Hongmei Wang
Fang Qiu
Jiangtao Luo

# Impact of unconditional cash transfers

## A replication study of the short-term effects in Kenya

January 2019

Replication Paper 20

Multisector

International Initiative for Impact Evaluation

## About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programs. We are the global leader in funding, producing and synthesizing high-quality evidence of what works, for whom, how, why and at what cost. We believe that using better and policy-relevant evidence helps to make development more effective and improve people's lives.

## 3ie Replication Paper Series

The 3ie Replication Paper Series is designed to be a publication and dissemination outlet for internal replication studies of development impact evaluations. Internal replication studies are those that reanalyze the data from an original paper in order to validate the results. The series seeks to publish replication studies with findings that reinforce or challenge the results of an original paper. To be eligible for submission, a replication study needs to be of a paper in 3ie's online Impact Evaluation Repository and needs to include a pure replication. 3ie invites formal replies from the original authors. These are published on the 3ie website together with the replication study.

The 3ie Replication Program also includes grant-making windows to fund replication studies of papers identified on our candidate studies list. Requests for proposals are issued one to two times a year. The candidate studies list includes published studies that are considered influential, innovative or counterintuitive. The list is periodically updated based on 3ie staff input and outside suggestions. The aim of the 3ie Replication Program is to improve the quality of evidence from development impact evaluations for use in policymaking and program design.

## About this report

The Bill & Melinda Gates Foundation helped fund this report. All content, errors and omissions are the sole responsibility of the authors and do not represent the opinions of 3ie, its donors or its Board of Commissioners. Please direct any comments or queries to the corresponding author, Hongmei Wang, PhD, at hongmeiwang@unmc.edu.

Suggested citation: Wang, H, Qui, Fang, Q and Luo, J, 2018. *Impact of unconditional cash transfers: a replication study of the short-term effects in Kenya.* 3ie Replication Paper 20. Washington, DC: International Initiative for Impact Evaluation (3ie). Available at: https://doi.org/10.23846/RPS0020

3ie Replication Paper Series executive editor: Marie Gaarder
Managing editor: Anna Heard
Production manager: Brigid Monaghan
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: John F McGill and Akarsh Gupta

# Impact of unconditional cash transfers: a replication study of the short-term effects in Kenya

Hongmei Wang
University of Nebraska Medical Center, College of Public Health

Fang Qiu
University of Nebraska Medical Center, College of Public Health

Jiangtao Luo
University of Nebraska Medical Center, College of Public Health

**Replication Paper 20**

**January 2019**

International Initiative for Impact Evaluation

## Summary

More evidence is needed on the impact of unconditional cash transfers on economic and health measures, especