	1.24	107.49**	5.58***	2.00
	(3.59)	(1.14)	(11.48)	(4.43)	(1.29)	(0.91)	(48.20)	(2.12)	(1.92)
Cash 500	9.76**	0.23	52.13***	3.70	0.60	2.65**	65.50	4.38**	4.00*
	(4.65)	(1.34)	(13.73)	(4.79)	(1.41)	(1.08)	(39.93)	(2.03)	(2.35)
Cash 750	4.58	-0.18	38.85***	4.44	1.54	3.14***	78.63*	2.23	0.95
	(3.47)	(1.19)	(11.81)	(5.05)	(1.93)	(1.12)	(41.76)	(1.54)	(1.68)
Control mean	18.68	5.25	54.13	38.11	4.81	11.47	266.01	6.46	7.03
Control SD	64.44	19.82	154.30	92.96	21.67	13.78	590.66	32.68	31.04
p-value (all three equal)	0.558	0.835	0.473	0.989	0.007	0.285	0.766	0.264	0.466
Observations	2,595	2,595	2,595	2,595	2,595	2,595	2,595	2,595	2,595
Panel B. Malawi									
Pooled cash treatment:									
Cash	3.08**	-0.16	17.94***	11.12**	2.73***	0.76*	24.69***	-0.92	-0.10
	(1.30)	(0.87)	(4.87)	(4.68)	(0.71)	(0.45)	(8.73)	(0.70)	(0.51)
Individual treatments by cash amount:									
Cash 250	1.97*	0.25	8.31	-2.17	1.74*	-0.09	29.68**	-1.30	0.25
	(1.05)	(1.08)	(6.22)	(5.52)	(0.97)	(0.62)	(12.03)	(0.86)	(0.74)
Cash 500	0.54	-1.62	11.49**	18.52***	2.91***	0.98*	9.56	-1.37	-0.55
	(0.88)	(1.05)	(5.68)	(6.50)	(0.96)	(0.56)	(11.82)	(0.94)	(0.70)
Cash 750	6.88*	0.92	34.67***	17.23**	3.56***	1.42*	35.28**	-0.05	0.01
	(3.53)	(1.33)	(7.91)	(7.64)	(1.11)	(0.74)	(14.04)	(1.05)	(0.75)
Control mean	6.28	7.47	63.39	48.59	16.65	9.51	111.23	5.02	3.71
Control SD	16.42	19.66	142.10	120.52	17.65	10.17	172.81	20.93	15.09
p-value (all three equal)	0.137	0.125	0.008	0.006	0.354	0.163	0.220	0.467	0.644
Observations	2,784	2,784	2,784	2,784	2,784	2,784	2,784	2,784	2,784
Panel C. Pooled									
Pooled cash treatment:									
Cash	4.63***	-0.14	29.05***	8.35***	1.31*	1.53***	53.97***	1.52*	1.03
	(1.48)	(0.62)	(4.64)	(2.99)	(0.70)	(0.42)	(14.71)	(0.77)	(0.74)
Individual treatments by cash amount									
Cash 250	3.19*	-0.15	19.43***	1.37	-0.37	0.56	66.45***	2.02*	0.97
	(1.81)	(0.78)	(6.42)	(3.58)	(0.81)	(0.54)	(24.17)	(1.14)	(1.00)
Cash 500	4.95**	-0.74	31.10***	12.41***	1.80**	1.79***	36.75*	1.42	1.52
	(2.33)	(0.86)	(7.45)	(4.23)	(0.86)	(0.60)	(20.41)	(1.10)	(1.21)
Cash 750	5.78**	0.48	36.79***	11.32**	2.54**	2.26***	58.85***	1.10	0.57
	(2.49)	(0.90)	(7.06)	(4.75)	(1.11)	(0.67)	(21.45)	(0.94)	(0.91)
Control mean	12.30	6.39	58.90	43.51	10.90	10.46	186.37	5.72	5.32
Control SD	46.82	19.77	148.19	108.13	20.57	12.10	436.62	27.28	24.23
p-value (all three equal)	0.606	0.519	0.126	0.032	0.008	0.057	0.541	0.757	0.765
Observations	5,379	5,379	5,379	5,379	5,379	5,379	5,379	5,379	5,379

Note: All outcomes are in USD and Winsorized at the 99th percentile. The endline was conducted about 18-22 months after first transfers were received in Liberia and 21-25 months in Malawi. Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

shows an increase in the harvest output and value of staple crops (Columns 1-2) as well as in value of non-staple (Column 3) and all crops (Column 4). In Liberia, most of the increase in the value of output comes via staple crops, while in Malawi the increases are comparable between staple and non-staple crops.

Columns 8-9 show the effects on business investment. We observe a $4 increase in business capital in Liberia, on a small base of $6.5; and no effect in Malawi. We also see a positive, but insignificant, change in revenue. While we previously showed no effect on non-agricultural income, including from business, it is possible that income is imperfectly measured, for example, because businesses were operated primarily by men. While we asked about total household income, female respondents might have underestimated male income.

Another possible channel is that households used the cash to purchase large quantities of storable commodities, such as maize or oil, and then used these over the two-year period. At endline, we measured self-reports and uses of large cash transfers (only available for Malawi). Table 2.A8 shows that most of the received transfers were cashed out. Table 2.A9 tabulates spends. Nearly half were spent on home repair or construction, and about $33 on food. Given that the monthly food expenditure of the control group in Malawi is $10 (Table 2.2), this does not seem likely to explain the persistent improvements in food security for 2 years.

2.3.5 Spillover effects

Prior work has shown spillover effects of cash transfers on non-beneficiaries, either through direct sharing (Angelucci and De Giorgi 2009) or via general equilibrium effects (Egger et al. 2019 and Filmer et al. forthcoming). In this section, we test for spillovers in our setting.

First, we present data on transfers and labor demand, reported by beneficiaries themselves in Figure 2.A4 (phone survey) and Table 2.A10 (endline). There is no evidence of an increase in transfers, even immediately post-disbursement.[20] There is no evidence of increased labor demand either: in Liberia, households are slightly more likely to hire casual labor, but the quantity of labor employed is small and insignificant.

Second, in Malawi and in Liberia Wave 2, we randomized the intensity of treatment, assigning clusters of villages to high-intensity, low-intensity, or pure control.[21] Using this design, we estimate spillovers as follows:

$$Y_{ivcs} = \beta Treat_c + \delta Cash_{vs} + \gamma Y_{ivs0} + \lambda_s + \varepsilon_{ivcs} \tag{2.3}$$

where the added subscript c indicates cluster, and errors are clustered by cluster. $Treat_c$ is an indicator for being a control household in a treated cluster.[22]

Results are presented in Table 2.4. In Columns 1-2, we see no evidence of increase in transfers. The point estimates on both transfers received and sent are negative (and significant at 10%) in Malawi, and negative but insignificant in Liberia. Columns 3-6 show the 4 main outcomes; none show statistically significant effects. However, because of the small number of clusters and the fact that this analysis does not include Liberia Wave 1, the confidence intervals include

[20] In fact, we find some weak evidence of increases in the amount of transfers *received*.

[21] In Malawi, we used an existing administrative unit "group village." Out of 104 group villages in our study sample, 49 were assigned to high-intensity (in which about 3/4 of villages were treated), 23 to low-intensity (in which about half of villages were treated), and 32 to pure control (in which no villages were treated). In Liberia, there is no comparable administrative unit, so we identified geographical clusters of villages using average distances between housing structures. While it was not possible to stratify the sample in this way for Wave 1, we were able to do this for Wave 2. We identified 70 village clusters, assigning 34 to high-intensity (in which about 80% of villages were treated), 22 to low-intensity (in which about a third were treated), and 14 to pure control (in which no villages were treated).

[22] High- and low-intensity clusters are pooled for this analysis, since sample sizes in each are small and there are no statistically significant differences in outcomes.

fairly substantial values (for example, the lower bound on food security is -0.25 SDs in Liberia and -0.10 standard deviations in Malawi; there is a similar pattern for most variables). Columns 7-8 follow our earlier analysis and show IPV and psychological well-being. We see weak evidence of effects. In particular, psychological well-being is 0.16 standard deviations higher in Liberia, and the coefficient on IPV is negative and borderline significant. We interpret these results as most likely due to statistical noise.

In sum, we find little evidence of spillovers, which is in contrast to other work: Angelucci and De Giorgi (2009) find that *Oportunidades* cash transfers increased consumption of the control households, which the authors attribute to sharing. One possible explanation for this contrast is due to program design: since *Oportunidades* had within-village targeting, the pressures to share may have been stronger. Similarly, Haushofer and Shapiro (2018), another study that uses within-village randomization, finds that non-beneficiaries in treatment villages lower their consumption, likely because they sell off productive assets. On the other hand, the transfers that we evaluate were universal within village.

2.3.6 Prices

Prior research has shown conflicting results on the effect of cash transfers on prices, with some showing minimal inflationary effects (i.e. Cunha et al. 2018, Aker et al. 2016 and Egger et al. 2019) and another study showing larger price increases (Filmer et al. forthcoming). Our study was designed to minimize price effects: randomization was spread out over a wide enough geographic area such that a small proportion of the overall population in any market catchment area was treated. In Liberia, we estimate that no more than 13% of the population

Table 2.4: Spillover Effects

	(1) Transfers Sent (past month)	(2) Transfers Received (past month)	(3) Food Security[a] (past year)	(4) Food Expend (past month)	(5) Non-food Expend (past month)	(6) Non-ag Income (past month)	(7) Any IPV[b] (past year)	(8) Psycho Well-being (past 2 weeks)
Panel A. Liberia								
Treat cluster	-1.16	-1.72	-0.08	-0.46	-2.88	-0.15	-0.08	0.16
	(0.80)	(1.06)	(0.08)	(1.88)	(2.50)	(2.74)	(0.05)	(0.10)
Cash village	-0.12	-0.04	0.33***	0.28	8.47***	2.66	-0.06	0.29***
	(0.46)	(0.63)	(0.06)	(1.20)	(2.30)	(2.90)	(0.04)	(0.05)
Pure control mean	3.31	4.04	0.03	29.61	34.80	8.09	0.40	-0.03
Pure control SD	16.90	19.55	0.98	22.03	37.36	19.79	0.49	1.04
Observations	1,867	1,867	1,867	1,867	1,867	1,867	860	1,867
Panel B. Malawi								
Treat cluster	-0.02	-0.44*	0.02	1.05	0.17	-0.24	0.01	0.06
	(0.08)	(0.24)	(0.07)	(0.66)	(0.82)	(1.06)	(0.02)	(0.06)
Cash village	0.02	0.37*	0.11**	-0.04	0.48	1.01	0.00	0.07
	(0.07)	(0.20)	(0.05)	(0.53)	(0.86)	(0.90)	(0.02)	(0.05)
Pure control mean	0.46	1.22	-0.01	9.08	12.08	9.59	0.19	-0.04
Pure control SD	1.96	4.89	0.99	9.80	14.40	21.56	0.39	1.03
Observations	2,784	2,784	2,784	2,784	2,784	2,784	1,829	2,784
Panel C. Pooled								
Treat cluster	-0.45	-0.93**	-0.03	0.47	-1.16	-0.21	-0.01	0.09
	(0.31)	(0.46)	(0.05)	(0.83)	(1.16)	(1.28)	(0.02)	(0.06)
Cash village	-0.05	0.20	0.20***	0.09	3.81***	1.61	-0.02	0.16***
	(0.20)	(0.29)	(0.04)	(0.59)	(1.21)	(1.33)	(0.02)	(0.04)
Pure control mean	1.49	2.25	0.01	16.55	20.34	9.04	0.26	-0.04
Pure control SD	10.39	12.48	0.98	18.30	27.54	20.94	0.44	1.03
Observations	4,651	4,651	4,651	4,651	4,651	4,651	2,689	4,651

Note: The endline was conducted about 18-22 months after first transfers were received in Liberia and 21-25 months in Malawi. Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Monetary outcomes are in USD and Winsorized at the 99th percentile. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

[a] Food Security Index is standardized z-score of HDDS, FCS, HHS (negatively weighted), and FIES (negatively weighted), using inverse covariance weighting (Michael L. Anderson 2008b) relative to the control mean and SD in each country.

[b] Includes only women and those for whom IPV was measured in face-to-face interviewing.

was treated in any market catchment area; in Malawi, it was about 7%.

As discussed earlier, we set up a data-collection protocol to identify causal price changes. Starting just before cash disbursal, we collected monthly data on prices (in 23 treatment and 57 comparison markets in Liberia, and 10 treatment and 85 comparison in Malawi) for a representative basket of food items based on each country's most recent LSMS.[23] From each market, we enrolled 2 vendors per item for surveying, calling them once a month throughout the study.[24] We classify control markets as being "nearby" or "distant", based on the median distance to the nearest treatment market.[25]

In Appendix C, we plot price changes relative to the pre-transfer period for (a) staple grains and vegetables ("select items"), and (b) staple grains. We find no evidence of increased prices due to the cash transfers. For Liberia, Figure 2.C1 and Figure 2.C2 show the simple average change in weighted prices after the transfer, relative to the pre-transfer period, for each wave separately. Overall, prices remain fairly stable throughout the entire period, and follow similar patterns in treatment, nearby, and distant markets. The situation is similar for Malawi, in that price changes, while substantial, are not differential by treatment status.

[23]These items are cassava, cassava flour, imported rice, okra, onion, palm oil, and salt for Liberia; beans, groundnut, maize flour, maize kernel, onion, salt, sugar, sweet potato, tomato, and unpacked rice for Malawi. Altogether, these accounted for about 80% of the food expenditure in each country at baseline.

[24]Vendors received airtime credit (worth \$0.50-\$1 per call) for answering the survey. In the survey, we asked about all items on the list that the vendor sold herself, as well as prices of items she did not sell. We use these "indirect" prices only in cases where the direct price is entirely missing for that market-month pair.

[25]The median distance to the nearest treated market is 48 km in both countries.

2.4 Conclusion

We use high-frequency panel data to measure the evolution and persistence of the effect of UCTs, focusing specifically on food security, expenditures, and income. We find lasting effects of transfers on food security (up to 1 year in Liberia and 2 years in Malawi), though we find fleeting effects on expenditures and no effect on non-agricultural income. We present suggestive evidence that investment in productive capital may be a likely pathway to generate these effects.

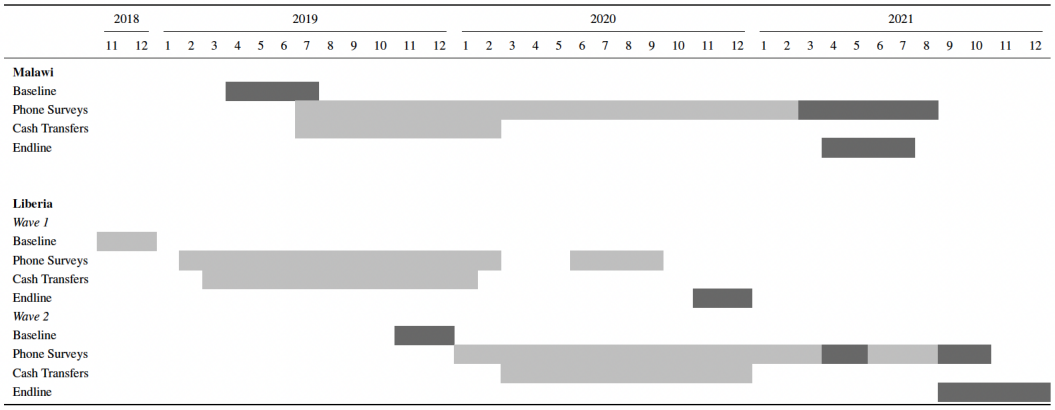
Our results confirm that the measurement of treatment effects of interventions such as cash will vary over time. While the size of estimated effects attenuates substantially after a few months, our results suggest that (in Malawi at least) effects on food security are similar and stable from 8 months on, suggesting that the somewhat arbitrary timing of many endlines around a year from disbursement may actually meaningfully capture lasting food security.[26] An open question for future work is what the effects become in the next period of time, and whether they may very well go to null at a future date.

[26]For example, the endlines in Haushofer and Shapiro (2016) and McIntosh and Zeitlin (2021) were 9 and 13 months after the final transfers respectively. The first endline in Blattman et al. (2014) was at 2 years, a timeline similar to ours.

2.5 Appendix

Appendix A

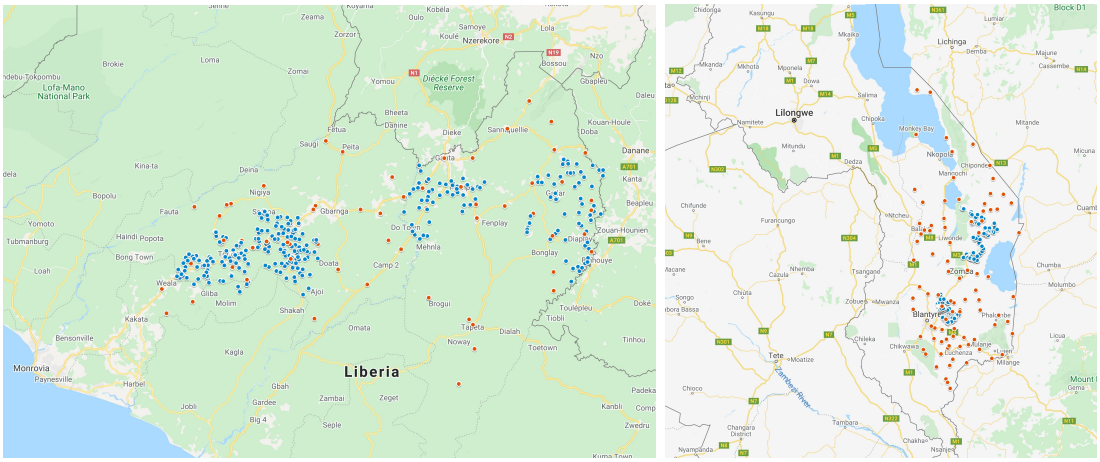
Figure 2.A1: Timeline of Cash Transfer Disbursements and Survey Activities



Note: Darker grey blocks indicate the survey rounds where module order randomization was conducted and thus data for which are included for analysis in this paper.

Figure 2.A2: Map of Study Villages and Markets in Liberia and Malawi

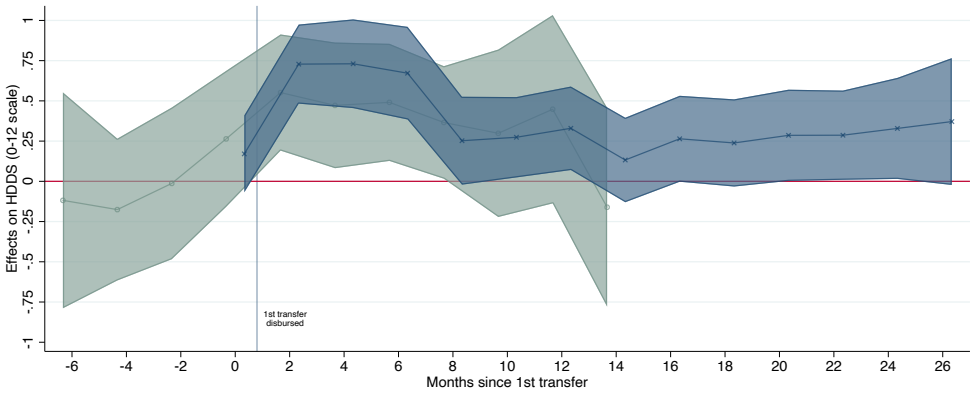
(a) Liberia **(b)** Malawi



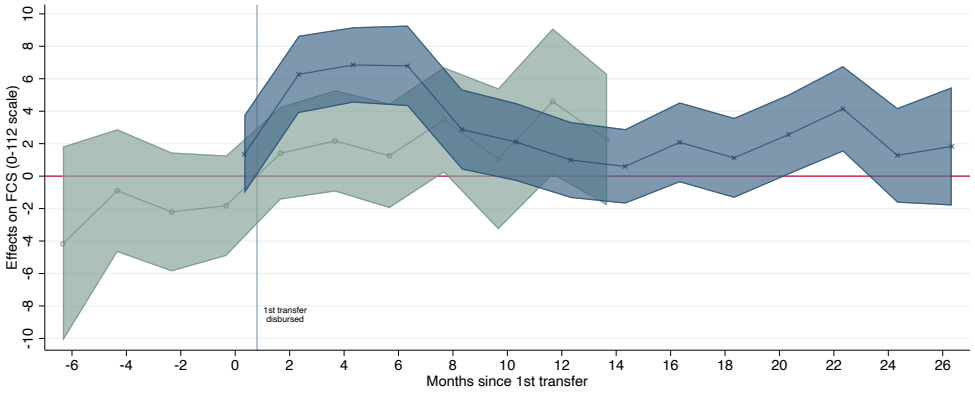
Note: Blue dots refer to villages, and orange dots markets. For Liberia, there are 300 villages and 80 markets. For Malawi, there are 300 villages and 95 markets.

Figure 2.A3: Effects on Individual Components of Food Security Index (HDDS, FCS, and HHS)

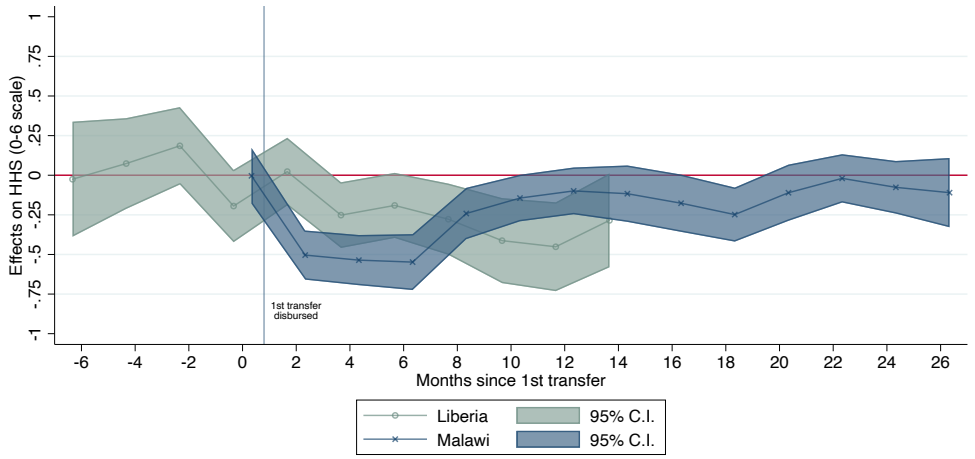
(a) Household Dietary Diversity Score (baseline control mean = 5.7, 5.4)



(b) Food Consumption Score (baseline control mean = 48.0, 46.3)



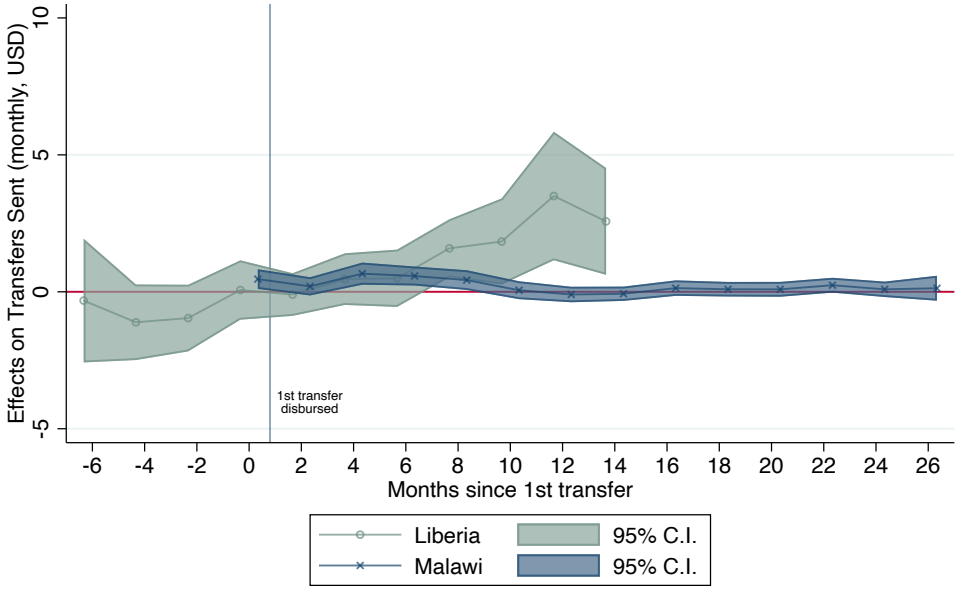
(c) Household Hunger Scale (baseline control mean = 1.2, 1.2)



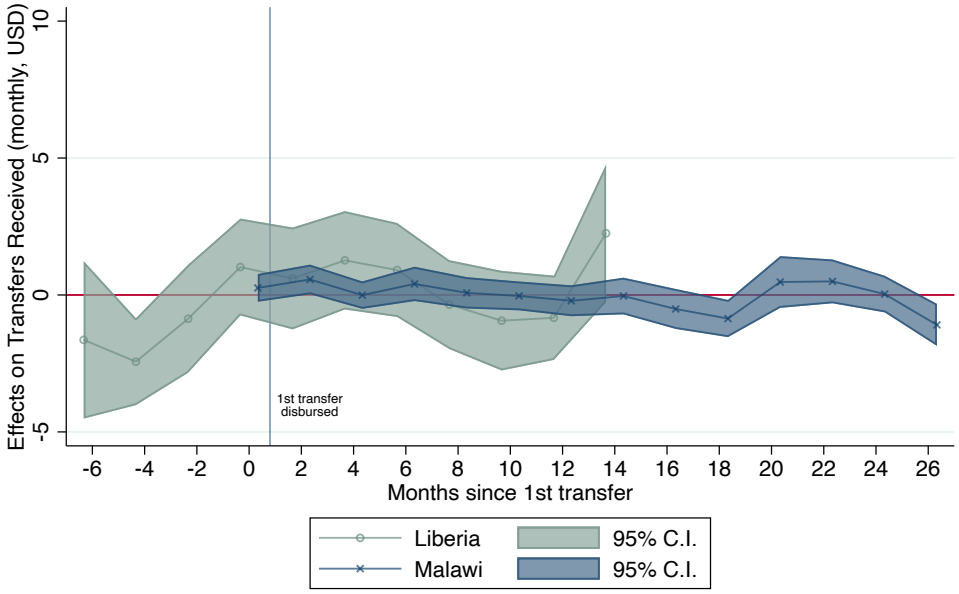
Note: Regressions include baseline measurement of outcome and strata fixed effects. Standard errors clustered at village level. First transfer for each treatment household was made across July-October 2019 for Malawi and March-September 2020 for Liberia (Wave 2). Sample includes 596 households in Malawi and 358 in Liberia (Wave 2).

Figure 2.A4: Transfers

(a) Transfers Sent



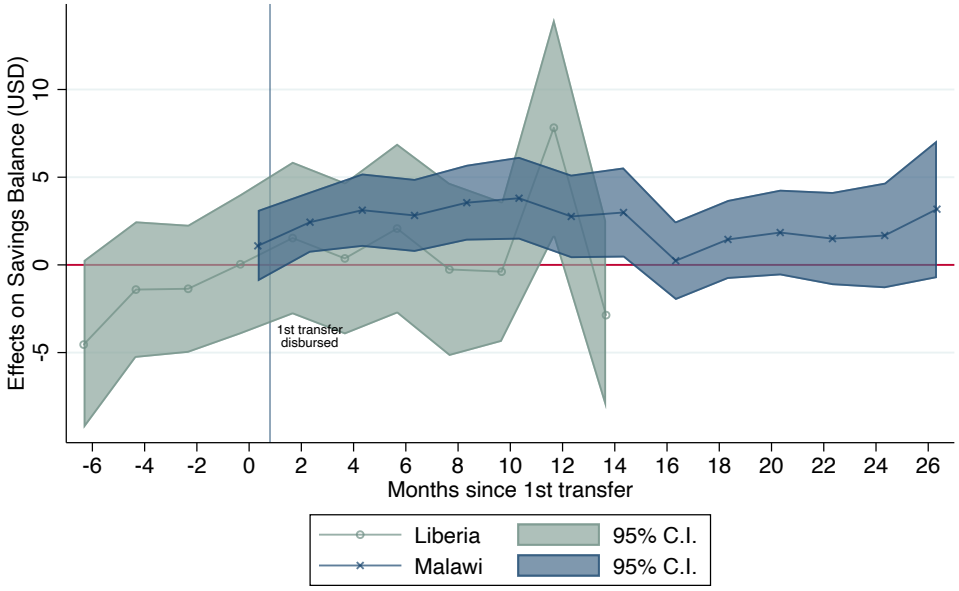
(b) Transfers Received



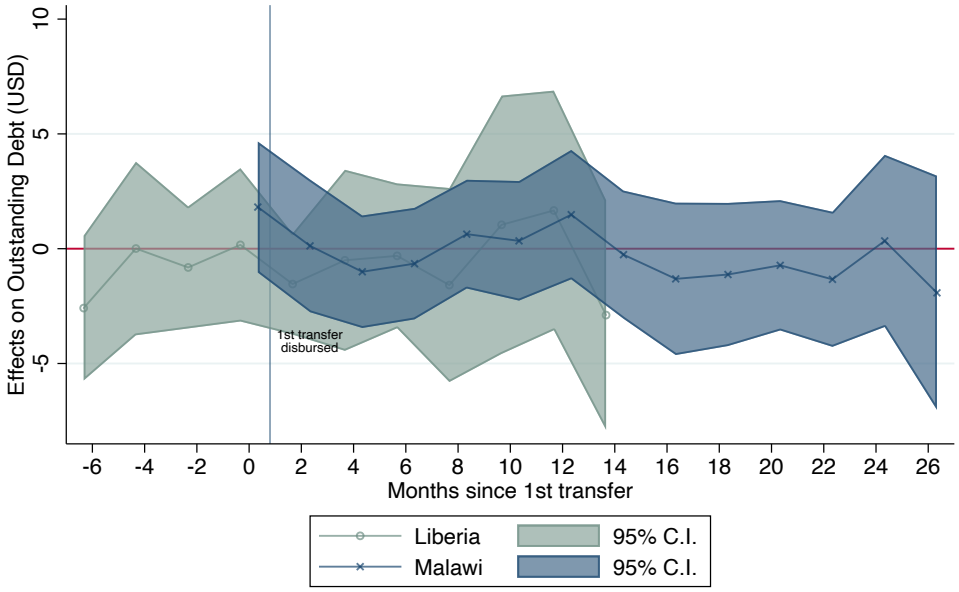
Note: Data comes from phone surveys, and transfers are measured ovr the past month. Regressions include baseline measurement of outcome and strata fixed effects. Standard errors clustered at village level. First transfer for each treatment household was made across July-October 2019 for Malawi and March-September 2020 for Liberia (Wave 2). Sample includes 596 households in Malawi and 358 in Liberia (Wave 2).

Figure 2.A5: Savings and Debt

(a) Savings Balance

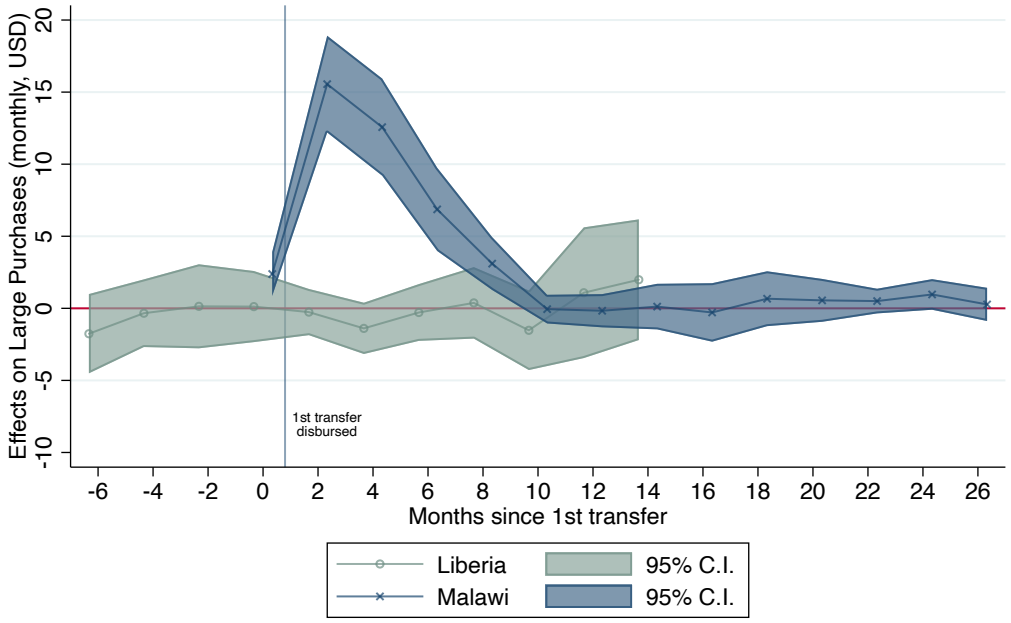


(b) Debt Amount



Note: Regressions include baseline measurement of outcome and strata fixed effects. Standard errors clustered at village level. First transfer for each treatment household was made across July-October 2019 for Malawi and March-September 2020 for Liberia (Wave 2). Sample includes 596 households in Malawi and 358 in Liberia (Wave 2).

Figure 2.A6: Large Purchases



Note: Regressions include baseline measurement of outcome and strata fixed effects. Standard errors clustered at village level. First transfer for each treatment household was made across July-October 2019 for Malawi and March-September 2020 for Liberia (Wave 2). Sample includes 596 households in Malawi and 358 in Liberia (Wave 2).

Table 2.A1: Attrition in Endline Survey

	(1)	(2)	(3)	(4)
	=1 if completed endline survey		=1 if completed IPV survey at endline[a]	
	Liberia	Malawi	Liberia	Malawi
Cash	-0.00	0.01	0.02	0.04*
	(0.01)	(0.01)	(0.02)	(0.02)
Control mean	0.96	0.94	0.69	0.66
Overall mean	0.96	0.95	0.70	0.68
Observations	2,715	2,944	2,595	2,784

Note: Regressions include strata fixed effects. Standard errors clustered at village level in parentheses.
[a] Sample restricted to female respondents.

Table 2.A2: Attrition in Phone Surveys

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	=1 if completed survey in following survey round													=1 if	% of
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	13th	$\geq$ 1R	rounds
Panel A. Malawi															
Cash	0.01	-0.03	-0.03	0.01	-0.03	-0.02	-0.04*	-0.03	-0.02	0.04	-0.00	-0.00	-0.00		-0.01
	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)		(0.02)
Control mean	0.97	0.95	0.87	0.91	0.93	0.95	0.96	0.94	0.89	0.84	0.86	0.80	0.62	1.00	0.88
Overall mean	0.97	0.94	0.85	0.91	0.92	0.93	0.94	0.93	0.87	0.86	0.86	0.80	0.61	1.00	0.88
Observations	596	596	596	596	596	596	596	596	596	596	596	596	596	596	596
Panel B. Liberia (Wave 2)															
Cash	-0.01	-0.01	-0.03	-0.05	-0.04	0.02	-0.03	-0.07	0.02					-0.06*	-0.02
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)					(0.04)	(0.04)
Control mean	0.75	0.68	0.70	0.71	0.64	0.60	0.55	0.48	0.45					0.90	0.62
Overall mean	0.74	0.68	0.69	0.69	0.62	0.62	0.54	0.44	0.46					0.87	0.61
Observations	416	416	416	416	416	416	416	416	416					416	416

Note: Each survey round is two months, where half of the sample is called in the even month and the other in the odd month. Regressions include strata fixed effects. Standard errors clustered at village level in parentheses.

Table 2.A3: Pooled treatment effects (phone surveys)

	(1) Food Security[a] (past year)	(2) Food Expend (past month)	(3) Non-food Expend (past month)	(4) Non-ag Income[b] (past month)	(5) Transfers Sent (past month)	(6) Transfers Received (past month)	(7) Savings Balance	(8) Outstanding Loan
Panel A. Liberia								
Pooled cash treatment:								
Cash	0.17***	0.59	7.33***	-0.05	0.57**	0.16	0.45	-0.00
	(0.05)	(0.95)	(2.66)	(0.80)	(0.28)	(0.47)	(0.63)	(0.74)
Individual treatments by cash amount:								
Cash 250	0.04	-0.52	0.71	-0.48	-0.09	-0.58	-0.70	-0.49
	(0.07)	(1.46)	(3.83)	(1.03)	(0.32)	(0.59)	(0.73)	(0.91)
Cash 500	0.29***	1.24	8.35**	1.62	0.72*	0.43	1.68*	0.04
	(0.08)	(1.19)	(3.51)	(1.08)	(0.43)	(0.71)	(0.97)	(1.12)
Cash 750	0.18**	0.96	12.66***	-1.52	1.05**	0.59	0.17	0.43
	(0.07)	(1.49)	(4.16)	(1.13)	(0.47)	(0.62)	(0.87)	(1.04)
Control mean	0.28	21.76	33.46	8.74	2.54	4.64	4.14	4.50
Control SD	0.87	16.99	37.62	14.19	5.35	8.53	13.82	15.61
No. of respondents	497	497	497	497	497	497	497	497
Observations	2,925	2,925	2,925	2,925	2,925	2,925	2,925	2,925
Panel B. Malawi								
Pooled cash treatment:								
Cash	0.23***	2.98***	9.86***	0.27	0.19**	-0.03	2.15***	-0.91
	(0.04)	(0.70)	(1.54)	(0.65)	(0.08)	(0.20)	(0.82)	(0.86)
Individual treatments by cash amount:								
Cash 250	0.15***	2.59**	8.03***	1.42	0.09	-0.08	1.36	0.28
	(0.05)	(1.05)	(2.31)	(1.05)	(0.11)	(0.22)	(1.36)	(1.35)
Cash 500	0.26***	1.76*	8.94***	-1.00	0.06	0.10	1.87*	-1.63
	(0.07)	(0.92)	(2.37)	(0.76)	(0.10)	(0.36)	(1.01)	(1.16)
Cash 750	0.28***	4.65***	12.72***	0.38	0.42***	-0.10	3.26***	-1.39
	(0.06)	(0.95)	(2.02)	(0.96)	(0.14)	(0.22)	(1.25)	(1.11)
Control mean	0.07	15.15	26.35	9.39	0.51	1.13	7.04	10.23
Control SD	0.91	13.69	32.45	12.99	1.87	4.34	13.71	18.98
No. of respondents	596	596	596	596	596	596	596	596
Observations	6,781	6,784	6,784	6,784	6,784	6,784	6,784	6,784
Panel C. Pooled								
Pooled cash treatment:								
Cash	0.21***	2.25***	9.15***	0.17	0.30***	0.03	1.55**	-0.63
	(0.03)	(0.57)	(1.34)	(0.51)	(0.10)	(0.20)	(0.63)	(0.64)
Individual treatments by cash amount								
Cash 250	0.12***	1.51*	5.57***	0.88	0.04	-0.23	0.95	0.11
	(0.04)	(0.87)	(2.00)	(0.81)	(0.12)	(0.23)	(1.01)	(1.00)
Cash 500	0.28***	1.62**	9.07***	-0.17	0.27*	0.23	1.51*	-1.14
	(0.05)	(0.74)	(1.97)	(0.64)	(0.15)	(0.34)	(0.85)	(0.88)
Cash 750	0.25***	3.69***	12.94***	-0.19	0.60***	0.09	2.22**	-0.84
	(0.05)	(0.81)	(1.89)	(0.76)	(0.17)	(0.24)	(0.96)	(0.85)
Control mean	0.14	17.18	28.54	9.19	1.13	2.21	6.15	8.46
Control SD	0.90	15.10	34.28	13.38	3.48	6.17	13.80	18.20
No. of respondents	1,093	1,093	1,093	1,093	1,093	1,093	1,093	1,093
Observations	9,706	9,709	9,709	9,709	9,709	9,709	9,709	9,709

Note: Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Monetary outcomes are in USD and Winsorized at the 99th percentile. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

Table 2.A4: Effects of Cash Transfers on IPV, Psychological Well-Being, and Self-reported Resilience

	(1) Any IPV[a] (past year)	(2) Psychological Well-being (past 2 weeks)	(3) Resilience (past year)
Panel A. Liberia			
Pooled cash treatment:			
Cash	-0.08***	0.34***	0.09**
	(0.03)	(0.04)	(0.04)
Individual treatments by cash amount:			
Cash 250	-0.05	0.28***	0.14***
	(0.04)	(0.06)	(0.05)
Cash 500	-0.10***	0.36***	0.07
	(0.04)	(0.06)	(0.05)
Cash 750	-0.08**	0.37***	0.06
	(0.04)	(0.05)	(0.06)
Control mean	0.38	-0.00	0.00
Control SD	0.49	1.00	1.00
p-value (all three equal)	0.560	0.402	0.403
Observations	1,229	2,595	2,595
Panel B. Malawi			
Pooled cash treatment:			
Cash	0.01	0.10**	0.12***
	(0.01)	(0.04)	(0.04)
Individual treatments by cash amount:			
Cash 250	0.01	0.04	0.10*
	(0.02)	(0.06)	(0.06)
Cash 500	0.01	0.11*	0.11**
	(0.02)	(0.06)	(0.05)
Cash 750	-0.01	0.16**	0.15***
	(0.02)	(0.06)	(0.04)
Control mean	0.18	0.00	0.00
Control SD	0.39	1.00	1.00
p-value (all three equal)	0.599	0.277	0.674
Observations	1,829	2,784	2,784
Panel C. Pooled			
Pooled cash treatment:			
Cash	-0.03**	0.21***	0.11***
	(0.01)	(0.03)	(0.03)
Individual treatments by cash amount			
Cash 250	-0.01	0.16***	0.12***
	(0.02)	(0.04)	(0.04)
Cash 500	-0.03*	0.23***	0.09**
	(0.02)	(0.04)	(0.04)
Cash 750	-0.03*	0.26***	0.10***
	(0.02)	(0.04)	(0.04)
Control mean	0.27	0.00	0.00
Control SD	0.44	1.00	1.00
p-value (all three equal)	0.575	0.117	0.848
Observations	3,058	5,379	5,379

Note: The endline was conducted about 18-22 months after first transfers were received in Liberia and 21-25 months in Malawi. Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Monetary outcomes are in USD and Winsorized at the 99th percentile. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

[a] Includes only women and those for whom IPV was measured in face-to-face interviewing.

Table 2.A5: Disaggregated treatment effects on livestock holdings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Goats		Chickens		Pigs		Cattle		Total value of livestock (USD)
	Herd size	Value (USD)	Herd size	Value (USD)	Herd size	Value (USD)	Herd size	Value (USD)	
Panel A. Liberia									
Pooled cash treatment:									
Cash	0.07**	3.04*	0.48**	2.06**	0.00	-0.67			3.93
	(0.03)	(1.74)	(0.19)	(0.80)	(0.03)	(1.58)			(3.44)
Individual treatments by cash amount:									
Cash 250	0.05	2.45	0.57*	2.37**	-0.02	-1.15			3.65
	(0.05)	(2.71)	(0.29)	(1.13)	(0.04)	(2.10)			(4.43)
Cash 500	0.05	2.92	0.61**	2.43**	0.02	-0.24			3.70
	(0.05)	(2.65)	(0.26)	(1.16)	(0.04)	(1.99)			(4.79)
Cash 750	0.10**	3.74	0.26	1.37	-0.00	-0.62			4.44
	(0.05)	(2.41)	(0.28)	(1.18)	(0.05)	(2.25)			(5.05)
Control mean	0.20	9.73	3.15	10.92	0.16	8.06	0.00	0.00	38.11
Control SD	0.81	43.76	4.62	19.12	0.79	43.25	0.00	0.00	92.96
p-value (all three equal)	0.692	0.922	0.528	0.723	0.708	0.926			0.989
Observations	2,595	2,595	2,595	2,595	2,595	2,595	2,595	2,595	2,595
Panel B. Malawi									
Pooled cash treatment:									
Cash	0.22***	7.05***	0.48***	3.00***	0.00	0.17	-0.00	0.58	11.12**
	(0.05)	(1.82)	(0.17)	(0.97)	(0.02)	(0.99)	(0.01)	(2.04)	(4.68)
Individual treatments by cash amount:									
Cash 250	0.07	1.86	0.08	0.54	-0.03	-1.57	-0.01	-0.20	-2.17
	(0.07)	(2.22)	(0.19)	(1.10)	(0.02)	(1.24)	(0.01)	(2.39)	(5.52)
Cash 500	0.29***	9.36***	0.63***	3.96***	0.03	1.75	-0.00	0.19	18.52***
	(0.08)	(3.07)	(0.23)	(1.37)	(0.03)	(1.59)	(0.01)	(2.49)	(6.50)
Cash 750	0.30***	10.03***	0.73**	4.56***	-0.00	0.33	0.00	1.78	17.23**
	(0.08)	(2.87)	(0.30)	(1.73)	(0.02)	(1.50)	(0.01)	(3.25)	(7.64)
Control mean	0.57	17.68	2.45	10.36	0.09	4.50	0.03	6.64	48.59
Control SD	1.35	46.64	4.19	20.83	0.49	28.20	0.24	49.86	120.52
p-value (all three equal)	0.013	0.019	0.025	0.025	0.151	0.170	0.724	0.838	0.006
Observations	2,784	2,784	2,784	2,784	2,784	2,784	2,784	2,784	2,784
Panel C. Pooled									
Pooled cash treatment:									
Cash	0.14***	5.15***	0.48***	2.58***	0.01	0.01	-0.00	0.29	8.35***
	(0.03)	(1.29)	(0.13)	(0.63)	(0.02)	(0.96)	(0.00)	(1.05)	(2.99)
Individual treatments by cash amount									
Cash 250	0.06	2.31	0.32*	1.44*	-0.02	-1.11	-0.00	-0.11	1.37
	(0.04)	(1.71)	(0.17)	(0.80)	(0.02)	(1.24)	(0.01)	(1.25)	(3.58)
Cash 500	0.17***	6.31***	0.62***	3.22***	0.04	1.22	-0.00	0.09	12.41***
	(0.05)	(2.13)	(0.17)	(0.91)	(0.03)	(1.35)	(0.01)	(1.29)	(4.23)
Cash 750	0.20***	6.88***	0.51**	3.08***	0.00	-0.07	0.00	0.92	11.32**
	(0.05)	(1.96)	(0.21)	(1.07)	(0.03)	(1.37)	(0.01)	(1.67)	(4.75)
Control mean	0.39	13.82	2.79	10.63	0.12	6.22	0.02	3.42	43.51
Control SD	1.14	45.43	4.42	20.01	0.66	36.33	0.18	35.92	108.13
p-value (all three equal)	0.022	0.102	0.352	0.179	0.182	0.322	0.728	0.837	0.032
Observations	5,379	5,379	5,379	5,379	5,379	5,379	5,379	5,379	5,379

Note: Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. Monetary outcomes are in USD and Winsorized at the 99th percentile. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

Table 2.A6: Crop Choice

	(1) =1 if planted or harvested staple crop	(2) Number of crops planted/harvested in the following: Non-staple cereals[b]	(3) Legumes[c]	(4) Fruit/ Vegetables[d]	(5) Other crops[e]	(6) All crops pooled
Panel A. Liberia						
Pooled cash treatment:						
Cash	0.04*	0.05*	0.02*	0.06	0.02*	0.20**
	(0.02)	(0.03)	(0.01)	(0.05)	(0.01)	(0.08)
Individual treatments by cash amount:						
Cash 250	0.04	0.02	0.03**	-0.05	0.04*	0.08
	(0.03)	(0.04)	(0.01)	(0.08)	(0.02)	(0.12)
Cash 500	0.05*	0.02	0.01	0.21***	0.02	0.31***
	(0.03)	(0.03)	(0.01)	(0.08)	(0.02)	(0.11)
Cash 750	0.04	0.11***	0.01	0.04	0.02	0.21*
	(0.03)	(0.04)	(0.01)	(0.07)	(0.02)	(0.11)
Control mean	0.55	0.47	0.03	0.71	0.05	1.83
Control SD	0.50	0.60	0.18	1.15	0.22	1.69
p-value (all three equal)	0.966	0.124	0.197	0.031	0.614	0.250
Observations	2,595	2,595	2,595	2,595	2,595	2,595
Panel B. Malawi						
Pooled cash treatment:						
Cash	0.01**	0.02	0.07**	0.01	0.00	0.11***
	(0.01)	(0.02)	(0.03)	(0.01)	(0.00)	(0.04)
Individual treatments by cash amount:						
Cash 250	0.01	-0.04	0.10*	0.02	-0.00	0.09
	(0.01)	(0.03)	(0.05)	(0.01)	(0.00)	(0.06)
Cash 500	0.01*	0.02	0.02	-0.00	0.01	0.06
	(0.01)	(0.03)	(0.05)	(0.01)	(0.01)	(0.06)
Cash 750	0.02***	0.07**	0.08*	0.00	0.00	0.18***
	(0.01)	(0.03)	(0.04)	(0.01)	(0.00)	(0.06)
Control mean	0.97	0.37	0.86	0.03	0.00	2.23
Control SD	0.18	0.56	0.78	0.19	0.04	1.02
p-value (all three equal)	0.439	0.032	0.367	0.368	0.310	0.273
Observations	2,784	2,784	2,784	2,784	2,784	2,784
Panel C. Pooled						
Pooled cash treatment:						
Cash	0.03***	0.03*	0.04**	0.03	0.01**	0.15***
	(0.01)	(0.02)	(0.02)	(0.03)	(0.01)	(0.04)
Individual treatments by cash amount						
Cash 250	0.03	-0.01	0.06**	-0.02	0.02*	0.08
	(0.02)	(0.03)	(0.03)	(0.04)	(0.01)	(0.07)
Cash 500	0.03**	0.02	0.01	0.10**	0.01	0.18***
	(0.01)	(0.02)	(0.03)	(0.04)	(0.01)	(0.06)
Cash 750	0.03**	0.09***	0.05*	0.02	0.01	0.20***
	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.06)
Control mean	0.77	0.42	0.46	0.36	0.02	2.03
Control SD	0.42	0.58	0.71	0.88	0.16	1.40
p-value (all three equal)	0.926	0.010	0.254	0.061	0.741	0.350
Observations	5,379	5,379	5,379	5,379	5,379	5,379

Note: The endline was conducted about 18-22 months after first transfers were received in Liberia and 21-25 months in Malawi. Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

[a] Rice in Liberia; maize in Malawi.

[b] Cassava, corn/maize in Liberia; rice, cassava in Malawi.

Table 2.A7: Agricultural Output

	(1)	(2)	(3)	(4)
	Staple crop:[a]			
	Amount harvested (kg)	Value of harvest (USD)	Total value of non-staple crops harvested (USD)	Total value of all crops harvested (USD)
Panel A. Liberia				
Pooled cash treatment:				
Cash	118.69***	78.15***	5.46	83.86***
	(35.32)	(23.25)	(15.71)	(28.63)
Individual treatments by cash amount:				
Cash 250	156.87***	103.28***	4.34	107.49**
	(60.32)	(39.72)	(21.56)	(48.20)
Cash 500	77.87	51.27	14.78	65.50
	(50.14)	(33.01)	(22.30)	(39.93)
Cash 750	121.50**	80.00**	-2.83	78.63*
	(50.69)	(33.37)	(24.02)	(41.76)
Control mean	249.03	163.96	102.05	266.01
Control SD	667.74	439.64	358.03	590.66
p-value (all three equal)	0.544	0.544	0.829	0.766
Observations	2,595	2,595	2,595	2,595
Panel B. Malawi				
Pooled cash treatment:				
Cash	41.45***	11.85***	13.03**	24.69***
	(15.34)	(4.39)	(6.40)	(8.73)
Individual treatments by cash amount:				
Cash 250	38.05*	10.88*	18.70**	29.68**
	(20.11)	(5.75)	(9.03)	(12.03)
Cash 500	25.04	7.16	2.32	9.56
	(20.39)	(5.83)	(7.96)	(11.82)
Cash 750	62.05**	17.74**	18.28*	35.28**
	(24.04)	(6.87)	(10.79)	(14.04)
Control mean	219.54	62.76	48.47	111.23
Control SD	325.07	92.93	122.64	172.81
p-value (all three equal)	0.401	0.401	0.213	0.220
Observations	2,784	2,784	2,784	2,784
Panel C. Pooled				
Pooled cash treatment:				
Cash	78.90***	44.02***	9.26	53.97***
	(19.06)	(11.63)	(8.31)	(14.71)
Individual treatments by cash amount				
Cash 250	94.41***	55.13***	11.84	66.45***
	(31.15)	(19.52)	(11.32)	(24.17)
Cash 500	51.30*	28.93*	8.06	36.75*
	(26.77)	(16.28)	(11.53)	(20.41)
Cash 750	91.30***	48.12***	7.84	58.85***
	(27.44)	(16.63)	(12.97)	(21.45)
Control mean	233.86	111.89	74.48	186.37
Control SD	520.51	317.47	265.81	436.62
p-value (all three equal)	0.400	0.479	0.952	0.541
Observations	5,379	5,379	5,379	5,379

Note: Sample restricted to households engaged in harvesting or planting at least one crop in the past season. The endline was conducted about 18-22 months after first transfers were received in Liberia and 21-25 months in Malawi. Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.
[a] Staple crop is rice in Liberia and maize in Malawi.

Table 2.A8: Self-reported transfer size, withdrawal, and transaction costs

	(1) =1 if received any $50+ transfers	(2) Total received amount (USD)	(3) Total cashout amount (USD)	(4) Total transport cost (USD)	(5) Total withdrawal fee (USD)
Pooled cash treatment:					
Cash	0.66***	295.80***	273.61***	3.82***	10.25***
	(0.02)	(15.06)	(14.13)	(0.30)	(0.59)
Individual treatments by cash amount:					
Cash 250	0.63***	139.40***	127.26***	1.59***	5.03***
	(0.03)	(8.13)	(7.73)	(0.15)	(0.34)
Cash 500	0.65***	289.71***	269.55***	3.85***	9.31***
	(0.03)	(14.24)	(13.57)	(0.34)	(0.59)
Cash 750	0.69***	464.83***	430.06***	6.11***	16.66***
	(0.03)	(24.55)	(22.84)	(0.65)	(1.06)
Control mean	0.02	1.31	1.00	0.00	0.04
Control SD	0.13	12.51	11.52	0.09	1.17
p-value (all three equal)	0.348	0.000	0.000	0.000	0.000
Observations	2,784	2,784	2,784	2,784	2,784

Note: Regressions include strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. Winsorized at the 99th percentile. Monetary values in USD.

Table 2.A9: Self-reported usage of cash transfers

	(1)	(2)
	Aggregate cash amount specified for following reason	
	Mean	SD
Food	32.79	37.39
Nondurables	11.58	46.95
Clothes	9.77	20.27
Education	6.41	21.15
Home repair/construction	106.03	102.70
Contributions	0.38	4.97
Health preventatives	4.67	19.86
Durables	20.26	53.25
Farming inputs	24.05	58.31
Total	215.95	145.34
Observations	968	

Note: Data is available for Malawi only. Observations restricted to households who reported any large cash transfers in 2019-2021.

Table 2.A10: Transfers and Labor Demand

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Transfers		Hiring casual laborers for:					
			Own farm			Self-owned business		
	Sent (USD)	Received (USD)	=1 if any	hours	payments (USD)	=1 if any	hours	payments (USD)
Panel A. Liberia								
Pooled cash treatment:								
Cash	-0.45	-0.34	0.03**	0.88	0.25	0.03	0.08	0.05
	(0.38)	(0.47)	(0.01)	(0.76)	(0.33)	(0.02)	(0.21)	(0.15)
Individual treatments by cash amount:								
Cash 250	-0.93**	-0.57	0.00	0.80	0.27	0.04	0.46	0.32
	(0.45)	(0.55)	(0.02)	(1.47)	(0.65)	(0.04)	(0.53)	(0.29)
Cash 500	-0.23	-0.40	0.06***	0.92	0.25	0.02	-0.12	-0.08
	(0.47)	(0.58)	(0.02)	(0.78)	(0.35)	(0.03)	(0.14)	(0.14)
Cash 750	-0.20	-0.05	0.02	0.94	0.24	0.03	-0.11	-0.09
	(0.47)	(0.67)	(0.02)	(0.98)	(0.35)	(0.03)	(0.15)	(0.14)
Control mean	2.00	2.63	0.08	2.87	1.15	0.03	0.18	0.16
Control SD	11.29	14.26	0.26	16.63	7.63	0.18	4.27	4.73
p-value (all three equal)	0.174	0.739	0.034	0.996	0.998	0.925	0.498	0.311
Observations	2,595	2,595	2,595	2,595	2,595	383	2,595	2,595
Panel B. Malawi								
Pooled cash treatment:								
Cash	0.01	0.17	0.01	-0.13	0.24	-0.00	0.53	0.10
	(0.06)	(0.19)	(0.01)	(0.36)	(0.19)	(0.01)	(0.42)	(0.08)
Individual treatments by cash amount:								
Cash 250	0.03	0.06	-0.00	-0.64**	-0.01	0.00	0.37	0.08
	(0.08)	(0.26)	(0.01)	(0.32)	(0.21)	(0.01)	(0.38)	(0.08)
Cash 500	-0.08	0.10	0.00	0.14	0.31	-0.01	-0.02	-0.01
	(0.07)	(0.27)	(0.01)	(0.60)	(0.33)	(0.01)	(0.10)	(0.02)
Cash 750	0.08	0.34	0.02	0.12	0.42	0.00	1.29	0.25
	(0.09)	(0.28)	(0.01)	(0.46)	(0.30)	(0.02)	(1.22)	(0.22)
Control mean	0.42	1.01	0.04	1.03	0.46	0.01	0.01	0.00
Control SD	1.84	4.31	0.19	10.00	3.46	0.12	0.20	0.14
p-value (all three equal)	0.124	0.682	0.396	0.066	0.364	0.296	0.352	0.306
Observations	2,784	2,784	2,784	2,784	2,784	411	2,784	2,784
Panel C. Pooled								
Pooled cash treatment:								
Cash	-0.23	-0.09	0.02**	0.36	0.24	0.01	0.31	0.08
	(0.19)	(0.25)	(0.01)	(0.42)	(0.19)	(0.01)	(0.24)	(0.08)
Individual treatments by cash amount								
Cash 250	-0.44*	-0.25	0.00	0.06	0.12	0.02	0.41	0.19
	(0.22)	(0.29)	(0.01)	(0.73)	(0.33)	(0.02)	(0.32)	(0.14)
Cash 500	-0.17	-0.15	0.03**	0.51	0.27	0.00	-0.07	-0.04
	(0.23)	(0.31)	(0.01)	(0.49)	(0.24)	(0.02)	(0.08)	(0.07)
Cash 750	-0.06	0.15	0.02*	0.53	0.33	0.02	0.61	0.08
	(0.23)	(0.36)	(0.01)	(0.54)	(0.23)	(0.02)	(0.64)	(0.13)
Control mean	1.19	1.80	0.06	1.93	0.79	0.02	0.09	0.08
Control SD	8.01	10.44	0.23	13.66	5.88	0.15	2.98	3.30
p-value (all three equal)	0.210	0.556	0.073	0.816	0.844	0.781	0.186	0.139
Observations	5,379	5,379	5,379	5,379	5,379	794	5,379	5,379

Note: Regressions include strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. Winsorized at the 99th percentile. Monetary values in USD.

Appendix B. Effects on Secondary Outcomes (End-line)

Table 2.B1: Labor Supply

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	For female and/or male household heads (past month):							
	Own farm		Casual labor		Own business		Other job	
	=1 if any	Number of hours	=1 if any	Number of hours	=1 if any	Number of hours	=1 if any	Number of hours
Panel A. Liberia								
Pooled cash treatment:								
Cash	-0.02	-2.29	-0.10***	-6.38***	0.02	-0.47	0.02	1.05*
	(0.02)	(3.05)	(0.02)	(1.56)	(0.02)	(1.56)	(0.01)	(0.62)
Individual treatments by cash amount:								
Cash 250	-0.02	-5.02	-0.10***	-5.08**	0.01	-0.43	0.01	0.76
	(0.03)	(3.84)	(0.03)	(2.22)	(0.02)	(2.06)	(0.01)	(0.88)
Cash 500	0.01	-3.14	-0.10***	-7.12***	0.02	-1.00	0.02	1.68*
	(0.03)	(3.51)	(0.02)	(1.82)	(0.02)	(2.05)	(0.01)	(1.00)
Cash 750	-0.05*	1.30	-0.10***	-6.93***	0.03	0.01	0.01	0.71
	(0.03)	(4.67)	(0.02)	(1.79)	(0.02)	(2.03)	(0.02)	(0.93)
Control mean	0.69	43.50	0.32	16.98	0.14	9.09	0.06	1.98
Control SD	0.46	70.13	0.47	39.93	0.35	38.66	0.24	14.84
p-value (all three equal)	0.208	0.446	0.993	0.624	0.786	0.905	0.795	0.688
Observations	2,595	2,595	2,595	2,595	2,595	2,595	2,595	2,595
Panel B. Malawi								
Pooled cash treatment:								
Cash	0.02	0.80	-0.04*	-1.38	-0.02	-0.34	0.01	-0.17
	(0.02)	(0.75)	(0.02)	(1.90)	(0.02)	(1.16)	(0.01)	(1.01)
Individual treatments by cash amount:								
Cash 250	0.00	-0.73	-0.01	0.27	-0.01	1.05	0.03*	0.96
	(0.03)	(0.87)	(0.03)	(2.57)	(0.02)	(1.79)	(0.01)	(1.36)
Cash 500	0.03	1.72	-0.07***	-2.10	-0.03	-0.82	0.01	-0.36
	(0.03)	(1.18)	(0.02)	(2.52)	(0.02)	(1.47)	(0.01)	(1.49)
Cash 750	0.04	1.42	-0.03	-2.34	-0.01	-1.28	-0.01	-1.14
	(0.03)	(1.19)	(0.03)	(2.63)	(0.02)	(1.48)	(0.01)	(1.21)
Control mean	0.59	9.85	0.43	21.90	0.22	6.21	0.06	3.69
Control SD	0.49	19.48	0.49	47.52	0.42	29.29	0.25	25.84
p-value (all three equal)	0.551	0.089	0.207	0.636	0.657	0.479	0.120	0.359
Observations	2,784	2,784	2,784	2,784	2,784	2,784	2,784	2,784
Panel C. Pooled								
Pooled cash treatment:								
Cash	0.00	-0.71	-0.07***	-3.72***	-0.00	-0.40	0.01	0.42
	(0.01)	(1.53)	(0.01)	(1.24)	(0.01)	(0.96)	(0.01)	(0.60)
Individual treatments by cash amount								
Cash 250	-0.01	-2.80	-0.06***	-2.22	-0.00	0.35	0.02**	0.88
	(0.02)	(1.90)	(0.02)	(1.73)	(0.01)	(1.36)	(0.01)	(0.82)
Cash 500	0.02	-0.64	-0.08***	-4.49***	-0.01	-0.91	0.01	0.63
	(0.02)	(1.81)	(0.02)	(1.58)	(0.02)	(1.24)	(0.01)	(0.91)
Cash 750	-0.01	1.34	-0.07***	-4.47***	0.01	-0.65	-0.00	-0.25
	(0.02)	(2.34)	(0.02)	(1.61)	(0.02)	(1.25)	(0.01)	(0.77)
Control mean	0.64	26.18	0.37	19.51	0.18	7.61	0.06	2.86
Control SD	0.48	53.52	0.48	44.06	0.39	34.19	0.25	21.24
p-value (all three equal)	0.423	0.228	0.396	0.407	0.771	0.686	0.260	0.462
Observations	5,379	5,379	5,379	5,379	5,379	5,379	5,379	5,379

Note: Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. Nonbinary outcomes are Winsorized at the 99th percentile.

Table 2.B2: Education

	(1)	(2)	(3)	(4)	(5)
			Missed school days (past year)		
	Proportion of children enrolled	Education expenditure (past 6 months)	for any reason	due to lack of money	Proportion of school days attended (past week)
Panel A. Liberia					
Pooled cash treatment:					
Cash	0.10***	8.92***	-3.59*	-1.43**	0.07**
	(0.02)	(2.36)	(1.94)	(0.60)	(0.03)
Individual treatments by cash amount:					
Cash 250	0.05	7.66*	-6.22***	-2.17***	0.07*
	(0.03)	(3.89)	(2.22)	(0.67)	(0.04)
Cash 500	0.11***	11.57***	-1.94	-1.18	0.07**
	(0.03)	(3.49)	(2.48)	(0.75)	(0.04)
Cash 750	0.13***	7.38**	-2.83	-0.99	0.06*
	(0.03)	(3.22)	(2.89)	(0.93)	(0.03)
Control mean	0.52	24.71	12.24	3.43	0.89
Control SD	0.45	48.16	43.64	14.12	0.28
p-value (all three equal)	0.104	0.578	0.192	0.272	0.933
Observations	1,871	1,871	1,876	1,876	245
Panel B. Malawi					
Pooled cash treatment:					
Cash	-0.01	-0.41	-0.76	-0.30*	0.01
	(0.01)	(0.93)	(0.57)	(0.16)	(0.01)
Individual treatments by cash amount:					
Cash 250	-0.02	-0.36	-0.38	-0.06	0.03**
	(0.01)	(1.54)	(0.90)	(0.23)	(0.01)
Cash 500	-0.01	-1.35	-1.15	-0.43**	-0.01
	(0.01)	(1.03)	(0.74)	(0.17)	(0.02)
Cash 750	-0.01	0.48	-0.75	-0.40**	0.02
	(0.01)	(1.21)	(0.75)	(0.17)	(0.01)
Control mean	0.93	10.78	7.28	0.99	0.91
Control SD	0.20	22.66	13.11	4.14	0.22
p-value (all three equal)	0.819	0.342	0.740	0.234	0.069
Observations	2,158	2,158	2,158	2,158	1,757
Panel C. Pooled					
Pooled cash treatment:					
Cash	0.04***	3.93***	-2.09**	-0.82***	0.02*
	(0.01)	(1.23)	(0.96)	(0.29)	(0.01)
Individual treatments by cash amount					
Cash 250	0.01	3.25	-3.04***	-1.02***	0.03***
	(0.02)	(1.98)	(1.15)	(0.34)	(0.01)
Cash 500	0.05***	4.78***	-1.54	-0.79**	-0.00
	(0.02)	(1.77)	(1.22)	(0.36)	(0.01)
Cash 750	0.06***	3.74**	-1.73	-0.67	0.02*
	(0.02)	(1.64)	(1.41)	(0.45)	(0.01)
Control mean	0.74	17.33	9.61	2.14	0.90
Control SD	0.40	37.55	31.50	10.21	0.23
p-value (all three equal)	0.094	0.801	0.454	0.689	0.110
Observations	4,029	4,029	4,034	4,034	2,002

Note: Sample restricted to households with any school-aged children (age 6-18). Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. Nonbinary outcomes are Winsorized at the 99th percentile.

Table 2.B3: Health Investment

	(1)	(2)	(3)	(4)
	Proportion of household members:			Average proportion of under-5 children with recommended vaccinations[b]
	sought preventative care (past 3 months)	slept under bednet (yesterday)	with any vaccinations (under 18)[a]	
Panel A. Liberia				
Pooled cash treatment:				
Cash	0.00	0.05***	0.01	-0.01
	(0.00)	(0.02)	(0.01)	(0.02)
Individual treatments by cash amount:				
Cash 250	0.01*	0.04	0.01	-0.03
	(0.01)	(0.02)	(0.01)	(0.03)
Cash 500	0.00	0.05**	0.02	-0.00
	(0.00)	(0.02)	(0.02)	(0.03)
Cash 750	-0.00	0.06***	-0.01	0.02
	(0.00)	(0.02)	(0.01)	(0.03)
Control mean	0.03	0.72	0.12	0.79
Control SD	0.08	0.42	0.24	0.28
p-value (all three equal)	0.203	0.575	0.255	0.336
Observations	2,595	2,595	2,228	643
Panel B. Malawi				
Pooled cash treatment:				
Cash	-0.01	0.00	-0.00	-0.01
	(0.00)	(0.02)	(0.01)	(0.01)
Individual treatments by cash amount:				
Cash 250	0.00	-0.03	-0.00	-0.02
	(0.01)	(0.02)	(0.01)	(0.02)
Cash 500	-0.01***	0.02	-0.02	0.01
	(0.01)	(0.02)	(0.01)	(0.01)
Cash 750	-0.00	0.01	0.01	-0.01
	(0.01)	(0.03)	(0.01)	(0.02)
Control mean	0.05	0.71	0.18	0.91
Control SD	0.12	0.40	0.28	0.18
p-value (all three equal)	0.050	0.179	0.147	0.140
Observations	2,784	2,784	2,516	966
Panel C. Pooled				
Pooled cash treatment:				
Cash	-0.00	0.02**	0.00	-0.01
	(0.00)	(0.01)	(0.01)	(0.01)
Individual treatments by cash amount				
Cash 250	0.01	0.00	0.00	-0.03
	(0.00)	(0.02)	(0.01)	(0.02)
Cash 500	-0.01	0.04**	-0.00	0.01
	(0.00)	(0.02)	(0.01)	(0.01)
Cash 750	-0.00	0.03**	0.00	0.00
	(0.00)	(0.02)	(0.01)	(0.02)
Control mean	0.04	0.71	0.15	0.86
Control SD	0.11	0.41	0.27	0.23
p-value (all three equal)	0.059	0.131	0.925	0.187
Observations	5,379	5,379	4,744	1,609

Note: Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses.
[a] Sample restricted to households with any member under 18.
[b] Sample restricted to households with any child under 5.

Table 2.B4: Health Resilience

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		If any HH member sick in the past month:					
	Number of illnesses per member (past month)	Proportion of sick members			Number of missed		Expenses on treatment
		treated at all	delayed treatment	not fully treat	work days	school days	
Panel A. Liberia							
Pooled cash treatment:							
Cash	-0.01		-0.02	-0.01	4.02*	0.12	7.52**
	(0.01)		(0.02)	(0.02)	(2.43)	(0.14)	(2.98)
Individual treatments by cash amount:							
Cash 250	-0.01		0.00	-0.01	5.25*	0.13	10.00*
	(0.01)		(0.03)	(0.03)	(3.18)	(0.24)	(5.41)
Cash 500	0.01		-0.02	-0.02	3.88	0.36	6.37
	(0.01)		(0.03)	(0.02)	(4.09)	(0.24)	(4.56)
Cash 750	-0.02*		-0.05*	-0.01	2.78	-0.19	6.16
	(0.01)		(0.03)	(0.03)	(3.11)	(0.13)	(4.08)
Control mean	0.10		0.10	0.91	11.85	0.29	14.73
Control SD	0.20		0.29	0.27	25.38	1.41	32.45
p-value (all three equal)	0.175		0.409	0.990	0.813	0.098	0.820
Observations	2,595		704	704	704	704	704
Panel B. Malawi							
Pooled cash treatment:							
Cash	-0.01	-0.00	-0.00	0.00	-5.64	0.27	0.09
	(0.01)	(0.02)	(0.00)	(0.01)	(5.75)	(0.62)	(0.20)
Individual treatments by cash amount:							
Cash 250	-0.00	0.01	-0.00	-0.00	-11.55*	-0.15	-0.14
	(0.02)	(0.02)	(0.01)	(0.01)	(6.55)	(0.79)	(0.29)
Cash 500	-0.01	-0.02	0.00	0.02*	-1.29	0.49	0.28
	(0.02)	(0.02)	(0.01)	(0.01)	(8.26)	(0.82)	(0.30)
Cash 750	-0.00	0.00	-0.01	-0.01	-3.87	0.48	0.13
	(0.01)	(0.02)	(0.01)	(0.01)	(7.92)	(0.98)	(0.26)
Control mean	0.22	0.87	0.01	0.96	28.55	6.61	1.81
Control SD	0.27	0.30	0.11	0.16	111.63	11.71	3.30
p-value (all three equal)	0.876	0.673	0.403	0.104	0.423	0.750	0.512
Observations	2,784	1,495	1,495	1,495	1,495	1,495	1,495
Panel C. Pooled							
Pooled cash treatment:							
Cash	-0.01	-0.00	-0.01	-0.00	-2.30	0.22	2.55**
	(0.01)	(0.02)	(0.01)	(0.01)	(4.00)	(0.43)	(1.00)
Individual treatments by cash amount							
Cash 250	-0.01	0.01	-0.00	-0.01	-6.50	-0.07	2.94*
	(0.01)	(0.02)	(0.01)	(0.01)	(4.71)	(0.55)	(1.72)
Cash 500	-0.00	-0.02	-0.00	0.01	0.89	0.46	2.46
	(0.01)	(0.02)	(0.01)	(0.01)	(5.74)	(0.55)	(1.57)
Cash 750	-0.01	0.00	-0.02*	-0.01	-1.31	0.28	2.23*
	(0.01)	(0.02)	(0.01)	(0.01)	(5.64)	(0.70)	(1.19)
Control mean	0.16	0.87	0.04	0.94	22.87	4.46	6.20
Control SD	0.25	0.30	0.19	0.21	92.20	10.00	20.06
p-value (all three equal)	0.684	0.673	0.344	0.394	0.437	0.710	0.934
Observations	5,379	1,495	2,199	2,199	2,199	2,199	2,199

Note: Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses.

Table 2.B5: Child Anthropometrics (Malawi only)

	(1) Height for age	(2) Weight for age	(3) MUAC for age
Pooled cash treatment:			
Cash	0.09	-0.00	0.02
	(0.09)	(0.06)	(0.06)
Individual treatments by cash amount:			
Cash 250	0.02	0.03	0.09
	(0.11)	(0.09)	(0.08)
Cash 500	0.09	-0.08	-0.09
	(0.15)	(0.07)	(0.08)
Cash 750	0.17	0.03	0.05
	(0.14)	(0.09)	(0.08)
Control mean	-1.63	-0.52	-0.30
Control SD	1.54	1.04	0.99
p-value (all three equal)	0.660	0.384	0.110
Observations	1,488	1,488	1,479

Note: Sample to restricted to children under 5. All measures are standardized z-scores using means and standard deviations from *WHO Child Growth Standards*. Regressions include strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

Table 2.B6: Social Capital

	(1)	(2)	(3)	(4)
	During difficult times[a]			
	=1 if your household could depend on:		=1 if the following could depend on your household:	
	relatives	non-relatives	relatives	non-relatives
Panel A. Liberia				
Pooled cash treatment:				
Cash	0.03	0.03	0.03	0.04
	(0.02)	(0.02)	(0.02)	(0.02)
Individual treatments by cash amount:				
Cash 250	0.01	0.03	0.01	0.04
	(0.03)	(0.03)	(0.03)	(0.03)
Cash 500	0.02	-0.00	0.03	0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Cash 750	0.07**	0.05*	0.06*	0.06*
	(0.03)	(0.03)	(0.03)	(0.03)
Control mean	0.60	0.36	0.64	0.45
Control SD	0.49	0.48	0.48	0.50
p-value (all three equal)	0.185	0.362	0.498	0.432
Observations	2,594	2,590	2,592	2,588
Panel B. Malawi				
Pooled cash treatment:				
Cash	0.00	0.02	0.04*	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Individual treatments by cash amount:				
Cash 250	0.03	0.03	0.06**	0.02
	(0.03)	(0.03)	(0.03)	(0.02)
Cash 500	-0.03	0.01	0.02	0.02
	(0.03)	(0.03)	(0.03)	(0.03)
Cash 750	0.01	0.03	0.03	-0.02
	(0.03)	(0.02)	(0.03)	(0.03)
Control mean	0.52	0.28	0.58	0.36
Control SD	0.50	0.45	0.49	0.48
p-value (all three equal)	0.196	0.756	0.397	0.377
Observations	2,783	2,777	2,781	2,764
Panel C. Pooled				
Pooled cash treatment:				
Cash	0.02	0.02*	0.04**	0.02
	(0.01)	(0.01)	(0.01)	(0.02)
Individual treatments by cash amount				
Cash 250	0.02	0.03	0.04*	0.03
	(0.02)	(0.02)	(0.02)	(0.02)
Cash 500	-0.01	0.00	0.03	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Cash 750	0.04*	0.04**	0.04*	0.02
	(0.02)	(0.02)	(0.02)	(0.02)
Control mean	0.56	0.32	0.61	0.40
Control SD	0.50	0.47	0.49	0.49
p-value (all three equal)	0.208	0.293	0.837	0.799
Observations	5,377	5,367	5,373	5,352

Note: Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses.

[a] Assistance includes financial or food support. Examples of difficult times include: loss of a family member, loss of income, hunger, drought, flood, conflict or similar events.

Table 2.B7: Public Goods Contributions

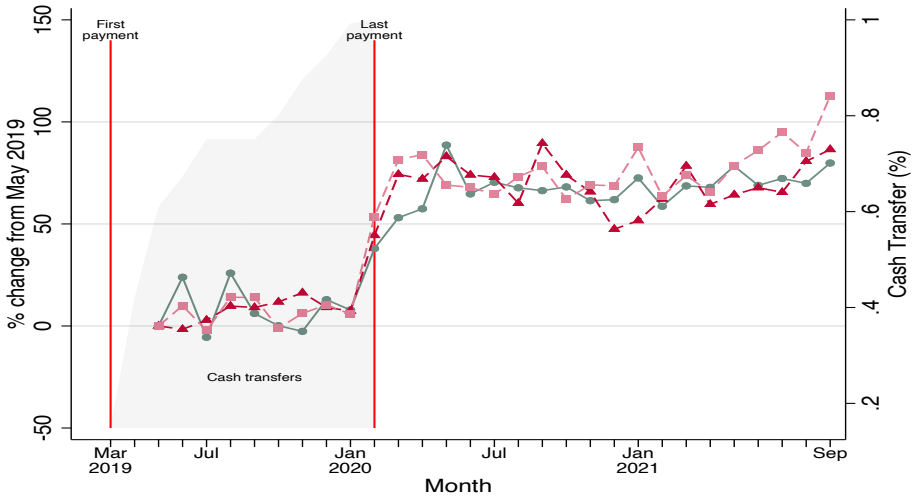
	(1)	(2)	(3)
	For community service activities (past 12 months)[a]:		
	Number of labor hours contributed	Cash contributions	Value of in-kind contributions
Panel A. Liberia			
Pooled cash treatment:			
Cash	1.25	0.02	0.03
	(1.09)	(0.02)	(0.03)
Individual treatments by cash amount:			
Cash 250	1.83	-0.01	0.00
	(1.69)	(0.02)	(0.04)
Cash 500	2.75	0.05*	0.06
	(1.78)	(0.03)	(0.04)
Cash 750	-0.84	0.03	0.03
	(1.17)	(0.03)	(0.04)
Control mean	6.59	0.10	0.14
Control SD	20.56	0.44	0.70
p-value (all three equal)	0.095	0.198	0.467
Observations	2,595	2,595	2,595
Panel B. Malawi			
Pooled cash treatment:			
Cash	0.19	0.00	-0.00
	(0.52)	(0.00)	(0.00)
Individual treatments by cash amount:			
Cash 250	0.30	-0.00	-0.00
	(0.68)	(0.00)	(0.00)
Cash 500	-0.18	0.00	-0.00
	(0.74)	(0.00)	(0.00)
Cash 750	0.45	0.00	-0.00
	(0.76)	(0.00)	(0.00)
Control mean	2.89	0.01	0.00
Control SD	12.67	0.07	0.03
p-value (all three equal)	0.770	0.156	0.756
Observations	2,784	2,784	2,784
Panel C. Pooled			
Pooled cash treatment:			
Cash	0.70	0.01	0.02
	(0.60)	(0.01)	(0.01)
Individual treatments by cash amount			
Cash 250	1.04	-0.00	0.00
	(0.89)	(0.01)	(0.02)
Cash 500	1.21	0.03*	0.03
	(0.95)	(0.01)	(0.02)
Cash 750	-0.16	0.01	0.02
	(0.70)	(0.02)	(0.02)
Control mean	4.68	0.05	0.07
Control SD	17.06	0.31	0.49
p-value (all three equal)	0.291	0.115	0.462
Observations	5,379	5,379	5,379

Note: Regressions include strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. Winsorized at the 99th percentile. Monetary values in USD.
[a] Cleaning/maintaining or repairing/building of: road/neighbourhood/bridge; schools; clean water/bathing, washing, sanitary facilities; irrigation canal/weir; house of worship/cemetery; village/neighbourhood facilities (meeting hall, office, gate, sports field); poor people dwellings; health facility.

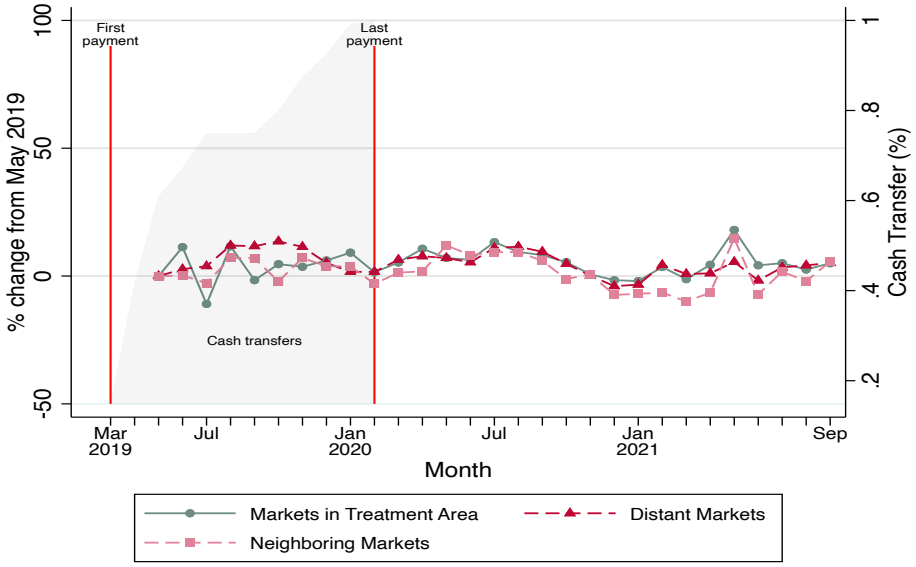
Appendix C. Prices

Figure 2.C1: Liberia Wave 1: Average Change in Prices of Select Items Relative to Pre-Treatment Level

a. Select Items



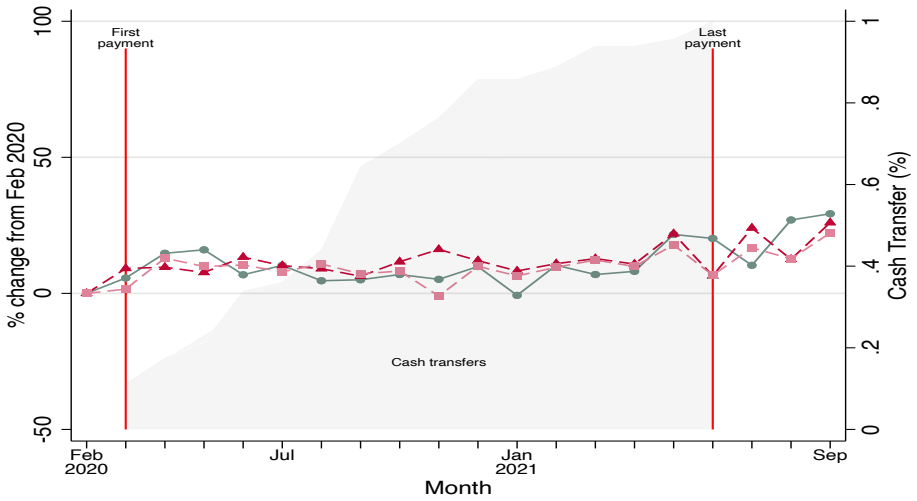
b. Staples



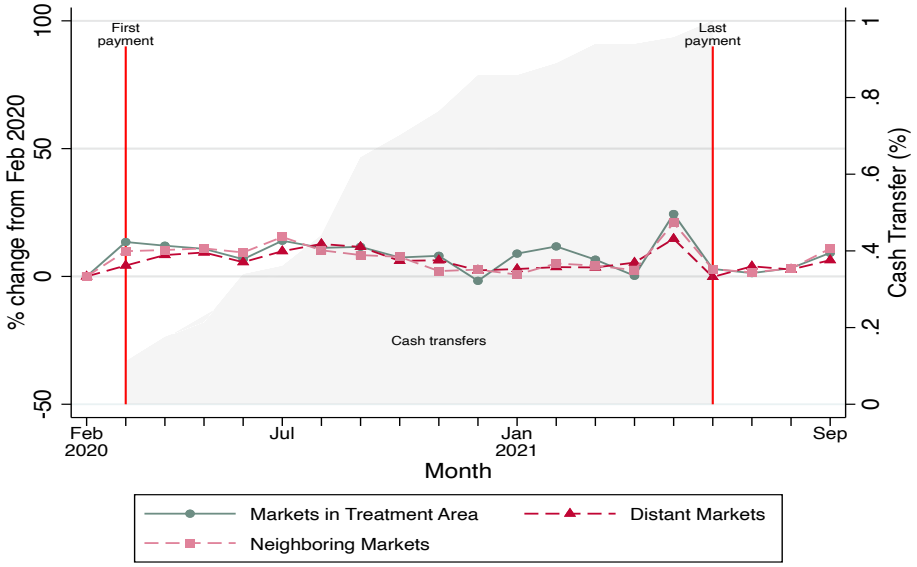
Note: The figure shows average change in prices relative to prices in May 2019 for Liberia Wave 1. First transfer to treatment households was made across March-May 2019 for Liberia Wave 1.There are 30 markets surveyed in Liberia Wave 1: 11 markets in treatment area, 7 in areas close to the treatment area, and 12 in distant areas. The sub-figure (a) shows the expenditure share weighted price of select items. The list of selected items for Liberia includes: cassava, cassava flour, dried fish, fresh fish, chicken, imported rice, okra, onion, palm oil, and salt. The list of items in staples include: cassava, cassava flour, and imported rice.

Figure 2.C2: Liberia Wave 2: Average Change in Prices of Select Items Relative to Pre-Treatment Level

a. Select Items



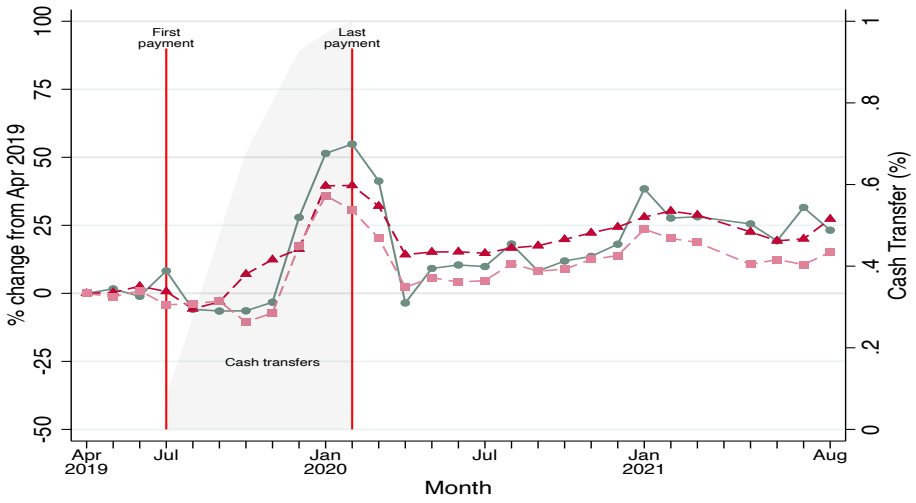
b. Staples



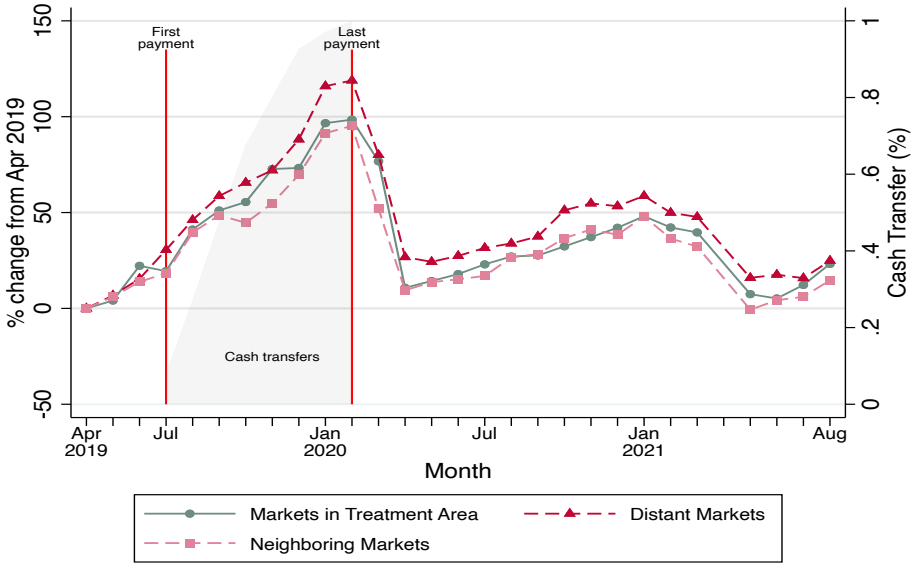
Note: The figure shows average change in prices relative to pre-treatment prices in February 2020 for Liberia Wave 2. First transfer to treatment households was made across March-September 2020 for Liberia Wave 2. There are 50 markets surveyed in Wave 2: 11 markets in treatment area, 22 in areas close to the treated area, and 17 in distant areas. The list of selected items for Liberia includes: cassava, cassava flour, dried fish, fresh fish, chicken, imported rice, okra, onion, palm oil, and salt. The list of items in staples include: cassava, cassava flour, and imported rice.

Figure 2.C3: Malawi: Average Change in Prices of Select Items Relative to Pre-Treatment Level

a. Select Items



b. Staples



Note: The figure shows average change in prices relative to prices in April 2019. First transfer to treatment households was made across July-October 2019. There are 95 markets surveyed in Malawi: 10 markets in treatment area, 42 in areas close to the treated area, and 43 in distant areas. The list of selected items for Malawi includes: beans, chicken, dried fish, eggs, groundnut, maize flour, maize kernel, onion, salt, sugar, sweet potato, tomato, and unpacked rice. The list of items in staples include: beans, maize flour, maize kernel, and sweet potato.

Appendix D: Disaggregated Primary outcomes

Table 2.D1: Food Security Index

	(1) HDDS[a] (yesterday)	(2) FCS[b] (past week)	(3) HHS[c] (past month)	(4) FIES[d] (past year)
Panel A. Liberia				
Pooled cash treatment:				
Cash	0.31***	-0.41	-0.28***	-0.73***
	(0.08)	(2.94)	(0.06)	(0.10)
Individual treatments by cash amount:				
Cash 250	0.21*	1.82	-0.12	-0.54***
	(0.11)	(1.26)	(0.08)	(0.13)
Cash 500	0.18*	-5.86	-0.30***	-0.71***
	(0.11)	(7.79)	(0.07)	(0.14)
Cash 750	0.55***	2.84**	-0.43***	-0.95***
	(0.09)	(1.39)	(0.08)	(0.16)
Control mean	5.36	47.32	1.34	6.50
Control SD	1.97	17.20	1.29	2.03
p-value (all three equal)	0.001	0.341	0.004	0.094
Observations	2,595	2,595	2,595	2,595
Panel B. Malawi				
Pooled cash treatment:				
Cash	0.03	0.65	-0.16***	-0.37***
	(0.07)	(0.59)	(0.04)	(0.11)
Individual treatments by cash amount:				
Cash 250	-0.02	0.24	-0.13**	-0.18
	(0.09)	(0.70)	(0.06)	(0.14)
Cash 500	0.06	0.36	-0.18***	-0.42***
	(0.10)	(0.78)	(0.06)	(0.15)
Cash 750	0.07	1.37	-0.17***	-0.51***
	(0.11)	(1.00)	(0.07)	(0.17)
Control mean	5.44	45.60	0.95	6.07
Control SD	1.80	14.62	1.28	2.75
p-value (all three equal)	0.674	0.555	0.737	0.189
Observations	2,784	2,784	2,784	2,784
Panel C. Pooled				
Pooled cash treatment:				
Cash	0.17***	0.14	-0.22***	-0.55***
	(0.06)	(1.45)	(0.04)	(0.08)
Individual treatments by cash amount				
Cash 250	0.10	0.98	-0.12**	-0.36***
	(0.07)	(0.71)	(0.05)	(0.10)
Cash 500	0.12*	-2.66	-0.24***	-0.59***
	(0.07)	(3.84)	(0.05)	(0.11)
Cash 750	0.30***	2.16***	-0.29***	-0.72***
	(0.07)	(0.82)	(0.05)	(0.12)
Control mean	5.40	46.43	1.14	6.28
Control SD	1.89	15.94	1.30	2.44
p-value (all three equal)	0.046	0.217	0.017	0.026
Observations	5,379	5,379	5,379	5,379

Note: In Columns 1, 2 and 5, higher values indicate improved food security; in Columns 3 and 4, lower values do. Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.
[a] Household Dietary Diversity Score (HDDS) ranges from 0 to 12 (FAO 2013).
[b] Food Consumption Score (FCS) is a weighted sum of the number of days (WFP 2008).
[c] Household Hunger Scale (HHS) ranges from 0 (less severe) to 6 (more severe) (Ballard et al. 2011).
[d] Food Insecurity Experience Scale (FIES) ranges from 0 (less insecure) to 8 (more insecure) (Cafiero et al. 2018).

Table 2.D2: Expenditure Categories

	(1) Food	(2) Nondurables	(3) Clothes	(4) Education	(5) Health	(6) Alcohol/ Tobacco	(7) Home repair	(8) Religious contribute	(9) Family events	(10) Nonmedical emergency
Panel A. Liberia										
Pooled cash treatment:										
Cash	0.68	0.75	1.72***	1.25***	1.41	-0.00	0.57**	0.34**	0.12	-0.05
	(0.89)	(0.60)	(0.45)	(0.31)	(0.96)	(0.05)	(0.26)	(0.15)	(0.33)	(0.12)
Individual treatments by cash amount:										
Cash 250	-1.13	-0.90	0.76	1.00**	1.78	0.01	0.23	0.32	-0.28	-0.01
	(1.17)	(0.74)	(0.58)	(0.49)	(1.79)	(0.07)	(0.33)	(0.20)	(0.46)	(0.20)
Cash 500	1.87	1.06	1.78***	1.69***	1.84	-0.10*	0.39	0.37	0.39	-0.14
	(1.37)	(0.81)	(0.58)	(0.48)	(1.51)	(0.05)	(0.35)	(0.24)	(0.48)	(0.14)
Cash 750	1.31	2.07**	2.64***	1.06**	0.60	0.08	1.09**	0.33	0.24	-0.01
	(1.11)	(0.94)	(0.68)	(0.41)	(1.11)	(0.07)	(0.42)	(0.22)	(0.45)	(0.16)
Control mean	26.91	11.40	5.58	3.14	4.94	0.29	1.12	1.27	3.44	0.45
Control SD	21.46	14.37	11.32	7.16	19.24	1.14	5.92	3.31	8.41	3.10
p-value (all three equal)	0.092	0.010	0.041	0.461	0.713	0.046	0.184	0.986	0.468	0.741
Observations	2,595	2,595	2,595	2,595	2,595	2,595	2,595	2,595	2,595	2,595
Panel B. Malawi										
Pooled cash treatment:										
Cash	0.45	0.14	0.16	-0.00	0.05	-0.00	0.17	0.01	-0.02	-0.01
	(0.47)	(0.27)	(0.22)	(0.13)	(0.11)	(0.01)	(0.16)	(0.06)	(0.08)	(0.03)
Individual treatments by cash amount:										
Cash 250	0.40	-0.03	-0.03	-0.01	-0.04	-0.01	0.10	-0.04	0.03	-0.04
	(0.60)	(0.37)	(0.31)	(0.20)	(0.17)	(0.01)	(0.20)	(0.09)	(0.11)	(0.04)
Cash 500	-0.09	-0.06	0.27	-0.12	0.12	0.00	0.24	0.05	-0.04	-0.01
	(0.65)	(0.35)	(0.32)	(0.14)	(0.17)	(0.01)	(0.22)	(0.09)	(0.12)	(0.05)
Cash 750	1.04	0.51	0.24	0.13	0.06	0.00	0.16	0.02	-0.07	0.02
	(0.75)	(0.43)	(0.34)	(0.17)	(0.15)	(0.01)	(0.23)	(0.09)	(0.10)	(0.05)
Control mean	9.56	4.96	2.18	1.47	1.09	0.05	0.94	1.01	0.45	0.13
Control SD	10.81	7.31	5.53	3.52	2.64	0.28	3.56	1.63	1.80	0.93
p-value (all three equal)	0.429	0.446	0.719	0.375	0.741	0.776	0.868	0.697	0.701	0.520
Observations	2,784	2,784	2,784	2,784	2,784	2,784	2,784	2,784	2,784	2,784
Panel C. Pooled										
Pooled cash treatment:										
Cash	0.56	0.43	0.92***	0.60***	0.71	-0.00	0.36**	0.17**	0.05	-0.03
	(0.49)	(0.33)	(0.25)	(0.17)	(0.47)	(0.02)	(0.15)	(0.08)	(0.16)	(0.06)
Individual treatments by cash amount										
Cash 250	-0.32	-0.44	0.36	0.46*	0.83	-0.00	0.16	0.13	-0.12	-0.03
	(0.65)	(0.41)	(0.32)	(0.26)	(0.87)	(0.04)	(0.19)	(0.11)	(0.23)	(0.10)
Cash 500	0.85	0.47	1.00***	0.76***	0.96	-0.04*	0.31	0.20	0.17	-0.07
	(0.75)	(0.43)	(0.33)	(0.25)	(0.73)	(0.03)	(0.20)	(0.13)	(0.24)	(0.07)
Cash 750	1.17*	1.28**	1.41***	0.58***	0.32	0.04	0.61**	0.17	0.08	0.01
	(0.66)	(0.51)	(0.38)	(0.22)	(0.54)	(0.04)	(0.24)	(0.12)	(0.23)	(0.08)
Control mean	17.98	8.09	3.83	2.28	2.96	0.16	1.03	1.13	1.90	0.29
Control SD	18.94	11.75	8.99	5.65	13.67	0.83	4.85	2.59	6.19	2.26
p-value (all three equal)	0.134	0.008	0.041	0.644	0.712	0.073	0.250	0.897	0.585	0.671
Observations	5,379	5,379	5,379	5,379	5,379	5,379	5,379	5,379	5,379	5,379

Note: Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. Monetary outcomes are in USD and Winsorized at the 99th percentile. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

Table 2.D3: Non-Agricultural Income

	(1)	(2)	(3)	(4)	(5)	(6)
	Casual labor		Self employment		Other income source	
	Self	Spouse	Self	Spouse	Self	Spouse
Panel A. Liberia						
Pooled cash treatment:						
Cash	-0.64**	-0.22	3.26*	1.39	0.51	0.85*
	(0.29)	(0.29)	(1.71)	(1.06)	(0.37)	(0.46)
Individual treatments by cash amount:						
Cash 250	-0.58	-0.05	5.93*	0.69	0.93*	0.23
	(0.39)	(0.49)	(3.48)	(1.07)	(0.54)	(0.54)
Cash 500	-0.95***	-0.08	3.34	3.36	-0.06	1.74**
	(0.31)	(0.47)	(2.37)	(2.56)	(0.65)	(0.87)
Cash 750	-0.38	-0.53	0.50	0.10	0.66	0.56
	(0.46)	(0.35)	(1.71)	(0.80)	(0.55)	(0.79)
Control mean	2.17	1.40	6.39	1.23	0.89	0.86
Control SD	7.01	5.66	36.52	14.71	7.42	7.67
p-value (all three equal)	0.369	0.600	0.226	0.411	0.449	0.317
Observations	2,595	2,595	2,595	2,595	2,595	2,595
Panel B. Malawi						
Pooled cash treatment:						
Cash	-0.38	0.46	-0.13	-0.05	0.22	0.38
	(0.24)	(0.39)	(0.74)	(0.55)	(0.22)	(0.46)
Individual treatments by cash amount:						
Cash 250	0.02	0.00	0.50	-0.49	0.07	2.04**
	(0.35)	(0.54)	(1.04)	(0.32)	(0.25)	(0.88)
Cash 500	-0.55*	0.76	-0.58	-0.62**	0.03	-0.26
	(0.33)	(0.54)	(1.06)	(0.31)	(0.17)	(0.60)
Cash 750	-0.62**	0.61	-0.33	1.01	0.58	-0.69*
	(0.28)	(0.55)	(1.07)	(1.45)	(0.53)	(0.41)
Control mean	2.66	2.68	4.21	0.57	0.25	1.43
Control SD	5.97	9.20	21.48	10.39	4.42	10.92
p-value (all three equal)	0.202	0.492	0.681	0.428	0.596	0.008
Observations	2,784	2,784	2,784	2,784	2,784	2,784
Panel C. Pooled						
Pooled cash treatment:						
Cash	-0.50***	0.13	1.48	0.65	0.38*	0.61*
	(0.19)	(0.24)	(0.93)	(0.59)	(0.21)	(0.33)
Individual treatments by cash amount						
Cash 250	-0.26	-0.03	2.93*	0.09	0.52*	1.18**
	(0.26)	(0.37)	(1.77)	(0.54)	(0.31)	(0.54)
Cash 500	-0.75***	0.35	1.20	1.28	0.01	0.70
	(0.23)	(0.36)	(1.27)	(1.25)	(0.32)	(0.53)
Cash 750	-0.50*	0.06	0.29	0.57	0.62	-0.08
	(0.27)	(0.33)	(1.01)	(0.84)	(0.38)	(0.44)
Control mean	2.42	2.06	5.27	0.89	0.56	1.16
Control SD	6.50	7.72	29.76	12.67	6.07	9.49
p-value (all three equal)	0.223	0.679	0.361	0.620	0.314	0.131
Observations	5,379	5,379	5,379	5,379	5,379	5,379

Note: Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. Monetary outcomes are in USD and Winsorized at the 99th percentile. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

Table 2.D4: Effects on Specific Categories of Intimate Partner Violence

	(1) Controlling Behavior	(2) Emotional IPV	(3) Physical IPV	(4) Sexual IPV
Panel A. Liberia				
Pooled cash treatment:				
Cash	-0.03	-0.09***	-0.03	-0.04***
	(0.03)	(0.03)	(0.02)	(0.01)
Individual treatments by cash amount:				
Cash 250	-0.05	-0.07*	-0.02	-0.05***
	(0.04)	(0.04)	(0.03)	(0.02)
Cash 500	0.00	-0.09**	-0.06**	-0.04**
	(0.04)	(0.04)	(0.03)	(0.02)
Cash 750	-0.06	-0.11***	-0.02	-0.04**
	(0.04)	(0.04)	(0.03)	(0.02)
Control mean	0.55	0.34	0.23	0.10
Control SD	0.50	0.48	0.42	0.31
p-value (all three equal)	0.477	0.671	0.333	0.922
Observations	1,229	1,229	1,229	1,229
Panel B. Malawi				
Pooled cash treatment:				
Cash	0.00	0.01	0.01	0.01
	(0.02)	(0.01)	(0.01)	(0.01)
Individual treatments by cash amount:				
Cash 250	0.03	0.02	0.01	0.02
	(0.02)	(0.02)	(0.01)	(0.01)
Cash 500	0.01	-0.00	0.02	0.01
	(0.03)	(0.02)	(0.01)	(0.01)
Cash 750	-0.03	0.00	-0.00	-0.01
	(0.03)	(0.02)	(0.01)	(0.01)
Control mean	0.31	0.14	0.08	0.07
Control SD	0.46	0.35	0.27	0.26
p-value (all three equal)	0.120	0.669	0.463	0.148
Observations	1,829	1,829	1,829	1,829
Panel C. Pooled				
Pooled cash treatment:				
Cash	-0.01	-0.03**	-0.01	-0.01
	(0.02)	(0.01)	(0.01)	(0.01)
Individual treatments by cash amount				
Cash 250	-0.00	-0.02	-0.00	-0.01
	(0.02)	(0.02)	(0.01)	(0.01)
Cash 500	0.01	-0.04**	-0.02	-0.01
	(0.02)	(0.02)	(0.01)	(0.01)
Cash 750	-0.04*	-0.04**	-0.01	-0.02**
	(0.02)	(0.02)	(0.01)	(0.01)
Control mean	0.42	0.23	0.14	0.08
Control SD	0.49	0.42	0.35	0.28
p-value (all three equal)	0.177	0.513	0.740	0.377
Observations	3,058	3,058	3,058	3,058

Note: Regressions include whether IPV was measured in ACASI or FTFI as well as baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses.

Table 2.D5: Effects on Types of Interpersonal Transfers

	(1)	(2)	(3)	(4)
	Transfers Sent		Transfers Received	
	Spouse	Non-spouse	Spouse	Non-spouse
Panel A. Liberia				
Pooled cash treatment:				
Cash	0.67	-0.45	2.91***	-0.34
	(0.55)	(0.38)	(0.78)	(0.47)
Individual treatments by cash amount:				
Cash 250	0.85	-0.93**	2.46**	-0.57
	(0.80)	(0.45)	(1.05)	(0.55)
Cash 500	0.43	-0.23	2.13**	-0.40
	(0.78)	(0.47)	(0.96)	(0.58)
Cash 750	0.71	-0.20	4.16***	-0.05
	(0.81)	(0.47)	(1.30)	(0.67)
Control mean	3.17	2.00	8.39	2.63
Control SD	9.80	11.29	15.00	14.26
p-value (all three equal)	0.911	0.174	0.349	0.739
Observations	1,794	2,595	1,794	2,595
Panel B. Malawi				
Pooled cash treatment:				
Cash	0.23	0.01	-0.25	0.17
	(0.19)	(0.06)	(0.73)	(0.19)
Individual treatments by cash amount:				
Cash 250	0.24	0.03	-0.20	0.06
	(0.31)	(0.08)	(1.03)	(0.26)
Cash 500	0.14	-0.08	-0.69	0.10
	(0.24)	(0.07)	(0.99)	(0.27)
Cash 750	0.31	0.08	0.17	0.34
	(0.28)	(0.09)	(1.14)	(0.28)
Control mean	0.97	0.42	9.04	1.01
Control SD	3.23	1.84	14.15	4.31
p-value (all three equal)	0.882	0.124	0.801	0.682
Observations	1,885	2,784	1,885	2,784
Panel C. Pooled				
Pooled cash treatment:				
Cash	0.44	-0.23	1.32**	-0.09
	(0.28)	(0.19)	(0.54)	(0.25)
Individual treatments by cash amount				
Cash 250	0.54	-0.44*	1.14	-0.25
	(0.42)	(0.22)	(0.75)	(0.29)
Cash 500	0.27	-0.17	0.69	-0.15
	(0.39)	(0.23)	(0.69)	(0.31)
Cash 750	0.51	-0.06	2.17**	0.15
	(0.42)	(0.23)	(0.87)	(0.36)
Control mean	2.06	1.19	8.72	1.80
Control SD	7.35	8.01	14.57	10.44
p-value (all three equal)	0.843	0.210	0.316	0.556
Observations	3,679	5,379	3,679	5,379

Note: Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Standard errors clustered at the village level in parentheses. Monetary outcomes are in USD and Winsorized at the 99th percentile.

Appendix E: Lump-sum and Flow Payments

Table 2.E1: Balance between Lump-sum and Flow within Matched Treatment Group

	(1)	(2)	(3)	(4)	(5)	(6)
	Liberia Wave 1			Liberia Wave 2		
	Lump-sum	Flow	p-value: difference	Lump-sum	Flow	p-value: difference
Panel A. Demographics						
=1 if female	0.82	0.83	0.839	0.75	0.76	0.739
=1 if currently married or has partner	0.77	0.84	0.135	0.91	0.90	0.777
Age	37.76	37.63	0.937	38.78	39.59	0.387
	[13.71]	[13.35]		[13.29]	[13.41]	
Years of education	1.93	1.56	0.275	3.14	3.39	0.372
	[3.12]	[2.80]		[3.85]	[3.96]	
Number of household members	4.09	4.39	0.179	4.68	4.95	0.077*
	[1.85]	[2.03]		[2.08]	[2.23]	
Panel B. Primary outcomes measured at baseline						
Food security index (z-score)	-0.45	-0.43	0.858	0.20	0.16	0.488
	[0.91]	[0.94]		[0.93]	[0.90]	
Total expenditure (monthly)	41.29	45.47	0.273	54.85	51.69	0.264
	[31.77]	[34.62]		[42.98]	[37.72]	
Food expenditure (monthly)	18.27	22.36	0.035**	20.97	19.59	0.204
	[16.37]	[17.37]		[16.25]	[14.78]	
Net value of durables, livestock, and financial assets	66.07	44.08	0.131	132.17	115.74	0.242
	[158.96]	[81.23]		[210.49]	[189.96]	
Non-agricultural income (monthly)	5.90	6.26	0.777	9.05	8.36	0.651
	[10.35]	[11.94]		[22.83]	[20.15]	
=1 if any IPV (past year)	0.33	0.37	0.515	0.55	0.56	0.738
Transfers received (monthly)	10.91	11.08	0.962	10.44	16.58	0.016**
	[17.09]	[14.42]		[12.54]	[21.82]	
Transfers sent (monthly)	7.12	9.67	0.521	13.29	14.90	0.601
	[13.70]	[20.99]		[23.46]	[25.72]	
Observations	151	153		393	430	

Note: Columns 1 and 4 present the mean for the subgroups for which we a match in GiveDirectly's database and are assigned to the lump-sum payment schedule; Columns 2 and 5 report the mean for those in the flow payment schedule; and Columns 3 and 6 report the p-values for testing mean difference. Standard deviations are in square brackets. Monetary outcomes are in USD and Winsorized at the 99th percentile. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

Table 2.E2: First Stage for Lump-sum / Flow Randomization

	(1) =1 if enrolled as Flow in GiveDirectly database
=1 if assigned to Flow	0.79***
	(0.02)
Assigned to Lump-sum: Mean	0.10
Observations	823

Note: This table is restricted to Liberia Wave 2 only. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

Table 2.E3: Difference in treatment effects of "lump-sum" and quarterly transfers

	(1) Food Security Index[a] (past year)	(2) Food Expenditures (past month)	(3) Non-food Expenditures (past month)	(4) Non-agricultural Income[b] (past month)
Panel A. Liberia Wave 1				
Pooled flow effect:				
Flow payments	0.04	-0.16	3.24	0.35
	(0.15)	(1.99)	(3.55)	(1.39)
Individual flow effects by cash amount:				
Cash 250 in Flow	0.01	-0.85	13.36*	4.08*
	(0.24)	(4.93)	(6.66)	(2.26)
Cash 500 in Flow	0.36	1.36	-3.35	-1.16
	(0.22)	(2.46)	(6.13)	(2.47)
Cash 750 in Flow	-0.31	-1.05	-0.67	-1.83
	(0.29)	(2.39)	(4.78)	(2.25)
Lump-sum: mean	0.34	22.53	29.01	5.71
Lump-sum: SD	1.13	20.16	31.31	14.41
p-value (all three equal)	0.172	0.776	0.149	0.158
Observations	304	304	304	304
Panel B. Liberia Wave 2				
Pooled cash treatment:				
Flow payments	0.07	0.01	1.27	-0.48
	(0.08)	(1.49)	(3.05)	(2.92)
Individual flow effects by cash amount:				
Cash 250 in Flow	0.20	2.11	3.55	-0.66
	(0.15)	(2.45)	(5.02)	(7.08)
Cash 500 in Flow	0.14	-2.38	-2.22	-2.22
	(0.13)	(3.03)	(5.45)	(3.48)
Cash 750 in Flow	-0.10	0.65	3.26	1.12
	(0.11)	(2.03)	(5.27)	(3.49)
Lump-sum: mean	0.26	30.09	40.07	11.47
Lump-sum: SD	1.08	22.77	44.96	38.04
p-value (all three equal)	0.196	0.515	0.692	0.796
Observations	823	823	823	823
Panel C. Pooled				
Pooled cash treatment:				
Flow payments	0.06	-0.05	1.75	-0.32
	(0.07)	(1.18)	(2.43)	(2.15)
Individual flow effects by cash amount				
Cash 250 in Flow	0.15	1.44	6.11	0.61
	(0.13)	(2.21)	(4.14)	(5.32)
Cash 500 in Flow	0.20*	-1.40	-2.43	-1.91
	(0.11)	(2.28)	(4.27)	(2.59)
Cash 750 in Flow	-0.17	0.07	2.16	0.28
	(0.11)	(1.61)	(4.09)	(2.62)
Lump-sum: mean	0.28	27.99	37.00	9.87
Lump-sum: SD	1.09	22.32	41.89	33.30
p-value (all three equal)	0.045	0.673	0.360	0.807
Observations	1,127	1,127	1,127	1,127

Note: Quarterly payments were implemented only in Liberia. The endline was conducted about 18-22 months after first transfers were received in Liberia. Regressions include baseline measurement, strata fixed effects, and indicator for market access treatment. Monetary outcomes are in USD and Winsorized at the 99th percentile. Standard errors clustered at the village level in parentheses. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

[a] Food Security Index is standardized z-score of HDDS, FCS, HHS (negatively weighted), and FIES (negatively weighted), using inverse covariance weighting (Michael L. Anderson 2008b) relative to the control mean and SD in each country.

Chapter 3

Exhaustive or Exhausting? Evidence on Respondent Fatigue in Long Surveys

3.1 Introduction

Many of the surveys that are administered in development economics or by multilateral agencies such as the World Bank to measure poverty or as part of evaluations are long and complicated, and require the sustained attention of a respondent for several hours. For any researcher who has observed such a survey, it is clear that some respondents disengage as the survey drags on, because they are exhausted, bored, or because their attention wanders. As a result, response quality during the later part of a long survey may suffer, a phenomenon known as survey fatigue.

While survey fatigue is well-documented in the literature,[1] until recently there has been comparatively little research to rigorously quantify its effects. In this paper, we provide such a quantification by randomizing the order in which modules appear in a long survey, generating exogenous variation in the time-into-survey when a particular question was asked. This random order of questions allows us to compare responses to the *same* question when it is asked sooner in the survey versus when it is asked later, and quantify the divergence in responses. We conduct this experiment within surveys administered at baseline and endline for a randomized evaluation of cash transfers in rural Liberia and Malawi (Aggarwal et al. 2022). These surveys were long, averaging about 2.5 hours, and the experimental randomization induced meaningful variation in the time it took to reach a specific question: the average time to reach a specific question was changed by as much as about 30 minutes as a result of the randomization.

We have two main findings. First, and consistent with other work, we find clear evidence of survey fatigue. We estimate survey fatigue separately for two

[1]For example, survey fatigue has its own entry in the Encyclopedia of Survey Research Methods.

ways of asking questions. The first is an "open-ended" method which we used for the questions in which there is no top code or pre-listed set of options. For example, for transfers given out, respondents were asked to provide the number of transfers that they gave, and could list as many or as few as they wanted. For such questions, we find that each additional hour of surveying causes a 26-64% decrease in the number of items listed. The second method, or "fixed list" method, is one in which the list of items was pre-coded. For example, in the food expenditures section, we generated a list of around 35 food items, and asked about each of these items separately. Survey fatigue might be reduced with this method, if the listing serves as a memory aid for those who need help with recall later in the survey as they begin to tire out. However, we still observe survey fatigue in this method, though much less than in the prior method: for every additional hour, respondents are about 10-19% more likely to report no value for a given item. While survey fatigue appears less prevalent when using the fixed list method, we are unable to definitively attribute this to the question type, since the method is not random – it is also possible that these categories are less subject to survey fatigue.

Second, we quantify the extent to which this skipping reduces the *value* of aggregate categories such as the total value of transfers or expenditures. For any skipped question, the value of that category would be set to zero by default, and so we would expect survey fatigue to lower aggregated values. This effect might be modest if the categories that are skipped tend to be more marginal. However, the effects we find are sizeable: for example, an additional hour of survey time reduces the value of food expenditures by 25%, and has even larger effects (in percentage terms) on smaller categories (such as transfers).

This paper contributes to a recent literature that experimentally evaluates the

effect of survey time on survey fatigue. Laajaj and Macours (2021) randomize the order of cognitive, non-cognitive and technical questions in a sample of farmers in Western Kenya but, unlike us, find no effect of survey time on reporting. Two other experiments were conducted contemporaneously to this study, and find similar results to ours. Ambler et al. (2021) randomize the order of a household labor supply module, where questions are asked about the labor supply of each household member, but the order in which the household members are listed was randomized. The authors find a 2% reduction in the number of activities reported when a household member is moved back by one position in the household roster. Abay et al. (2021) employ a methodology similar to ours, in which the authors randomize the placement of a dietary diversity module within a phone survey in Ethiopia. Like us, they find large effects: a 15 minute increase in survey time before the module leads to an 8-17% decline in reported dietary diversity.[2] Finally, in a similar but different design and different context, Backor et al. (2007) conduct a web-based time-use survey in the US in which an extra question is included at a random order, creating variation in how many hours had already been asked about when a particular question appeared in the survey. Similar to these other papers, the authors find that an additional hour lowers the number of activities reported in each subsequent hour by 5 percentage points.

While our experiment was not designed to explore *why* survey fatigue occurs, our data offers some suggestive evidence. Past research suggests that survey fatigue may be driven by people deliberately choosing to not answer questions in order to expedite the end of the survey, or if people become more likely to

[2]Another related paper is Kilic and Sohnesen (2019), who find that poverty incidence differs when measured in a short or a long survey in Malawi. However, in their case, since everybody got the same long survey or the same short survey, it is not possible to disentangle the effects of survey length from those of question order, i.e., when your responses are impacted by a question being preceded by another question (see here).

inadvertently make mistakes as they become tired. Some researchers have also conjectured that, over time, respondents learn that answering "no" to a question often invokes a "skip code" that will allow them to skip a number of follow-up questions. This behavior, known as "satisficing", has been documented in survey settings (Krosnick 1991). We have two pieces of evidence on this point. First, besides our in-person baseline and endline surveys, we also randomized the order of modules within phone surveys that we conducted with a subset of respondents repeatedly every 2 months. These surveys took about 30-40 minutes to complete. We only introduced question-order randomization in the phone surveys more than a year after the phone surveys had started, when each respondent had already answered several rounds of the phone survey. Therefore, at the time of the phone survey experiment, we would expect that respondents were already familiar with the structure of the surveys, including the mechanics of skip patterns, over time as they go through multiple rounds of the survey. If respondents were satisficing, they would answer fewer questions from the outset during the later rounds of the phone surveys, and there would be no evidence of experimental survey fatigue *within* a survey round. However, we find evidence of survey fatigue similar to our baseline and endline surveys, suggesting that this behavior is likely driven by cognitive burden as the survey progresses. On the other hand, we find some evidence that satisficing may also be at play. When we examine survey fatigue by topic, we find effects for both more and less memorable items; whereas if recall issues were the only channel, we would expect stronger fatigue effects for more easily forgettable categories (such as details of expenditures, as opposed to durable goods or livestock ownership). Our evidence therefore suggests both channels may be at play, though we leave a more definitive analysis to future work.

Finally, since our survey experiment is layered on top of another experimental

study (of cash transfers), we attempt to examine whether survey fatigue systematically reduces the measured treatment effects of the primary intervention. Our hypothesis is that the measured treatment effects will likely be attenuated in the presence of fatigue if one of the treatment arms has systematically more to report, for example, in Aggarwal et al. (2022), the cash transfer treatment arm reports having more assets. We find mixed evidence of the hypothesized attenuation, which is ultimately, entirely inconclusive as we are not sufficiently powered for this analysis. We leave this question to future research.

The rest of this paper proceeds as follows. Section 3.2 explains the data and experimental design, Section 3.3 presents results, and Section 3.4 concludes.

3.2 Data and Experimental Design

3.2.1 Setting

We use data from baseline and endline surveys conducted as part of a cash transfer RCT with the NGO GiveDirectly in Liberia and Malawi. In the experiment, the treatment group received cash transfers via mobile money. The average amount of the transfer was $500; however, the amount and other implementation details were varied experimentally – see our trial registry on the AEA website (Aggarwal et al. 2021a) and the paper describing the main experimental results (Aggarwal et al. 2022) for more details on the design of the underlying experiment.[3]

In each country, the project took place in rural areas, with universal targeting

[3]In both countries, the size of the transfer was varied between $250, $500, and $750. In addition, in Liberia, cash was disbursed either as a "lump-sum" or via quarterly payments. However, even the lump sum was disbursed in increments of $250 per month, so that cash was paid out over 3 months for the largest transfer.

in treatment villages (i.e. all households in treatment villages received transfers). For this reason, the total allocation to a village depends on its size; to ensure liquidity, the NGO decided to only include villages which were small. Operationally, we set a population threshold based on the most recent population census.[4] In Liberia, the study takes place in Bong and Nimba Counties; in Malawi, it takes place in Chiradzulu and Machinga Districts. In each country, the project enrolled 300 villages, with half selected for treatment.

In each village, we attempted to enroll 10 households into the survey sample.[5] We chose to target women for the study, though many questions were asked at the household level. Male heads were interviewed only when the female was not present, and would not be reachable within a few days; our sample was ultimately 76% female in Liberia and 94% in Malawi.

Two of the 10 sampled households in each village were further randomly sampled to participate in a monthly panel survey that was conducted over the phone and was designed to measure a pre-defined set of outcomes at a high frequency. While the major focus of these surveys was to measure food security, they also included questions on income, labor supply, transfers, savings, and credit. We designed these surveys such that each household was called every other month, but the 2 households in each village alternated months, such that each village provided a data point every month. The phone surveys took about 30-40 minutes to complete.

Figure 3.A1 shows the timeline of project activities.

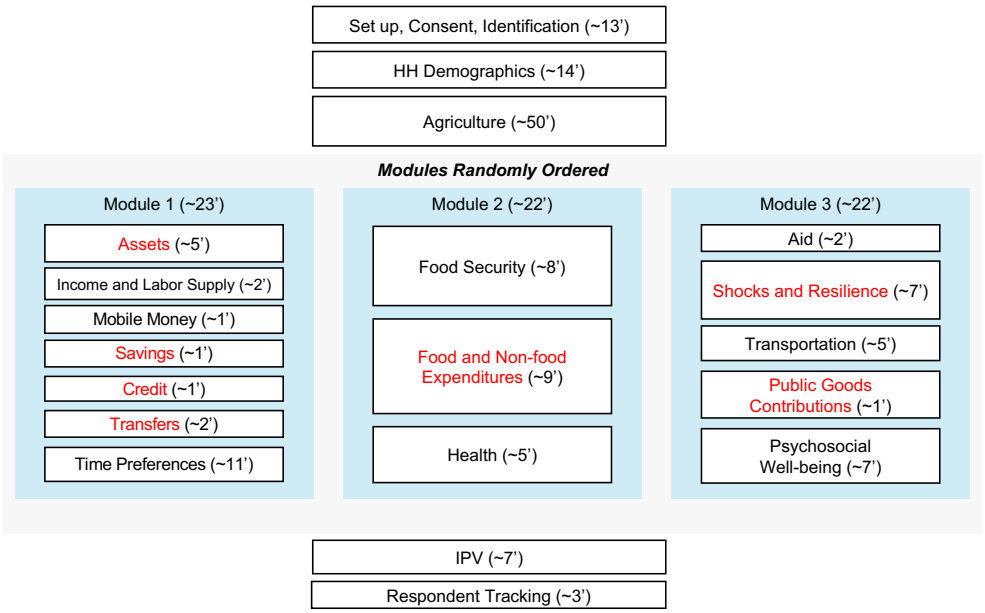
[4]In Malawi, the upper threshold was 100 household per village according to the 2008 national census. In Liberia, we conducted the experiment in two cohorts; the first cohort included villages that had up to 25 households in the 2008 national census, and the threshold for the second cohort was 125, reflecting the larger village sizes in the study region.

[5]It was not always possible to enroll 10 households per village. The total sample size is 2,715 in Liberia and 2,944 in Malawi

3.2.2 Question order randomization

This experiment takes place within baseline and endline surveys which are similar to the World Bank's LSMS surveys and take about 2.5 hours to complete on average. The surveys contain 19 self-contained sections, including household demographics, agriculture, income, expenditures, savings, assets, labor supply, shocks, and other topics. We show the full list of sections in Figure 3.1.

Figure 3.1: Sections in In-person Surveys

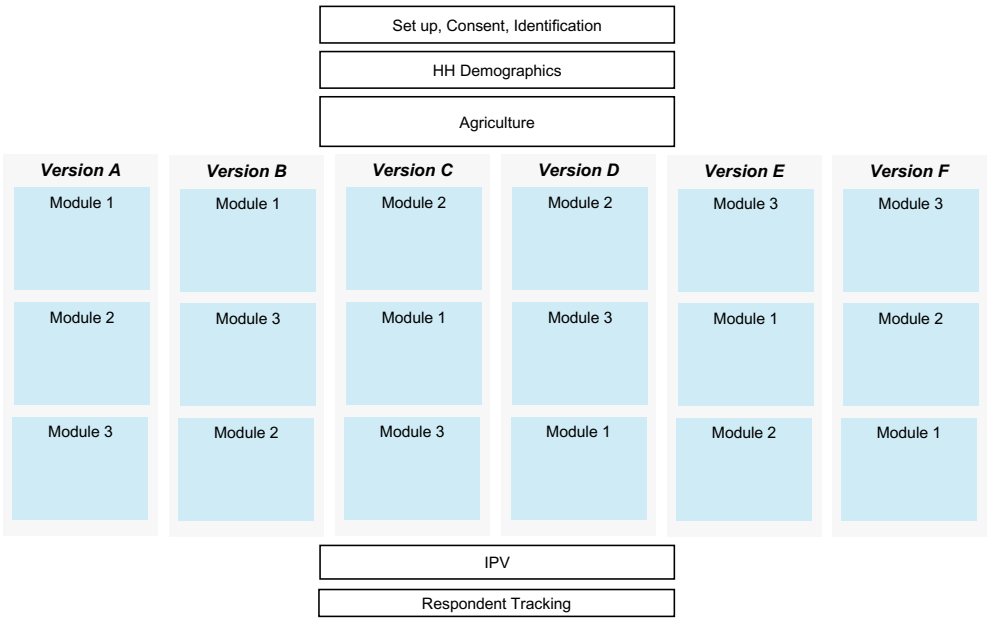


Note: Approximate duration for each section (in minutes) are reported in parentheses. In red are the sections for which survey questions are relevant for analysis in this paper.

The beginning of the survey (which included household identifying information, demographics, and agriculture) and the end of the survey (which had a section on intimate partner violence, followed by the collection of household tracking information) were the same across all versions. The remaining sections were grouped into 3 modules, and the order of these 3 modules was randomized, giving

us 6 versions of the survey (which we refer to as versions A-F – see Figure 3.2). The survey software records the amount of time elapsed (since beginning) at each question, allowing us to calculate the exact time at which a question appeared in the survey.

Figure 3.2: Randomized Order of Modules in In-person Surveys



Note: A respondent is randomly provided with one among Versions A-F. For every version, survey set-up, demographics, and agriculture come at the beginning, while IPV and respondent tracking are at the end.

The amount of time it takes to progress through the survey varies depending on a number of factors, including respondent and enumerator characteristics, and the details of a household's circumstance. For example, because our survey had a focus on agriculture, a household which grew multiple crops would be asked a number of questions about each one of them. Table 3.A1 shows information on the average survey duration. The baseline and endline surveys took on average 2.3 and 2.7 hours respectively in Liberia; and 3 and 2.8 hours respectively in

Malawi. The standard deviation in survey time is sizeable, ranging from 0.7 to 1.1 hours. Figure 3.A2 shows a CDF of the time until completion of different points of the survey (using survey Version A only) for both countries and for both baseline and endline pooled together (i.e., for 4 country-survey combinations). The figure shows CDFs for various quantiles in the survey time distribution (i.e. relative to completing the question which makes up the p-th percentile of the overall distribution of time to survey completion). The CDFs show that even 10% into the survey, the standard deviation of time is already over 30 minutes and that for all percentiles, there are surveys that take a large amount of time. For example, about 10% of people take over 3 hours to even get halfway through the survey (Panel C).

Finally, although not the main focus of this paper, we also randomized survey order for the final 2-3 rounds of the phone survey. In order to do this, we randomized the location of the Expenditures and Transfers sections to appear at either the very beginning or the very end of the survey, and the order between the two sections, generating 4 possible permutations (Figure 3.A3). We return to this randomization in the discussion section, when we discuss possible explanations for survey fatigue.

Table 3.1 shows the effect of the randomized survey versions on the time until which the first question of each section was administered. The reported means and standard deviations at the bottom of the table are those pertaining to that section for Version A of the survey. As can be seen from this table, the module randomization introduced significant variation in the time-into-survey when a section starts. For example, looking at Column 1, we can see that the Assets section started just after the 80th minute on average for those who got Version A of the survey. However, the full range for when this section started ranges from

77th minute (version B) to 106th minute (Version F) - a difference of around 30 minutes. Similar range of difference is consistently observed across all sections.

We use the survey version that was used for each respondent as an instrument for the time-into-survey when a particular set of questions began to be asked of that respondent. While the validity of module randomization as an instrument is largely intuitive, we also show this formally: first-stage F-statistics are at the bottom of Table 3.1, and range from 35 to almost 200.

Table 3.1: Experimental variation in time before which sections were administered

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Time into survey (minutes) at the beginning of following section:						
	Module 1				Module 2	Module 3	
	Assets	Savings	Credit	Transfers	Expenditure	Shocks	Contributions
Version B	-3.30***	-2.84**	-2.79**	-2.63*	6.39***	-12.20***	-8.83***
	(1.28)	(1.36)	(1.36)	(1.37)	(1.37)	(1.47)	(1.50)
Version C	19.21***	17.71***	17.54***	17.67***	-16.83***	-4.74***	-3.24**
	(1.27)	(1.36)	(1.36)	(1.37)	(1.36)	(1.46)	(1.50)
Version D	23.96***	22.52***	22.24***	22.35***	-18.00***	-16.52***	-5.45***
	(1.27)	(1.36)	(1.36)	(1.37)	(1.36)	(1.46)	(1.50)
Version E	6.61***	6.01***	5.67***	6.04***	5.79***	-25.74***	-15.79***
	(1.27)	(1.35)	(1.36)	(1.37)	(1.36)	(1.46)	(1.50)
Version F	26.06***	24.56***	24.01***	24.07***	-8.38***	-27.57***	-14.49***
	(1.28)	(1.36)	(1.36)	(1.38)	(1.37)	(1.47)	(1.50)
Version A: Mean	80.01	93.47	93.89	95.61	109.39	125.53	134.72
Version A: SD	38.78	40.34	40.26	40.87	44.23	47.39	49.78
F-statistic: joint significance	197.30	151.42	146.55	143.33	127.92	114.25	35.05
Number of respondents	5,591	5,597	5,597	5,597	5,597	5,597	5,592
Observations	10,153	10,226	9,952	10,228	10,227	10,224	10,154

Note: The omitted group is Version A. Observations include in-person baseline and endline survey data. Regressions include a survey fixed effect (i.e. baseline and endline, for each country separately, as well as differentiating Wave 1 and 2 in Liberia). ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

3.2.3 Respondent characteristics and randomization check

Table 3.2 presents summary statistics for several basic demographic indicators, as well as comparisons across treatment groups. We present these statistics only for those indicators which were asked before the module randomization kicked in as the variables from the later sections would by definition be imbalanced under our central hypothesis for this paper. We show the balance across versions separately for the baseline and endline surveys, but pool them across the 2 countries. For each survey (baseline or endline), we show the mean and standard deviation (for non-binary variables) pertaining to Version A of the survey (chosen arbitrarily), followed by the p-value for the joint test of equality across all 6 versions of the survey. Panel A shows respondent characteristics. Almost 90% of the sample is female, three-quarters are married, and the average age is 41. Average years of education (for the respondent) is only 4.2, and 57% are literate (these last 2 variables were measured at baseline only).

Panel B shows household characteristics. At baseline, the average household has 4.8 members, and 96% were engaged in farming. About 40% of the sample live in a house with a thatch roof, and 80% live in a house with a mud floor. About 77% own their dwelling and only 2% have electricity. We cannot reject equality across treatments for all of these variables. Finally, Panel C shows the other experimental treatments. Cash was randomly given out to 50% of villages (and given that we sampled about 10 households per village, it was given, by design, to roughly 50% of the respondents). The phone surveys were administered to about 20% of the respondents. As expected, the survey experiment is orthogonal to both of these cross-randomized treatments.

Table 3.2: Summary Statistics and Randomization Check

	Baseline Survey		Endline Survey	
	Version A (Mean/SD)	p-value: test of equality over 6 versions	Version A (Mean/SD)	p-value: test of equality over 6 versions
Panel A. Respondent Characteristics				
=1 if female	0.87	0.188	0.89	0.308
=1 if currently married or has partner	0.76	0.970	0.74	0.188
Age	40.50	0.661	40.95	0.388
	(15.20)		(14.31)	
Years of education	4.18	0.553		
	(3.75)			
=1 if can read/write in English	0.57	0.667		
	(0.50)			
Panel B. Household Characteristics				
Number of household members	4.77	0.436	4.98	0.744
	(2.11)		(2.16)	
=1 if household engaged in farming past year	0.96	0.786	0.90	0.803
=1 if thatch roof	0.40	0.206	0.24	0.780
=1 if mud/dirt floor	0.80	0.848	0.77	0.392
=1 if owns dwelling	0.77	0.844	0.77	0.840
=1 if has electricity in dwelling	0.02	0.280	0.02	0.523
Panel C. Cross-randomized groups				
Cash Treatment Group	0.53	0.216	0.51	0.914
Phone survey group	0.21	0.640	0.22	0.655
Observations	4,879		5,349	

Note: Column 1 and 3 (Version A) represent control mean with standard deviation in parentheses. Columns 2 and 4 present p-values from the joint test of equality of the means for all the 6 survey versions, A-F.

3.3 Results

3.3.1 Quantifying survey fatigue

We start by examining the impacts of time-into-survey on the count of items or instances reported in response to the open-ended questions (questions described in Figure 3.A4). To do this, we run the following regression:

$$Y_{ics} = \beta Hours_{ics} + \phi_s + \varepsilon_{ics}, \tag{3.1}$$

These regressions are run separately for each category of questions (specifically, ROSCAs, VSLAs, transfers received, transfers sent, and credit purchases). Within each category, the unit of observation is at the *respondent-survey* level (i.e. there are 2 surveys for most respondents, baseline and endline, for each country). In the regression, Y_{ics} refers to the count of items reported by survey respondent i within category c in country-survey sample s, $Hours_{ics}$ denotes elapsed time into survey (in hours) at which category c is administered to respondent i, instrumented with the randomized module order (Versions A-F) that was fielded to the respondent, ϕ_s represents a survey fixed effects (i.e. country, baseline/endline, Waves 1 and 2 in Liberia), and ε_{ics} is the error term.

In this analysis, there is no reason to expect heterogeneity in responses based on outcomes – *ex ante*, we expect similar results for any question category. Therefore, to discipline our analysis, we present results exhaustively for every relevant outcome, and adjust the standard errors to account for a false discovery rate (FDR) using the procedure in Michael L Anderson (2008a). For each outcome, we present only q-values from this procedure, and statistical significance is ascertained only based on the q-values obtained after FDR correction.

Finally, please note that for ease of exposition, we run our analyses and interpret results in terms of 1-hour delays in the survey. It may be useful however, to slightly scale down these effects as the actual survey delays that we observe are slightly more modest, as shown in Table 3.1.

We present these results in Table 3.3. We show 5 outcomes: the number of Rotating Savings and Credit Associations (ROSCAs) and Village Savings and Loan Associations (VSLAs) that the respondent reported being part of in the savings section; the reported number of transfers received and given during the

past month; and the number of credit purchases during the past month.[6] Four out of 5 of these outcomes are statistically significant at 10% (and 2 are significant at 5%), even with the FDR adjustment. The effect sizes are large: an extra hour reduces the number of items by 26-64%. Because these surveys average 2.5 hours, this implies that the decision to place a question at the beginning rather than the end of the survey can have a large effect.

Table 3.3: Survey time and the number of items reported ("Open-ended" questions)

	(1)	(2)	(3)	(4)	(5)
	Number of distinct items reported for the following:				
	ROSCAs	VSLAs	Transfers received	Transfers given	Credit purchases
Hours into Survey	0.002	-0.058**	-0.074*	-0.209***	-0.095*
	[0.613]	[0.042]	[0.081]	[0.001]	[0.081]
Dependent variable: Mean	0.056	0.205	0.275	0.328	0.366
Hours into Survey: Mean	1.9	1.9	1.9	1.9	1.9
Hours into Survey: SD	1.0	1.0	1.0	1.0	1.0
Number of respondents	5,596	5,597	5,596	5,594	5,597
Observations	10,225	10,224	10,223	10,215	10,228

Note: There is 1 observation per respondent per survey. Baseline and endline surveys are pooled in each country, so for most individuals there are 2 observations. We report TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include a survey fixed effect (i.e. baseline and endline, for each country separately, as well as differentiating Wave 1 and 2 in Liberia). See Table 3.B1 for results by country and Table 3.C1 for results by survey type (baseline/endline). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

[6]For both transfers and credit purchases, some earlier survey versions included questions recalling for the past 3 months instead. Later for analysis on aggregated values, the monetary values collected from these versions are divided by 3, comparable to the past-month values.

Next, we investigate the impacts of elapsed survey time on choosing an item in questions asked via the fixed-list method (questions described in Figure 3.A5). Recall that our hypothesis is that going through a pre-set list of items may serve as an aid to memory (for example, it may be easier to remember if the enumerator asks the respondent whether her household consumed say, bananas in the past week than it would be to recall if the enumerator asks her to list all the items that the household consumed in the past week). We run the following regression:

$$Y_{icsj} = \beta Hours_{ics} + \phi_s + \varepsilon_{icsj}, \tag{3.2}$$

The main difference for this approach is that, for each category, there are multiple *items* where Y_{iscj} is a binary indicator of whether respondent i in survey sample s responded "yes" to having consumed/bought/experienced item j in category c of the survey, $Hours_{ics}$ elapsed time into survey (in hours) at the beginning of category c, instrumented with the randomized module order (Versions A-F), ϕ_s survey fixed effects (i.e. country, baseline/endline, Waves 1 and 2 in Liberia), and ε_{iscj} the error term. Like before, we adjust the standard errors for multiple testing, and report only the FDR-corrected q-values in our tables.

Table 3.4 presents this analysis for a set of 9 categories: livestock, farm tools, durable goods, savings, loans, food expenditures, non-durables expenditures, household shocks, and public goods contributions. Note that these regressions are at the category-item level, and so are much better powered than the previous set of regression results: we find that 4 of 9 outcomes are significant at 5% (and even of those not significant, nearly all are negative signed).[7] As we hypothesized, effect sizes are more moderately measured than for the "open-

[7]See Appendix B and Appendix C for heterogeneity in these results by country and by survey type (baseline or endline).

ended" questions, ranging from 10-19% for the statistically significant outcomes. Nevertheless, survey fatigue is clearly evident here as well.[8]

Table 3.4: Survey time and the probability of reporting an item ("Fixed list" questions)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	=1 if item is selected (not skipped):								
	Livestock	Farm tools	Durable	Savings	Loans	Food expend	Non-durables	Shocks	Public goods
Hours into Survey	-0.007*	-0.004	-0.001	-0.002	0.000	-0.025***	-0.038***	-0.025***	-0.002
	[0.081]	[0.236]	[0.613]	[0.555]	[0.613]	[0.001]	[0.001]	[0.001]	[0.613]
Dependent variable: Mean	0.072	0.154	0.176	0.060	0.020	0.203	0.249	0.130	0.050
Hours into Survey: Mean	1.7	1.7	1.8	1.9	1.9	1.9	1.9	2.0	2.0
Hours into Survey: SD	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2
Number of items: Liberia	11	21	20	12	14	37	11	16	9
Number of items: Malawi	15	20	22	11	14	35	11	17	9
Number of respondents	5,594	5,594	5,594	5,597	5,597	5,597	5,597	5,597	5,349
Observations	134,831	208,281	212,373	114,045	138,711	366,947	112,497	166,524	48,141

Note: Each column represents a different category of questions in the survey. Each category includes multiple items (e.g., livestock includes 11 types of animals). We report TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include a survey fixed effect (i.e. baseline and endline, for each country separately, as well as differentiating Wave 1 and 2 in Liberia). See Table 3.B2 for results by country and Table 3.C2 for results by survey type (baseline/endline). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets

One advantage that our data provides over the remainder of the literature on this topic is that we have repeated observations of the same person as our phone surveys are a panel, and even our in-person measurements were taken twice, at baseline and at endline (except for Liberia Wave 1, for which the survey order experiment was introduced only for endline surveys). We can use these repeat measurements in a fixed-effects set-up to control for all individual specific traits that may affect survey responses. We show these in Appendix D for our phone

[8]Please note, however, that in both Table 3.3 and Table 3.4, the effect sizes in percent terms are slightly overestimated due to the fact that the dependent variable means are calculated across all versions and are therefore, depressed due to survey duration effects. Nevertheless, the effects are large enough in an absolute sense to be economically meaningful.

surveys as well as the in-person surveys. We find no meaningful differences in these tables relative to the regressions without fixed effects.

Finally, we hypothesize that survey fatigue may not not evolve linearly, but instead, there may be an inflexion point beyond which there is a change in the slope. We investigate this in Appendix E, where, for the outcomes which show significant effects of fatigue in Table 3.3 and Table 3.4, we show a scatter plot and a non-linear fit through these scatter points. The evidence varies, depending on the outcome in question, although the underlying scatter points suggest that a linear fit provides a good approximation of respondent behavior.

We note however, that our range of hours into survey begins only at 1.5 hours (or more) as the initial sections were fixed across all respondents. It is possible, therefore, that non-linearities may have set in before then. As a result, we leave a fuller investigation of non-linear effects of fatigue to future research.

3.3.2 Effect of survey fatigue on aggregated values

The prior section implies that aggregated values of categories such as expenditures or transfers will be attenuated by survey fatigue; in this section, we quantify this attenuation. We run regressions identical to Equation 3.1, except that the dependent variable is now in dollar amounts, rather than counts; in addition, results are shown for both open-ended and fixed list questions.

Results are shown in Table 3.5. We find that the vast majority (9 of 11) of point estimates are negative, more than half of which (5) are significant at conventional statistical significance levels despite being corrected for multiple hypothesis testing. In addition, 2 of the coefficients - those for farm tools and public goods - are marginally significant at 17% and 13% respectively. Moreover, the effect sizes

are economically meaningful. Focusing on just the statistically significant effects, the coefficient magnitudes range from 25% of the mean (for food expenditure) to 86% (for transfers given).

In some cases, effect sizes for reported monetary values (as shown in Table 3.5) are much larger in percent terms than they are for the counts that were collected via the open-ended and the fixed-list questions in Table 3.3 and Table 3.4, respectively. This is especially true for some of the small categories such as transfers given, where an extra hour reduces the value by \$0.59, on a base of just \$0.69, or 86%, while the effect of an hour on the count in Table 3.3 is a reduction of 0.21 transfers on a base of 0.33 (or 64%). But even for a larger category like food, the percent decline in value is 25%, compared to 12% in skipping in Table 3.4. One possible explanation is that fatigue causes respondents to report lower values (because the value questions come after the counts). This is consistent with studies such as Brzozowski et al. (2017), who show that recall errors in surveys tend to not be mean zero, but are in fact, negatively correlated with true behavior - i.e., when respondents make mistakes, they tend to overstate the low values and understate the high values.

Table 3.5: Survey Time and Reported Total Monetary Value of Aggregated Categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Total value of reported items for the following:										
	Livestock	Farm Tools	Durables	Savings	Loans	Food Expend	Non-durables	Public goods	Transfers received	Transfers given	Credit purchases
Hours into Survey	-13.47	-1.32	4.68	-1.47	0.73	-4.12***	-2.52***	-0.10	-0.51**	-0.59***	-0.65***
	[0.344]	[0.172]	[0.344]	[0.367]	[0.344]	[0.001]	[0.001]	[0.130]	[0.011]	[0.001]	[0.002]
Dependent variable: Mean	95.78	10.48	58.11	15.52	6.40	16.22	7.93	0.14	0.95	0.69	0.81
Hours into Survey: Mean	1.7	1.6	1.8	1.9	1.9	1.9	2.0	2.1	1.9	1.9	1.9
Hours into Survey: SD	1.0	1.3	1.0	1.0	1.0	1.3	1.3	1.6	1.0	1.0	1.0
Number of respondents	5,594	5,349	5,594	5,597	5,597	5,597	5,597	5,349	5,597	5,597	5,597
Observations	10,189	5,349	10,189	10,226	9,952	10,227	10,227	5,349	10,228	10,228	10,228

Note: All values in USD. There is 1 observation per respondent per survey. Baseline and endline surveys are pooled in each country, so for most individuals there are 2 observations. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include a survey fixed effect (i.e. baseline and endline, for each country separately, as well as differentiating Wave 1 and 2 in Liberia). For transfers and credit purchases, some earlier survey versions included questions recalling for the past 3 months instead of past month. See Table 3.B3 for results by country and Table 3.C3 for results by survey type (baseline/endline). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

3.3.3 Effect of survey time on estimated treatment effects of cash

An important implication of these results is that the effects of any program might be attenuated if effects are measured later in the survey. This may happen through two distinct channels: (1) survey fatigue may proportionally reduce the number of items mentioned by respondents, in which case treatment-control differences will become smaller (in absolute value, though not in percentages) if measured later in the survey; or (2) if there exist non-linearities, for example if there is some threshold level of cognitive load that the treatment group is more likely to encounter because they have more to report, treatment effects can be attenuated in both absolute and percentage terms.

To understand the interaction of fatigue with the primary treatment, we examine if the effect of the cash transfer differs when outcomes are measured later in the survey, by regressing outcomes on cash, time into the survey, and their interaction. Specifically, we run the following regressions analogous to Equation 3.1

and Equation 3.2, but with cash interactions.

$$Y_{isc} = \beta Hours_{isc} + \gamma Cash_{v(i)} + \kappa Cash_{v(i)} \times Hours_{isc} + \phi_s + \psi_m + \varepsilon_{isc}, \quad (3.3)$$

$$Y_{iscj} = \beta Hours_{isc} + \gamma Cash_{v(i)} + \kappa Cash_{v(i)} \times Hours_{isc} + \phi_s + \psi_m + \varepsilon_{iscj}, \quad (3.4)$$

where $Cash_{v(i)}$ denotes whether a village v received cash transfers, ϕ_s represent country-wave sample fixed effects,[9] ψ_m represent fixed effects for the cash randomization strata. All other notation is the same as before. In these regressions, we demean the hours variable.

Please note that there is also an alternative way of interpreting these regressions - which is if the cash transfer treatment has an effect on fatigue. This could happen if, for example, better nutrition afforded by the cash improves respondents' cognitive capacity. The coefficient κ will capture either effect - of cash on fatigue or of fatigue on cash treatment coefficients.

We show the results from these regressions in tables Table 3.6 for open-ended questions and in Table 3.7 for fixed-list questions; Table 3.A3 shows results for the aggregated categories.

[9]There is no survey type fixed effects separately for baseline and endline because the cash effects are measurable only at endline

Table 3.6: Effect of survey time on measured treatment effects of cash ("Open-ended questions")

	(1)	(2)	(3)	(4)	(5)
	Number of distinct items reported for the following				
	ROSCAs	VSLAs	Transfers received	Transfers given	Credit purchases
Time into Survey (hr)	-0.02	-0.13	0.03	-0.15	-0.22
	[0.594]	[0.228]	[0.705]	[0.228]	[0.228]
Cash × Time into Survey (hr)	-0.02	0.21	-0.29	-0.05	0.16
	[1.000]	[0.276]	[0.135]	[1.000]	[1.000]
Cash	0.01*	0.04*	0.02	0.05**	-0.02
	[0.087]	[0.098]	[0.152]	[0.021]	[0.178]
Control Mean	0.03	0.24	0.16	0.15	0.34
Hours into Survey: Mean	0.0	-0.0	-0.0	-0.0	0.0
Hours into Survey: SD	0.9	0.9	0.9	0.9	0.9
Observations	3,961	3,962	3,962	3,958	3,962

Note: Regressions include baseline measurement of outcome, fixed effects for cash treatment randomization strata, and country-wave fixed effects. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values (calculated from p-values based on standard errors clustered at village level) in brackets.

We find no compelling evidence of a tempering effect of fatigue on the cash effects (or of cash on the fatigue effects). We conjecture that this is perhaps because statistical power is limited since this analysis can only be conducted on the endline, effectively halving our sample size, and because the cash treatment requires standard error clustering at the village level. Moreover, the interaction effect is defined only for the cash treatment group. We leave a further evaluation of this to future work.

Table 3.7: Effect of survey time on measured treatment effects of cash ("Fixed-list questions")

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	=1 if tem is selected (not skipped)								
	Livestock	Farm tools	Durable	Savings	Loans	Food expend	Non-durables	Shocks	Public goods
Time into Survey (hr)	0.01	-0.00	-0.01	-0.01	-0.01	-0.02	-0.08	0.00	-0.01
	[0.591]	[0.766]	[0.591]	[0.256]	[0.256]	[0.591]	[0.222]	[1.000]	[0.594]
Cash × Time into Survey (hr)	-0.02	-0.01	-0.02	0.00	0.01	0.01	0.08	-0.02	0.00
	[0.304]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.276]	[1.000]	[1.000]
Cash	0.01***	0.01*	0.02***	0.01***	0.00	0.01*	0.02***	-0.01*	-0.00
	[0.001]	[0.072]	[0.001]	[0.001]	[0.266]	[0.087]	[0.003]	[0.072]	[0.376]
Control Mean	0.06	0.13	0.15	0.05	0.02	0.18	0.20	0.08	0.04
Hours into Survey: Mean	1.7	1.7	1.8	1.9	1.9	1.9	2.0	2.2	2.2
Hours into Survey: SD	0.9	0.9	0.9	1.0	0.9	0.9	0.9	1.0	1.0
Observations	54,714	80,419	82,023	44,761	51,489	141,028	43,582	63,392	35,658

Note: Regressions include baseline measurement of outcome, fixed effects for cash treatment randomization strata, and country-wave fixed effects. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values (calculated from p-values based on standard errors clustered at village level) in brackets.

3.3.4 Descriptive Evidence on Pathways

In this subsection, we investigate whether the practice known as "satisficing" is likely an explanation behind the observed pattern of results. Satisficing is a term used to describe the phenomenon where respondents may be answering questions in such a way that helps them avoid or shorten follow-ups, and therefore, reduce survey length (see Roberts et al. 2019 for a review of the evidence about this behavior). In this case, satisficing would entail responding "no" to questions, or list fewer number of items such as transfers, in order to avoid follow-up questions on those items. Satisficing requires that respondents learn that answering "no" to a question reduces the number of follow-up questions, and so can only be present if respondents learn this pattern over the course of the survey, or if fatigue makes

people more likely to satisfice. Empirically, if respondents already suspect that answering no to particular questions will lessen the number of follow-up questions and behave strategically from the start, then satisficing will not be detectable even though it is present.

While our study was not set up to answer this question, we produce two pieces of descriptive evidence. First, as mentioned in subsection 3.2.1, we randomly selected 20% of our sample to participate in phone surveys, which contained a subset of questions from the in-person surveys and began shortly after the baseline survey. Respondents were called once every 2 months for about 16-26 months (or 8-13 rounds). After deciding to implement the survey-order randomization into the longer in-person surveys, we later decided to also randomize the order in the phone surveys. Importantly, the randomization began around the 8th round of the survey in Liberia and the 11th in Malawi, so respondents already had lots of experience with the questionnaire.[10] If satisficing is an explanation, we would expect survey fatigue to be minimal in this experiment (since based on prior experience, people would be equally able to skip questions wherever they appeared in the survey).[11] The randomization was very similar to the longer surveys, though less involved: specifically, as shown in Figure 3.A3, we varied the location of the expenditure and transfers sections within the survey.

Results are shown in Table 3.8.[12] Columns 1-5 analyze responses to open-ended questions, and Columns 6-9 show outcomes for questions that follow the fixed list pattern. To study these, we run the same regressions as in (3.1) and (3.2)

[10] See Figure 3.A1 for the specific survey rounds when order randomization was implemented.

[11] Another implication of survey fatigue is that the total survey time, and thus the value of categories, should decline over time as respondents learn the skip codes. However, we have no way of testing this since the number of rounds is colinear with time trends.

[12] In Table 3.A4, we show the impacts on the value of aggregated categories, a replication of the analysis that we show in Table 3.5.

respectively, except that the outcomes are now drawn from the phone survey.

Table 3.8: Survey Fatigue in Phone Surveys

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Number of distinct items reported for the following:					=1 if item is selected (not skipped):			
	ROSCAs	VSLAs	Transfers received	Transfers given	Credit purchases	Savings	Loans	Food expend	Non-durables
Hours into Survey	0.050	0.308	0.091	-0.246	-0.346*	0.048	-0.014	-0.103***	-0.069*
	[0.315]	[0.108]	[0.308]	[0.105]	[0.091]	[0.108]	[0.185]	[0.001]	[0.091]
Dependent variable: Mean	0.088	0.372	0.205	0.190	0.283	0.140	0.031	0.216	0.351
Hours into Survey: Mean	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Hours into Survey: SD	0.1	0.1	0.1	0.2	0.2	0.1	0.1	0.1	0.1
Number of respondents	780	780	780	780	780	780	780	780	780
Observations	1,762	1,762	1,762	1,762	1,762	18,678	24,654	63,059	20,083

Note: For columns 1-5, there is 1 observations per respondent per survey. For most individuals, 2-3 rounds of phone surveys are included in this table. For columns 6-9, each column represents a separate set of questions and each set includes multiple items (e.g., food expenditure includes 35 types of food). All regressions include a survey fixed effect (i.e., country and Wave 1 and 2 in Liberia). Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Contrary to the predictions of a satisficing hypothesis, we find evidence of negative effects of survey duration on both the counts and the value of objects/outcomes reported by the respondents. This is similar qualitatively to the results in the main survey, for which respondents had much less experience. In fact, we find that our observed fatigue effects over the phone are similar in magnitude to those documented in Abay et al. (2021), who find that a 15 minute delay in the timing of the food consumption module leads to an 8-17% decline in the household dietary diversity score (similar to the effect sizes we document). Moreover, for the items that we measure in-person as well as over the phone, we find that the fatigue effects are in fact, much stronger over the phone than they are in person. For example, staying with the example of food expenditure, in Table 3.4, we document a fatigue effect of about 10% for an hour delay during an in-person survey, but this effect is of the order of 50% over the phone. This is in line with the evidence laid out in Abay et al. (2021), who show that survey fatigue comes about much sooner over the phone relative to in person surveys.

Second, as suggested by an anonymous referee, we note that Table 3.5 and Appendix F show survey fatigue effects on different categories of items. In earlier work such as Ambler et al. (2021) and Abay et al. (2021), researchers have found larger effects on less memorable items and smaller effects on more memorable ones. However, we find evidence consistent with a nearly across-the-board negative effect of fatigue, rather than differential effect based on salience. While not definitive, this result muddies the picture, since it is more consistent with satisficing than with cognitive burden.

Ultimately then, we do not have a definitive piece of evidence on pathways. Instead we conclude that both effects may be at play, and we leave a fuller investigation to future work.

3.4 Conclusion

In this paper, we randomize the order of questions asked as part of the baseline and endline surveys of a cash transfer experiment to provide evidence on the impact of survey duration on the quality of responses elicited during the survey. Our results point to strong fatigue effects, on the order of a 10-64% reduction in the count of reported items, which leads to even bigger effects on the reported monetary values of categories that aggregate over these items.

Is there a way for these findings to inform survey design? Survey fatigue is not a recent discovery, and practitioners suggest a variety of remedies to address this concern, most of which boil down to fielding shorter surveys, or splitting surveys into multiple shorter versions. For example, Aggarwal et al. (2021b) is an example of a multi-day baseline survey. Other strategies involve sacrificing the scope of data collection, for example by splitting the survey into shorter versions,

administering only one of the versions to each respondent, and imputing responses to the unasked questions (Herzog and Bachman 1981; Raghunathan and Grizzle 1995). Another strategy is to replace ordinal questions with binary ones (Dolnicar et al. 2011). However, each of these remedies comes with its own set of problems, either in terms of detail and measurement error, or in terms of cost.

While we have no easy fixes to recommend, an obvious remedial step would be to place the most important questions (for example, those about the primary outcomes in an RCT), as early as possible in the survey. Relatedly, it may also be good survey practice for enumerators to suggest taking a short break before they start asking important questions that are placed later in the survey. This may be an important consideration especially for interventions in which the primary outcome is sensitive (for example, intimate partner violence, which was placed at the end of these surveys for exactly this reason).[13] Researchers often choose to place such sensitive questions later in the survey to allow respondents some time to become familiar with the enumerator and with the survey, but this paper suggests that this consideration should be balanced against the risk of survey fatigue.

In general, it may make sense for enumerators to be trained to pay more attention to signs of fatigue and disengagement, and for survey protocols to have a set of remedial actions to take in such a scenario, like taking a break or playing a short game. Future research should identify such remedial actions.

Another implication from this paper is that, for those working with secondary data collected via long surveys, such as the LSMS or the DHS surveys, it may be useful to recognize that cross-country comparisons or even within country comparisons across survey waves may be complicated because of varying survey

[13]See Park et al. (2022b) and Park and Kumar (2022) for related work on the pitfalls of measuring IPV in this and a related sample in Liberia.

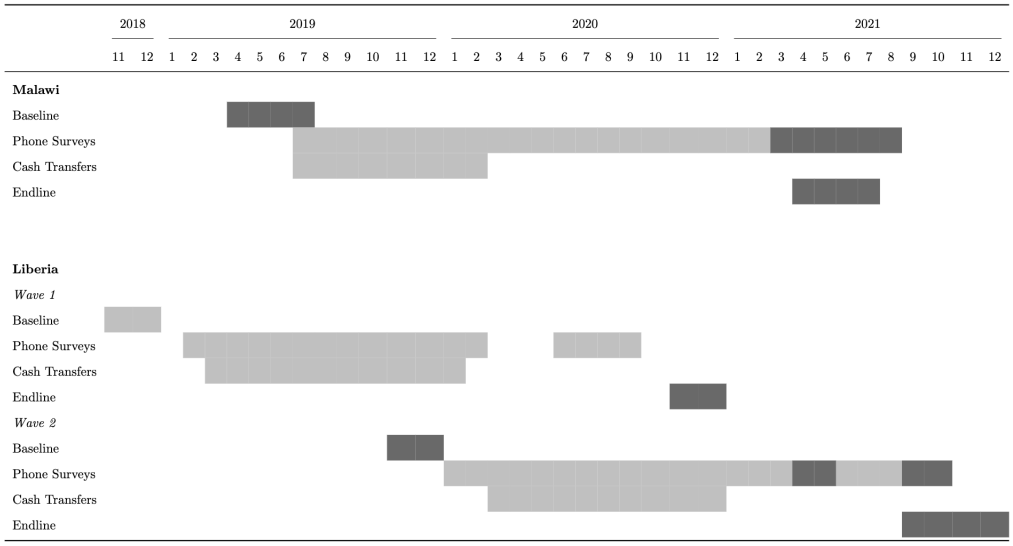
duration. It may be important to design panel surveys such that outcomes are measured at similar points in the survey over waves.

Finally, we note that in addition to the cognitive decline faced by respondents, enumerators are also human participants in a survey and may be constrained by mental bandwidth in the administration of long surveys. In this paper, we have no way of disentangling the effects of fatigue on enumerators from those on respondents as both start and end the survey together. However, measuring these effects separately as well as identifying remedies should be a focus of future research.

3.5 Appendix

Appendix A: Main Appendix Figures and Tables

Figure 3.A1: Timeline of Survey Activities



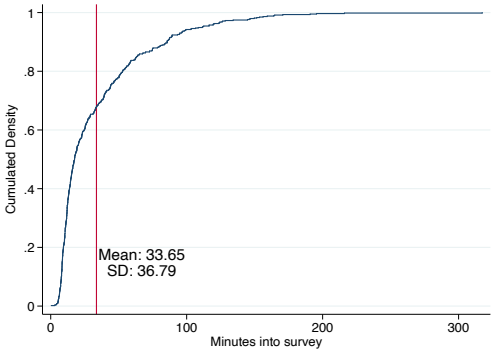
Note: Darker grey blocks indicate the survey rounds where module order randomization was conducted and thus data for which are included for analysis in this paper.

Note: Darker grey blocks indicate the survey rounds where module order randomization was conducted and thus data for which are included for analysis in this paper.

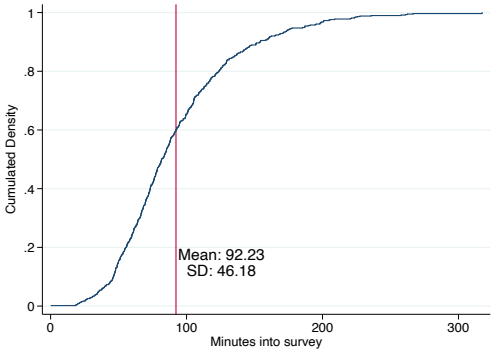
Figure 3.A2: Distribution of Survey Time

Distribution of time to reach the question where on average the survey is:

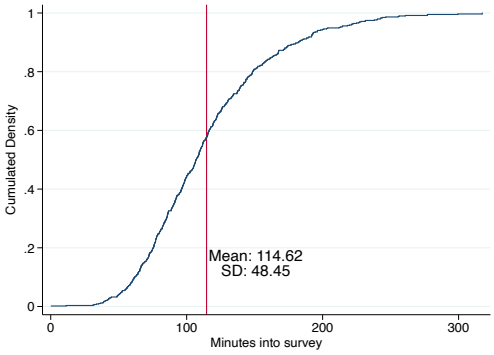
(a) 10% completed



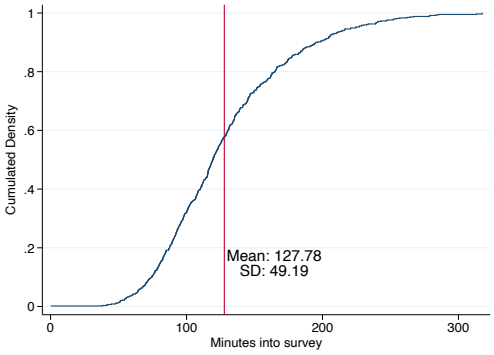
(b) 25% completed



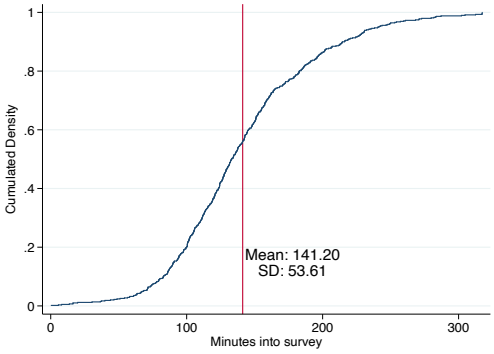
(c) Median



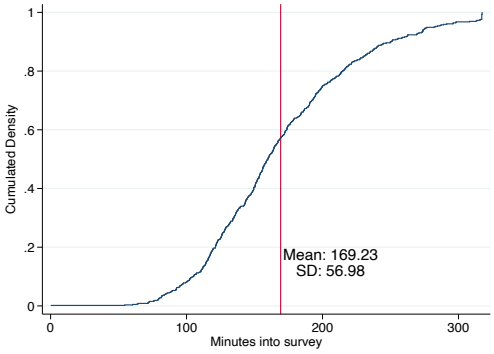
(d) 75th percentile



(e) 90th percentile

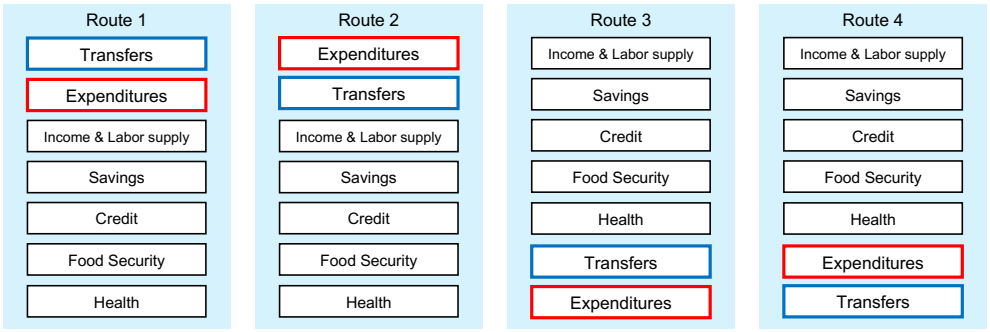


(f) Total survey length



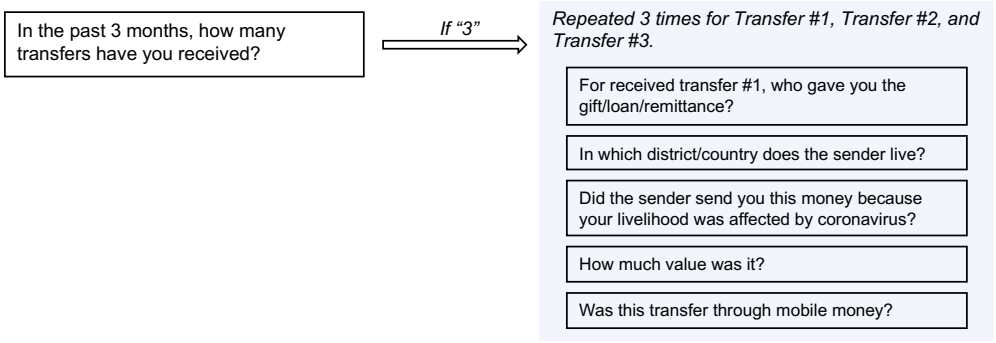
Note: Based on Version A only.

Figure 3.A3: Randomized Order of Modules in Phone Surveys



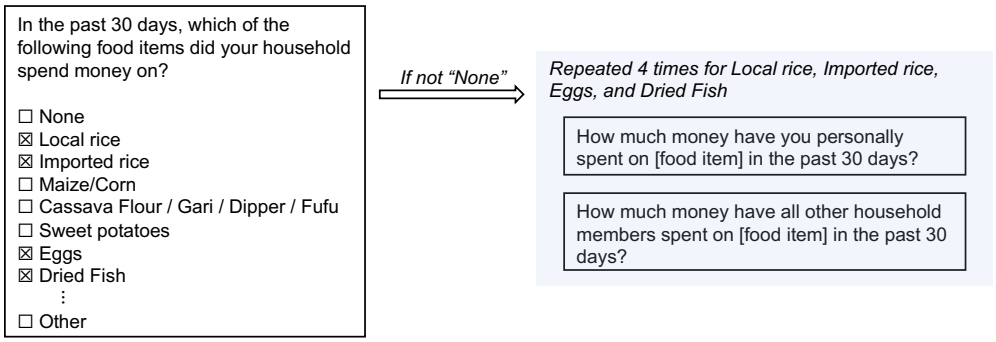
Note: A respondent is randomly provided with one among Routes 1-4.

Figure 3.A4: Example of "Open-Ended" Question Order



H

Figure 3.A5: Example of "Fixed List" Question



H

Table 3.A1: Average Duration by Survey Versions (in hours)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Survey Version						Overall
	A	B	C	D	E	F	
Panel A: Liberia							
Baseline	2.28	2.27	2.24	2.31	2.29	2.24	2.27
	(0.69)	(0.65)	(0.69)	(0.75)	(0.67)	(0.70)	(0.69)
Endline	2.73	2.64	2.74	2.68	2.72	2.77	2.71
	(1.04)	(1.05)	(1.12)	(1.02)	(1.09)	(1.16)	(1.08)
Panel B: Malawi							
Baseline	3.15	3.03	3.06	3.03	3.01	3.04	3.05
	(1.02)	(0.89)	(0.93)	(0.92)	(0.91)	(0.90)	(0.93)
Endline	2.75	2.81	2.80	2.76	2.75	2.78	2.77
	(0.80)	(0.82)	(0.81)	(0.79)	(0.82)	(0.82)	(0.81)

Note: Standard deviations in parentheses.

Table 3.A2: Experimental variation in time before sections were administered (phone surveys)

	(1)	(2)	(3)	(4)
	Time into survey (minutes) at the beginning of following section:			
	Savings	Credit	Transfers	Expenditure
Version B	-0.17	-0.08	8.66***	-1.45***
	(0.32)	(0.34)	(0.34)	(0.29)
Version C	-9.14***	-9.03***	10.48***	9.95***
	(0.31)	(0.33)	(0.34)	(0.28)
Version D	-9.66***	-9.59***	17.52***	8.53***
	(0.31)	(0.32)	(0.33)	(0.28)
Version A: Mean	15.47	16.53	3.21	4.81
Version A: SD	6.89	6.98	3.10	3.70
F-statistic: joint significance	585.88	523.70	941.79	837.01
Number of respondents	780	780	779	780
Observations	1,762	1,762	1,759	1,760

Note: Observations include only phone survey data. Regressions include country-sample fixed effects. ***, **, and * represent significance at 1%, 5%, and 10%, respectively.

Table 3.A3: Effect of survey time on measured treatment effects on monetary value of aggregated categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Expenditure		Assets					Transfers	
	Food	Nondurables	Livestock	Farm tools	Durables	Savings	Loans	Given	Received
Time into Survey (hr)	-0.55	-0.39	6.94	-2.47	-7.09	-7.71	-3.98	0.14	-1.94
	[1.000]	[1.000]	[1.000]	[0.999]	[1.000]	[1.000]	[0.999]	[1.000]	[1.000]
Cash × Time into Survey (hr)	-0.21	1.12	-45.35	0.30	36.20	-5.41	7.99	-2.70	-2.15
	[1.000]	[1.000]	[1.000]	[1.000]	[0.962]	[1.000]	[0.286]	[0.286]	[1.000]
Cash	0.19	0.27	26.00**	1.47***	21.02***	4.56***	-0.19	0.26	1.33**
	[0.197]	[0.227]	[0.034]	[0.004]	[0.001]	[0.007]	[0.545]	[0.197]	[0.019]
Control Mean	3.08	6.30	90.00	9.75	56.32	8.68	6.94	1.66	6.85
Control SD	4.90	9.37	367.73	10.66	138.21	55.59	19.14	6.58	14.53
Hours into Survey: Mean	1.9	2.0	1.7	1.7	1.8	1.9	1.9	1.9	1.9
Hours into Survey: SD	0.9	0.9	0.8	0.8	0.8	0.9	0.8	0.9	0.9
Observations	3,962	3,962	3,962	3,962	3,962	3,962	3,687	3,962	3,962

Note: Regressions include baseline measurement of outcome, fixed effects for cash treatment randomization strata, and country fixed effects. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values (calculated from p-values based on standard errors clustered at village level) in brackets.

Table 3.A4: Effect of survey time of total value of aggregated categories, phone surveys

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total value of reported items for the following:						
	Savings	Loans	Food Expend	Non-durables	Transfers received	Transfers given	Credit purchases
Hours into Survey	8.56	-2.45	-9.63**	-2.62	2.68*	-0.73	-2.38*
	[0.200]	[0.399]	[0.036]	[0.200]	[0.067]	[0.173]	[0.061]
Dependent variable: Mean	10.19	8.63	13.37	7.49	1.55	0.59	1.28
Hours into Survey: Mean	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Hours into Survey: SD	0.1	0.1	0.1	0.1	0.1	0.2	0.2
Number of respondents	780	780	780	780	780	780	780
Observations	1,762	1,762	1,762	1,762	1,762	1,762	1,762

Note: Observations at respondent level, and regressions include sample fixed effects (i.e., country and Wave 1 and 2 in Liberia). Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Appendix B: Heterogeneity by Country

Table 3.B1: Heterogeneity by Country in Open Ended Questions Question

	(1)	(2)	(3)	(4)	(5)
	Number of distinct items reported for the following:				
	ROSCAs	VSLAs	Transfers received	Transfers given	Credit purchases
Panel A. Liberia					
Hours into Survey	-0.001	-0.036	-0.037	-0.157**	-0.180
	[0.786]	[0.127]	[0.331]	[0.019]	[0.389]
Dependent variable: Mean	0.106	0.063	0.297	0.381	0.349
Hours into Survey: Mean	1.7	1.7	1.8	1.8	1.8
Hours into Survey: SD	1.2	1.2	1.2	1.2	1.2
Number of respondents	2,652	2,653	2,652	2,650	2,653
Observations	4,500	4,500	4,498	4,494	4,501
Panel B. Malawi					
Hours into Survey	0.004	-0.077	-0.122	-0.240**	-0.016
	[0.786]	[0.127]	[0.331]	[0.019]	[0.389]
Dependent variable: Mean	0.017	0.316	0.258	0.285	0.380
Hours into Survey: Mean	2.0	2.0	2.1	2.1	2.1
Hours into Survey: SD	0.8	0.8	0.8	0.8	0.8
Number of respondents	2,944	2,944	2,944	2,944	2,944
Observations	5,725	5,724	5,725	5,721	5,727

Note: Observations at respondent level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include country-sample fixed effects. ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.B2: Heterogeneity by Country in Fixed List Questions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	=1 if item is selected (not skipped):								
	Livestock	Farm tools	Durable	Savings	Loans	Food expend	Non-durables	Shocks	Public goods
Panel A. Liberia									
Hours into Survey	-0.012	0.002	0.001	-0.002	0.001	-0.033***	-0.053**	-0.011	0.006
	[0.405]	[0.411]	[0.741]	[0.731]	[0.661]	[0.005]	[0.019]	[0.136]	[0.385]
Dependent variable: Mean	0.097	0.196	0.171	0.048	0.009	0.189	0.234	0.065	0.075
Hours into Survey: Mean	1.5	1.6	1.6	1.7	1.7	1.7	1.7	1.8	1.8
Hours into Survey: SD	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.5
Number of respondents	2,653	2,653	2,653	2,653	2,653	2,653	2,653	2,653	2,566
Observations	49,511	94,521	90,020	54,012	62,397	166,537	49,511	72,016	23,094
Panel B. Malawi									
Hours into Survey	-0.002	-0.011	-0.001	-0.002	-0.001	-0.016***	-0.025**	-0.028	-0.014
	[0.405]	[0.411]	[0.741]	[0.731]	[0.661]	[0.005]	[0.019]	[0.136]	[0.385]
Dependent variable: Mean	0.057	0.119	0.179	0.071	0.028	0.214	0.261	0.180	0.028
Hours into Survey: Mean	1.9	1.9	1.9	2.0	2.0	2.1	2.1	2.2	2.2
Hours into Survey: SD	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.9
Number of respondents	2,941	2,941	2,941	2,944	2,944	2,944	2,944	2,944	2,783
Observations	85,320	113,760	122,353	60,033	76,314	200,410	62,986	94,508	25,047

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include country-sample fixed effects and question-item level fixed effects. ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.B3: Heterogeneity by country on total monetary values of aggregated categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Total value of reported items for the following:										
	Livestock	Farm Tools	Durables	Savings	Loans	Food Expend	Non-durables	Public goods	Transfers received	Transfers given	Credit purchases
Panel A. Liberia											
Hours into Survey	-22.28	0.14	-3.04	-0.47	1.06	-6.05**	-4.10*	-0.15	-0.59*	-0.87**	-1.25
	[0.520]	[0.359]	[0.396]	[0.426]	[0.517]	[0.028]	[0.079]	[0.386]	[0.088]	[0.014]	[0.311]
Dependent variable: Mean	155.11	11.12	53.59	27.39	4.62	21.42	10.59	0.28	1.50	1.23	1.39
Hours into Survey: Mean	1.5	1.5	1.6	1.7	1.8	1.7	1.8	1.9	1.8	1.8	1.8
Hours into Survey: SD	1.3	1.6	1.3	1.2	1.2	1.8	1.8	2.2	1.2	1.2	1.2
Number of respondents	2,653	2,566	2,653	2,653	2,653	2,653	2,653	2,566	2,653	2,653	2,653
Observations	4,501	2,566	4,501	4,501	4,501	4,501	4,501	2,566	4,501	4,501	4,501
Panel B. Malawi											
Hours into Survey	-4.48	-3.86	15.58	-1.51	0.22	-2.41**	-1.08*	-0.01	-0.39*	-0.26**	-0.02
	[0.520]	[0.359]	[0.396]	[0.426]	[0.517]	[0.028]	[0.079]	[0.386]	[0.088]	[0.014]	[0.311]
Dependent variable: Mean	48.83	9.89	61.68	6.18	7.87	12.13	5.83	0.02	0.52	0.26	0.36
Hours into Survey: Mean	1.9	1.8	1.9	2.0	2.0	2.1	2.1	2.3	2.1	2.1	2.1
Hours into Survey: SD	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.9	0.8	0.8	0.8
Number of respondents	2,941	2,783	2,941	2,944	2,944	2,944	2,944	2,783	2,944	2,944	2,944
Observations	5,688	2,783	5,688	5,725	5,451	5,726	5,726	2,783	5,727	5,727	5,727

Note: Observations at respondent level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include country-sample fixed effects. ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Appendix C: Heterogeneity by Survey type

Table 3.C1: Heterogeneity by Survey (Baseline or Endline) in Open-Ended Questions

	(1)	(2)	(3)	(4)	(5)
	Number of distinct items reported for the following:				
	ROSCAs	VSLAs	Transfers received	Transfers given	Credit purchases
Panel A. Baseline surveys					
Hours into Survey	0.020	-0.076	-0.085	-0.267***	-0.043
	[0.348]	[0.201]	[0.212]	[0.010]	[0.207]
Dependent variable: Mean	0.067	0.204	0.382	0.494	0.414
Hours into Survey: Mean	2.0	2.0	2.0	2.0	2.0
Hours into Survey: SD	0.8	0.8	0.8	0.8	0.8
Number of respondents	4,877	4,875	4,874	4,870	4,879
Observations	4,877	4,875	4,874	4,870	4,879
Panel B. Endline surveys					
Hours into Survey	-0.010	-0.029	-0.061	-0.139***	-0.159
	[0.348]	[0.201]	[0.212]	[0.010]	[0.207]
Dependent variable: Mean	0.046	0.205	0.178	0.176	0.323
Hours into Survey: Mean	1.8	1.8	1.8	1.9	1.9
Hours into Survey: SD	1.2	1.2	1.2	1.2	1.2
Number of respondents	5,348	5,349	5,349	5,345	5,349
Observations	5,348	5,349	5,349	5,345	5,349

Note: Observations at respondent level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include country-sample fixed effects. ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.C2: Heterogeneity by Survey (Baseline or Endline) in Fixed List Questions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	=1 if item is selected (not skipped):								
	Livestock	Farm tools	Durable	Savings	Loans	Food expend	Non-durables	Shocks	Public goods
Panel A. Baseline surveys									
Hours into Survey	-0.008	-0.003	0.008	0.005	0.002	-0.032**	-0.044**	-0.026	
	[0.243]	[0.355]	[0.172]	[0.257]	[0.355]	[0.011]	[0.018]	[0.106]	
Dependent variable: Mean	0.076	0.166	0.191	0.069	0.022	0.227	0.289	0.194	
Hours into Survey: Mean	1.9	1.8	1.9	2.0	2.0	1.9	2.0	2.0	
Hours into Survey: SD	0.7	0.7	0.7	0.7	0.8	0.8	0.7	0.8	
Number of respondents	4,840	4,840	4,840	4,877	4,877	4,878	4,878	4,875	
Observations	64,860	98,735	102,610	52,640	67,977	174,600	53,658	80,940	
Panel B. Endline surveys									
Hours into Survey	-0.005	-0.005	-0.012	-0.009	-0.002	-0.017**	-0.032**	-0.017	-0.002
	[0.243]	[0.355]	[0.172]	[0.257]	[0.355]	[0.011]	[0.018]	[0.106]	[0.404]
Dependent variable: Mean	0.068	0.144	0.162	0.053	0.018	0.180	0.212	0.070	0.050
Hours into Survey: Mean	1.6	1.6	1.7	1.8	1.8	1.8	1.9	2.0	2.0
Hours into Survey: SD	1.1	1.2	1.1	1.2	1.2	1.2	1.2	1.2	1.2
Number of respondents	5,349	5,349	5,349	5,349	5,073	5,349	5,349	5,349	5,349
Observations	69,971	109,546	109,763	61,405	70,734	192,347	58,839	85,584	48,141

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include country-sample fixed effects and question-item level fixed effects. ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.C3: Heterogeneity by survey type (baseline or endline) on total monetary values of aggregated categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Total value of reported items for the following:										
	Livestock	Farm Tools	Durables	Savings	Loans	Food Expend	Non-durables	Public goods	Transfers received	Transfers given	Credit purchases
Panel A. Baseline surveys											
Hours into Survey	-15.61		6.40	-4.45	1.78	-4.65**	-2.50**		-0.36	-0.44**	-0.27*
	[0.212]		[0.287]	[0.295]	[0.229]	[0.046]	[0.050]		[0.107]	[0.041]	[0.063]
Dependent variable: Mean	51.24		46.06	18.16	6.60	16.93	8.31		0.83	0.71	0.54
Hours into Survey: Mean	1.8		1.8	2.0	2.0	1.9	2.0		2.0	2.0	2.0
Hours into Survey: SD	0.7		0.7	0.8	0.8	0.8	0.7		0.8	0.8	0.8
Number of respondents	4,840		4,840	4,877	4,878	4,878	4,878		4,879	4,879	4,879
Observations	4,840		4,840	4,877	4,878	4,878	4,878		4,879	4,879	4,879
Panel B. Endline surveys											
Hours into Survey	-11.42	-1.32	3.29	0.33	-0.53	-3.06**	-2.06**	-0.10	-0.59	-0.69**	-1.06*
	[0.212]	[0.140]	[0.287]	[0.295]	[0.229]	[0.046]	[0.050]	[0.127]	[0.107]	[0.041]	[0.063]
Dependent variable: Mean	136.08	10.48	69.01	13.11	6.21	15.57	7.57	0.14	1.05	0.67	1.06
Hours into Survey: Mean	1.6	1.6	1.7	1.8	1.8	1.9	1.9	2.1	1.8	1.9	1.9
Hours into Survey: SD	1.3	1.3	1.3	1.2	1.2	1.7	1.7	1.6	1.2	1.2	1.2
Number of respondents	5,349	5,349	5,349	5,349	5,074	5,349	5,349	5,349	5,349	5,349	5,349
Observations	5,349	5,349	5,349	5,349	5,074	5,349	5,349	5,349	5,349	5,349	5,349

Note: Observations at respondent level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include country-sample fixed effects. ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Appendix D: Robustness to Household Fixed Effects

Table 3.D1: Survey Time and Probability of Reporting an Item in Phone Surveys (with Household FE)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Number of distinct items reported for the following:					=1 if item is selected (not skipped):			
	ROSCAs	VSLAs	Transfers received	Transfers given	Credit purchases	Savings	Loans	Food expend	Non-durables
Hours into Survey	0.081	-0.047	-0.145	-0.062	-0.199	0.012	-0.009	-0.090***	-0.124
	[1.000]	[1.000]	[1.000]	[1.000]	[0.888]	[1.000]	[1.000]	[0.010]	[0.160]
Dependent variable: Mean	0.074	0.394	0.200	0.175	0.255	0.140	0.031	0.216	0.351
Hours into Survey: Mean	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Hours into Survey: SD	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Number of respondents	610	610	610	610	610	780	780	780	780
Observations	1,592	1,592	1,592	1,592	1,592	18,678	24,654	63,059	20,083

Note: For columns 1-4, observations at respondent-question-item level, and regressions include country-sample fixed effects, question-item level fixed effects and household level fixed effects. Regressions drop singleton observations (there are 170 of these). For columns 5-9, observations at respondent level, and regressions include country-sample fixed effects. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.D2: Survey time and the number of items reported in in-person surveys (with Household FE)

	(1)	(2)	(3)	(4)	(5)
	Number of distinct items reported for the following:				
	ROSCAs	VSLAs	Transfers received	Transfers given	Credit purchases
Hours into Survey	-0.013	-0.087**	-0.149**	-0.327***	-0.060
	[0.757]	[0.032]	[0.032]	[0.001]	[0.757]
Dependent variable: Mean	0.055	0.220	0.273	0.330	0.372
Hours into Survey: Mean	1.9	1.9	1.9	2.0	2.0
Hours into Survey: SD	1.0	1.0	1.0	1.0	1.0
Number of respondents	4,629	4,627	4,627	4,621	4,631
Observations	9,258	9,254	9,254	9,242	9,262

Note: Observations at respondent level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include country-sample fixed effects and household level fixed effects. Regressions drop singleton observations (there are 966 of these). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.D3: Survey time and the probability of reporting an item in in-person surveys (with Household FE)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	=1 if item is selected (not skipped):							
	Livestock	Farm tools	Durable	Savings	Loans	Food expend	Non-durables	Shocks
Hours into Survey	-0.005	-0.003	-0.001	-0.002	0.002	-0.020***	-0.028**	-0.021***
	[0.710]	[0.822]	[1.000]	[0.847]	[0.757]	[0.001]	[0.023]	[0.001]
Dependent variable: Mean	0.072	0.154	0.176	0.060	0.020	0.203	0.249	0.130
Hours into Survey: Mean	1.7	1.7	1.8	1.9	1.9	1.9	1.9	2.0
Hours into Survey: SD	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Number of respondents	5,594	5,594	5,594	5,597	5,597	5,597	5,597	5,597
Observations	134,831	208,281	212,373	114,045	138,711	366,947	112,497	166,524

Note: Observations at respondent-question-item level. Each regression is an IV regression, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). Regressions include country-sample fixed effects, question-item level fixed effects and household level fixed effects. ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Appendix E: Non-linearities in the Relationship Between Survey Time and the Probability of Skipping

Figure 3.E1: Probability of selection against the predicted time to reach the question

In-person Surveys:

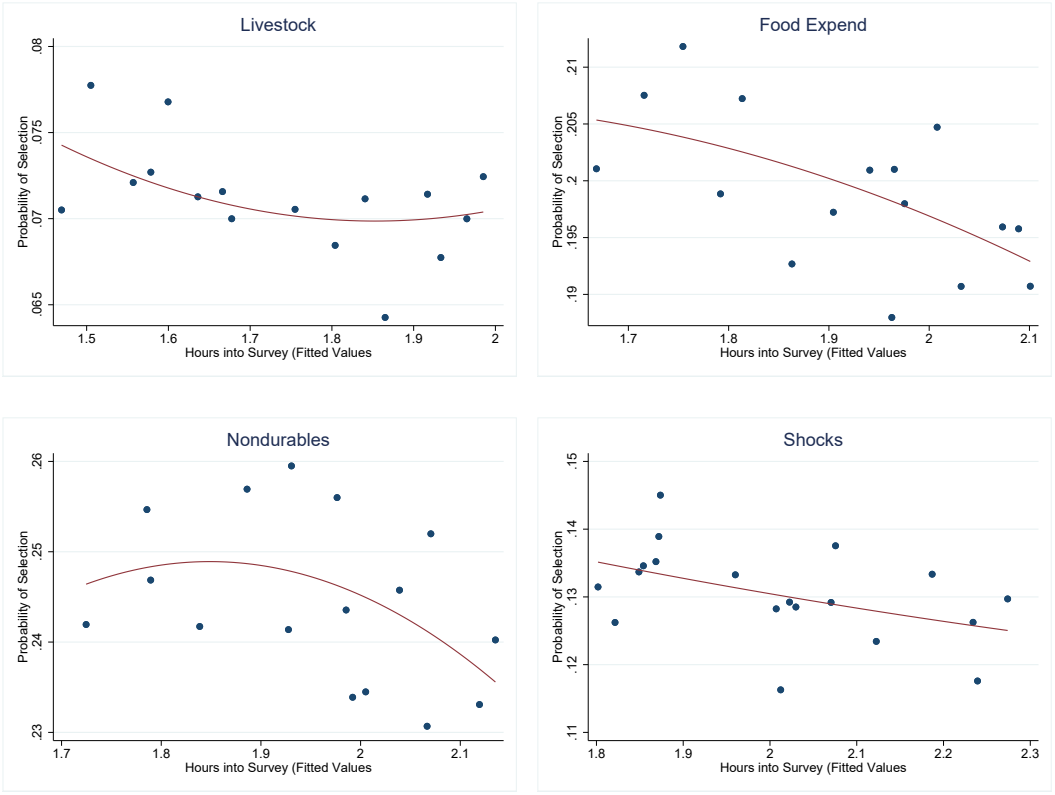
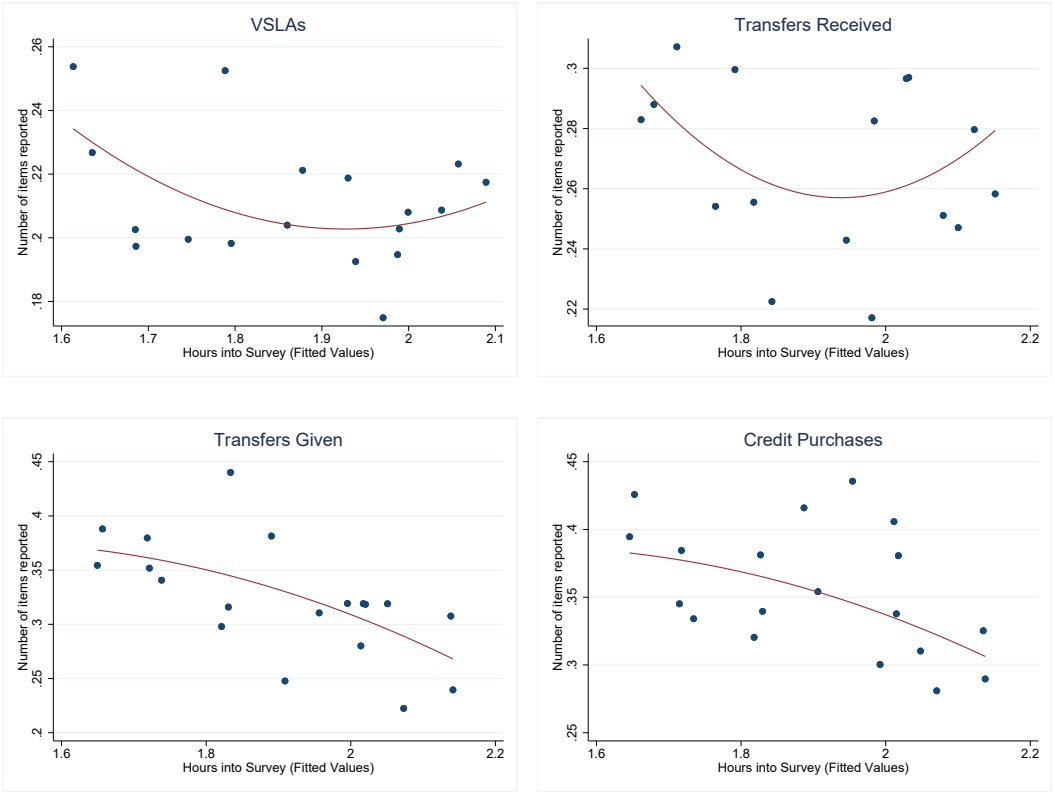


Figure 3.E2: Number of items reported against the predicted time to reach the question

In-person Surveys:



Appendix F: Analysis of Fatigue on Disaggregated Categories

Table 3.F1: Livestock

	(1)	(2)	(3)	(4)	(5)
	=1 if item is selected (not skipped):				
	Goat	Pig	Chicken	Dog	Goat (local)
Hours into Survey	-0.025	-0.002	-0.040	-0.007	-0.001
	(0.263)	(0.900)	(0.114)	(0.752)	(0.974)
	[0.541]	[0.895]	[0.396]	[0.787]	[0.895]
Dependent variable: Mean	0.121	0.060	0.534	0.126	0.214
Hours into Survey: Mean	1.5	1.5	1.7	1.5	1.9
Hours into Survey: SD	1.2	1.2	1.0	1.2	0.8
Number of respondents	2,653	2,653	5,594	2,653	2,941
Observations	4,501	4,501	10,189	4,501	5,688

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.F2: Farm Tools

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	=1 if item is selected (not skipped):									
	Hand hoes	Cut-lass	Sho-vels	Dig-gers	Axes	Fill-ing Tools	Cans/ Buckets	Ping-alays	Knives	Hooks
Hours into Survey	-0.046	-0.006	0.014	0.015	-0.040	-0.011	-0.042	0.033	-0.031	0.034
	(0.011)	(0.777)	(0.452)	(0.201)	(0.086)	(0.504)	(0.072)	(0.118)	(0.185)	(0.190)
	[0.124]	[0.787]	[0.697]	[0.508]	[0.367]	[0.697]	[0.367]	[0.396]	[0.504]	[0.504]
Dependent variable: Mean	0.855	0.575	0.165	0.055	0.283	0.118	0.355	0.107	0.450	0.174
Hours into Survey: Mean	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.6	1.7	1.6
Hours into Survey: SD	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2	1.0	1.2
Number of respondents	5,594	5,594	5,594	5,594	5,594	5,594	5,594	2,653	5,594	2,653
Observations	10,189	10,189	10,189	10,189	10,189	10,189	10,189	4,501	10,189	4,501

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.F3: Saving Places

	(1)	(2)	(3)	(4)
	=1 if item is selected (not skipped):			
	Saving group	Cash home	VSLA	Live-stock
Hours into Survey	-0.025	0.002	-0.046	0.015
	(0.366)	(0.918)	(0.021)	(0.490)
	[0.559]	[0.895]	[0.148]	[0.697]
Dependent variable: Mean	0.193	0.193	0.175	0.078
Hours into Survey: Mean	1.7	1.9	1.9	2.0
Hours into Survey: SD	1.2	1.0	1.0	0.8
Number of respondents	2,653	5,597	5,597	2,944
Observations	4,501	10,226	10,226	5,725

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.F4: Loan Sources

	(1)	(2)
	=1 if item is selected (not skipped):	
	Neighbors or friends	VSLA
Hours into Survey	-0.002	-0.009
	(0.901)	(0.571)
	[0.895]	[0.743]
Dependent variable: Mean	0.079	0.098
Hours into Survey: Mean	1.9	1.9
Hours into Survey: SD	1.0	1.0
Number of respondents	5,597	5,470
Observations	9,950	9,361

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.F5: Food Expenditures (Part-I)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	=1 if item is selected (not skipped):												
	Local rice	Imported rice	Maize/ Corn	Cassava roots	Cassava Flour	Sweet potatoes	Irish potatoes	Dried Beans	Groundnut	Palm nuts	Palm oil	Tomatoes	Onions
Hours into Survey	-0.001	-0.074	0.012	0.007	-0.016	0.027	-0.020	-0.017	-0.031	-0.042	-0.043	-0.020	-0.057
	(0.946)	(0.011)	(0.604)	(0.696)	(0.313)	(0.270)	(0.135)	(0.459)	(0.171)	(0.107)	(0.352)	(0.339)	(0.054)
	[0.895]	[0.124]	[0.754]	[0.787]	[0.553]	[0.543]	[0.396]	[0.697]	[0.487]	[0.390]	[0.559]	[0.559]	[0.278]
Dependent variable: Mean	0.069	0.451	0.192	0.096	0.091	0.294	0.060	0.192	0.187	0.086	0.520	0.525	0.524
Hours into Survey: Mean	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.7	1.7	1.9	1.9
Hours into Survey: SD	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2	1.2	1.0	1.0
Number of respondents	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	2,653	2,653	5,597	5,597
Observations	10,227	10,227	10,227	10,227	10,227	10,227	10,227	10,227	10,227	4,501	4,501	10,227	10,227

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.F6: Food Expenditures(Part-II)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	=1 if item is selected (not skipped):												
	Okra	Bananas	Oranges	Eggs	Goat meat	Chicken	Dried Fish	Fresh Fish	Salt	Sugar	Breads	Other Veg.	Vita/ Maggi
Hours into Survey	-0.040	-0.022	-0.007	-0.031	-0.019	-0.021	-0.029	-0.029	-0.026	-0.061	-0.062	-0.034	-0.002
	(0.076)	(0.245)	(0.657)	(0.170)	(0.219)	(0.360)	(0.309)	(0.223)	(0.130)	(0.029)	(0.007)	(0.162)	(0.949)
	[0.367]	[0.508]	[0.787]	[0.487]	[0.508]	[0.559]	[0.553]	[0.508]	[0.396]	[0.167]	[0.124]	[0.487]	[0.895]
Dependent variable: Mean	0.179	0.121	0.082	0.198	0.076	0.261	0.635	0.335	0.904	0.362	0.186	0.261	0.868
Hours into Survey: Mean	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.7
Hours into Survey: SD	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.2
Number of respondents	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	2,653
Observations	10,227	10,227	10,227	10,227	10,227	10,227	10,227	10,227	10,227	10,227	10,227	10,227	4,501

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.F7: Non-Durables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	=1 if item is selected (not skipped):							
	Transport-ation	Air-time	Home supplies	Personal hygiene	Cleaning supplies for home	Kitchen supplies	Cosmetics	Barber
Hours into Survey	-0.123***	-0.020	-0.044	0.003	-0.011	-0.061	-0.066	-0.040
	(0.000)	(0.490)	(0.082)	(0.931)	(0.695)	(0.011)	(0.014)	(0.098)
	[0.003]	[0.697]	[0.367]	[0.895]	[0.787]	[0.124]	[0.128]	[0.367]
Dependent variable: Mean	0.473	0.407	0.234	0.620	0.300	0.187	0.270	0.214
Hours into Survey: Mean	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9
Hours into Survey: SD	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Number of respondents	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597
Observations	10,227	10,227	10,227	10,227	10,227	10,227	10,227	10,227

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Table 3.F8: Shocks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	=1 if item is selected (not skipped):												
	Flood	Drought	Land erosion	Food inflation	Loss of belongings	Lack of inputs for crops	Crop disease	Pesticide	Lack of inputs for livestock	Livestock disease	Low crop/ livetock prices	Severe illness in family	Death in household
Hours into Survey	-0.006	-0.049	-0.018	-0.064	-0.024	-0.058	-0.026	-0.059	-0.011	-0.021	0.000	-0.063*	-0.046
	(0.812)	(0.024)	(0.348)	(0.023)	(0.094)	(0.008)	(0.214)	(0.017)	(0.439)	(0.128)	(0.979)	(0.001)	(0.004)
	[0.826]	[0.148]	[0.559]	[0.148]	[0.367]	[0.124]	[0.508]	[0.134]	[0.697]	[0.396]	[0.895]	[0.052]	[0.124]
Dependent variable: Mean	0.308	0.164	0.125	0.387	0.063	0.181	0.163	0.252	0.071	0.061	0.061	0.125	0.083
Hours into Survey: Mean	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Hours into Survey: SD	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Number of respondents	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597	5,597
Observations	10,224	10,224	10,224	10,224	10,224	10,224	10,224	10,224	10,224	10,224	10,224	10,224	10,224

Note: Observations at respondent-question-item level. Reported are TOT estimates, where elapsed time into survey (in hours) is instrumented with the randomized module order (Versions A-F). ***, **, and * represent significance at 1%, 5%, and 10%, respectively, based on the false discovery rate (FDR) sharpened q-values in brackets.

Bibliography

Abay, Kibrom A., Guush Berhane, John Hoddinott, and Kibrom Tafere Hirfrfot (2021). "Assessing response fatigue in phone surveys: Experimental evidence on dietary diversity in Ethiopia". World Bank Policy Research Working Paper No. 9636.

Aggarwal, Shilpa, Jenny Aker, Dahyeon Jeong, Naresh Kumar, David S. Park, Jonathan Robinson, and Alan Spearot (2021a). "The Effect of Cash Transfers and Market Access on Households in Rural Liberia and Malawi". AEA RCT Registry.

Aggarwal, Shilpa, Jenny Aker, Dahyeon Jeong, Naresh Kumar, David Sungho Park, Jonathan Robinson, and Alan Spearot (2020a). "The Effect of Cash Transfers and Market Access on Households in Rural Liberia and Malawi". AEA RCT Registry.

— (2022). "The Dynamic Effects of Cash Transfers: Evidence from Rural Liberia and Malawi". Unpublished.

Aggarwal, Shilpa, Valentina Brailovskaya, and Jonathan Robinson (forthcoming). "Saving for Multiple Financial Needs: Evidence from Lockboxes and Mobile Money in Malawi". *Review of Economics and Statistics*.

Aggarwal, Shilpa, Rebecca Dizon-Ross, and Ariel Zucker (2021b). "Incentivizing Behavior Change: The Role of Time Preferences". National Bureau of Economic Research Working Paper No. 27079.

Aggarwal, Shilpa, Dahyeon Jeong, Naresh Kumar, David Sungho Park, Jonathan Robinson, and Alan Spearot (2020b). "Did COVID-19 market disruptions disrupt food security? Evidence from households in rural Liberia and Malawi". Working Paper w27932. National Bureau of Economic Research.

Aker, Jenny, Rachid Boumnijel, Amanda McClelland, and Niall Tierney (2016). "Payment Mechanisms and Anti-Poverty Programs: Evidence from a Mobile Money Cash Transfer Experiment in Niger". *Economic Development and Cultural Change* 65 (1): 1–37.

Ambler, Kate, Sylvan Herskowitz, and Mywish K. Maredia (2021). "Are we done yet? Response fatigue and rural livelihoods". *Journal of Development Economics* 153: 102736.

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

— (2008b). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". *Journal of the American Statistical Association* 103 (484): 1481–1495.

Angelucci, Manuela (2008). "Love on the Rocks: Domestic Violence and Alcohol Abuse in Rural Mexico". *The B.E. Journal of Economic Analysis & Policy* 8 (1).

Angelucci, Manuela and Giacomo De Giorgi (2009). "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?" *American Economic Review* 99 (1): 486–508.

Bacchus, L. J., Meghna Ranganathan, Charlotte Watts, and Karen Devries (2018). "Recent intimate partner violence against women and health: a systematic review and meta-analysis of cohort studies". *BMJ Open* 8 (7): e019995.

Backor, Kristen, Saar Golde, and Norman Nie (2007). "Estimating Survey Fatigue in Time Use Study". SIQSS Working Paper.

Ballard, Teri, Jennifer Coates, Anne Swindale, and Megan Deitchler (2011). *Household Hunger Scale: Indicator Definition and Measurement Guide.*

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman (2020). "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa". *American Economic Journal: Applied Economics* 12 (1): 210–259.

Barker, Nathan, Gharad T. Bryan, Dean Karlan, Angela Ofori-Atta, and Christopher R. Udry (2021). "Mental Health Therapy as a Core Strategy for Increasing Human Capital: Evidence from Ghana". Working Paper 29407. National Bureau of Economic Research.

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

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez (2020). "The Long-Term Impacts of Grants on Poverty: Nine-Year Evidence from Uganda's Youth Opportunities Program". *American Economic Review: Insights* 2 (3): 287–304.

Blattman, Christopher, Eric P. Green, Julian Jamison, M. Christian Lehmann, and Jeannie Annan (2016). "The Returns to Microenterprise Support among the Ultrapoor: A Field Experiment in Postwar Uganda". *American Economic Journal: Applied Economics* 8 (2): 35–64.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan (2017). "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia". *American Economic Review* 107 (4): 1165–1206.

Bobonis, Gustavo J., Melissa Gonzalez-Brenes, and Roberto Castro (2013). "Public Transfers and Domestic Violence: The Roles of Private Information and Spousal Control". *American Economic Journal: Economic Policy* 5 (1): 179–205.

Brzozowski, Matthew, Thomas F. Crossley, and Joachim K. Winter (2017). "A comparison of recall and diary food expenditure data". *Food Policy* 72: 53–61.

Buller, Ana Maria, Amber Peterman, Meghna Ranganathan, Alexandra Bleile, Melissa Hidrobo, and Lori Heise (2018). "A Mixed-Method Review of Cash Transfers and Intimate Partner Violence in Low- and Middle-Income Countries". *The World Bank Research Observer* 33 (2): 218–258.

Cafiero, Carlo, Sara Viviani, and Mark Nord (2018). "Food security measurement in a global context: The food insecurity experience scale". *Measurement*.

Cunha, Jesse M, Giacomo De Giorgi, and Seema Jayachandran (2018). "The Price Effects of Cash Versus In-Kind Transfers". *The Review of Economic Studies* 86 (1): 240–281.

Deaton, Angus (1997). *The analysis of household surveys*. The World Bank.

Devries, K. M., Joelle Y. Mak, Loraine J. Bacchus, Jennifer C. Child, Gail Falder, Max Petzold, Jill Astbury, and Charlotte H. Watts (2013a). "Intimate Partner Violence and Incident Depressive Symptoms and Suicide Attempts: A Systematic Review of Longitudinal Studies". *PLOS Medicine* 10 (5): e1001439.

Devries, K. M., Joelle YT Mak, Claudia Garcia-Moreno, Max Petzold, James C Child, Gail Falder, Stephen Lim, Loraine J Bacchus, Rebecca E Engell, Lisa Rosenfeld, et al. (2013b). "The global prevalence of intimate partner violence against women". *Science* 340 (6140): 1527–1528.

Dhar, Diva, Tarun Jain, and Seema Jayachandran (2018). "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India". Working Paper 25331. National Bureau of Economic Research.

Dolnicar, Sara, Bettina Grün, and Friedrich Leisch (2011). "Quick, Simple and Reliable: Forced Binary Survey Questions". *International Journal of Market Research* 53 (2): 231–252.

Domingo, Pilar, Rebecca Holmes, Alina Rocha Menocal, and Nicola Jones (2013). "Assessment of the evidence of links between gender equality, peacebuilding and statebuilding". ODI (Overseas Development Institute) Report.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael W. Walker (2019). "General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya". Working Paper w26600. National Bureau of Economic Research.

Ellsberg, M., L. Heise, R. Peña, S. Agurto, and A. Winkvist (2001). "Researching domestic violence against women: methodological and ethical considerations". *Studies in Family Planning* 32 (1): 1–16.

Ellsberg, Mary, Diana J. Arango, Matthew Morton, Floriza Gennari, Sveinung Kiplesund, Manuel Contreras, and Charlotte Watts (2015). "Prevention of violence against women and girls: what does the evidence say?" *The Lancet* 385 (9977): 1555–1566.

FAO (2013). *Guidelines for measuring household and individual dietary diversity*.

Filmer, Deon, Jed Friedman, Eeshani Kandpal, and Junko Onishi (forthcoming). "Cash Transfers, Food Prices, and Nutrition Impacts on Ineligible Children". *Review of Economics and Statistics*.

Fink, Günther, B. Kelsey Jack, and Felix Masiye (2020). "Seasonal Liquidity, Rural Labor Markets, and Agricultural Production". *American Economic Review* 110 (11): 3351–3392.

Fulu, Emma, Rachel Jewkes, Tim Roselli, and Claudia Garcia-Moreno (2013). "Prevalence of and factors associated with male perpetration of intimate partner violence: findings from the UN Multi-country Cross-sectional Study on Men and Violence in Asia and the Pacific". *The Lancet Global Health* 1 (4): e187–e207.

Garcia-Moreno, Claudia, Christina Pallitto, Karen Devries, Heidi Stockl, Charlotte Watts, and Naeema Abrahams (2013). *Global and Regional Estimates of Violence Against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-partner Sexual Violence*. World Health Organization.

Ghandour, Reem M., Jacquelyn C. Campbell, and Jacqueline Lloyd (2015). "Screening and Counseling for Intimate Partner Violence: A Vision for the Future". *Journal of Women's Health* 24 (1): 57–61.

Green, Eric P., Christopher Blattman, Julian Jamison, and Jeannie Annan (2015). "Women's entrepreneurship and intimate partner violence: A cluster randomized trial of microenterprise assistance and partner participation in post-conflict Uganda (SSM-D-14-01580R1)". *Social Science & Medicine* 133: 177–188.

Haushofer, Johannes, Robert Mudida, and Jeremy P. Shapiro (2020). "The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being". Tech. rep. w28106. National Bureau of Economic Research.

Haushofer, Johannes, Charlotte Ringdal, Jeremy P Shapiro, and Xiao Yu Wang (2019). "Income Changes and Intimate Partner Violence: Evidence from Unconditional Cash Transfers in Kenya". Working Paper 25627. National Bureau of Economic Research.

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

— (2018). "The Long-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya". Working Paper.

Heise, Lori (1998). "Violence Against Women: An Integrated, Ecological Framework". *Violence Against Women* 4 (3): 262–290.

Herzog, A. Regula and Jerald G. Bachman (1981). "Effects of Questionnaire Length on Response Quality". *The Public Opinion Quarterly* 45 (4): 549–559.

Hidrobo, Melissa and Lia Fernald (2013). "Cash transfers and domestic violence". *Journal of Health Economics* 32 (1): 304–319.

Hidrobo, Melissa, Amber Peterman, and Lori Heise (2016). "The Effect of Cash, Vouchers, and Food Transfers on Intimate Partner Violence: Evidence from a Randomized Experiment in Northern Ecuador". *American Economic Journal: Applied Economics* 8 (3): 284–303.

Jeong, Dahyeon, Shilpa Aggarwal, Jonathan Robinson, Naresh Kumar, Alan Spearot, and David S. Park (2022). "Exhaustive or Exhausting? Evidence on Survey Fatigue in Long Surveys". National Bureau of Economic Research Working Paper No. 30439.

Jones, Nicola, Janice Cooper, Elizabeth Presler-Marshall, and David Walker (2014). "The fallout of rape as a weapon of war". ODI (Overseas Development Institute) Report.

Kabeer, Naila and Hugh Waddington (2015). "Economic impacts of conditional cash transfer programmes: a systematic review and meta-analysis". *Journal of Development Effectiveness* 7 (3): 290–303.

Kilic, Talip and Thomas Pave Sohnesen (2019). "Same Question But Different Answer: Experimental Evidence on Questionnaire Design's Impact on Poverty Measured by Proxies". *Review of Income and Wealth* 65 (1): 144–165.

Krosnick, Jon A. (1991). "Response strategies for coping with the cognitive demands of attitude measures in surveys". *Applied Cognitive Psychology* 5 (3): 213–236.

Laajaj, Rachid and Karen Macours (2021). "Measuring skills in developing countries". *Journal of Human Resources* 56 (4): 1254–1295.

Lee, David S. (2009). "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects". *The Review of Economic Studies* 76 (3): 1071–1102.

Machisa, Mercilene T., Nicola Christofides, and Rachel Jewkes (2017). "Mental ill health in structural pathways to women's experiences of intimate partner violence". *PLOS ONE* 12 (4): e0175240.

McIntosh, Craig and Andrew Zeitlin (2021). "Cash versus Kind: Benchmarking a Child Nutrition Program against Unconditional Cash Transfers in Rwanda". arXiv: 2106.00213 [econ.GN].

— (2022). "Using household grants to benchmark the cost effectiveness of a USAID workforce readiness program". *Journal of Development Economics* 157: 102875.

Mel, Suresh de, David McKenzie, and Christopher Woodruff (2014). "Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka". *Journal of Development Economics* 106: 199–210.

Millán, Teresa Molina, Tania Barham, Karen Macours, John A Maluccio, and Marco Stampini (2019). "Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence". *The World Bank Research Observer* 34 (1): 119–159.

Moser, Caroline and Ailsa Winton (2002). "Violence in the Central American Region: Towards an Integrated Framework for Violence Reduction". ODI (Overseas Development Institute) Working Paper.

Mummolo, Jonathan and Erik Peterson (2019). "Demand Effects in Survey Experiments: An Empirical Assessment". *American Political Science Review* 113 (2): 517–529.

Omanyondo, Marie-Claire (2005). "Sexual Gender-based Violence and Health Facility Needs Assessment". WHO Report.

Park, David Sungho, Shilpa Aggarwal, Jenny Aker, Dahyeon Jeong, Naresh Kumar, Jonathan Robinson, and Alan Spearot (2021). "Private but Misunderstood? Evaluating the Effects ofSelf-Interviewing to Measure Intimate Partner Violence in a CashTransfer Experiment". Working Paper.

Park, David Sungho, Shilpa Aggarwal, Dahyeon Jeong, Naresh Kumar, Jonathan Robinson, and Alan Spearot (2022a). "Private but Misunderstood? Evidence on Measuring Intimate Partner Violence via Self-Interviewing in Rural Liberia and Malawi". National Bureau of Economic Research Working Paper No. 29584.

— (2022b). "Private but Misunderstood? Evidence on Measuring Intimate Partner Violence via Self-Interviewing in Rural Liberia and Malawi". Unpublished.

Park, David Sungho and Naresh Kumar (2022). "Reducing Intimate Partner Violence: Evidence from a Multifaceted Female Empowerment Program in Urban Liberia". Unpublished.

Parker, Susan W. and Petra E. Todd (2017). "Conditional Cash Transfers: The Case of Progresa/Oportunidades". *Journal of Economic Literature* 55 (3): 866–915.

Peterson, Cora, Megan C. Kearns, Wendy LiKamWa McIntosh, Lianne Fuino Estefan, Christina Nicolaidis, Kathryn E. McCollister, Amy Gordon, and Curtis Florence (2018). "Lifetime Economic Burden of Intimate Partner Violence Among U.S. Adults". *American Journal of Preventive Medicine* 55 (4): 433–444.

Quidt, Jonathan de, Johannes Haushofer, and Christopher Roth (2018). "Measuring and Bounding Experimenter Demand". *American Economic Review* 108 (11): 3266–3302.

Raghunathan, Trivellore E. and James E. Grizzle (1995). "A Split Questionnaire Survey Design". *Journal of the American Statistical Association* 90 (429): 54–63.

Ranganathan, Meghna, Lori Heise, Amber Peterman, Shalini Roy, and Melissa Hidrobo (2021). "Cross-disciplinary intersections between public health and economics in intimate partner violence research". *SSM - Population Health* 14: 100822.

Roberts, Caroline, Emily Gilbert, Nick Allum, and Léïla Eisner (2019). "Research Synthesis: Satisficing in Surveys: A Systematic Review of the Literature". *Public Opinion Quarterly* 83 (3): 598–626.

Roy, Shalini, Melissa Hidrobo, John Hoddinott, and Akhter Ahmed (2019). "Transfers, Behavior Change Communication, and Intimate Partner Violence: Post-program Evidence from Rural Bangladesh". *The Review of Economics and Statistics* 101 (5): 865–877.

Smith, S.G., J. Chen, K.C. Basile, L.K. Gilbert, M.T. Merrick, N. Patel, M. Walling, and A. Jain (2017). "The National Intimate Partner and Sexual Violence Survey (NISVS): 2010-2012 State Report". Tech. rep.

Steenkamp, Chrissie (2005). "The Legacy of War: Conceptualizing a 'Culture of Violence' to Explain Violence after Peace Accords". *The Round Table* 94 (379): 253–267.

Trevillion, Kylee, Sian Oram, Gene Feder, and Louise M. Howard (2012). "Experiences of Domestic Violence and Mental Disorders: A Systematic Review and Meta-Analysis". *PLOS ONE* 7 (12): e51740.

USAID (2019). *Feed the Future Indicator Handbook*.

WFP (2008). *Food consumption analysis: Calculation and use of the food consumption score in food security analysis*.

WHO (2012). "Understanding and addressing violence against women: Intimate partner violence". World Health Organization Report.

— (2016). "Ethical and safety recommendations for intervention research on violence against women". Tech. rep.

— (2021). "Violence Against Women Prevalence Estimates, 2018: Global, regional and national prevalence estimates for intimate partner violence against women and global and regional prevalence estimates for non-partner sexual violence against women". Tech. rep.

Women, UN (2013). "The Contribution of UN Women to Increasing Women's Leadership and Participation in Peace and Security and in Humanitarian Response". UN Women Evaluation Report.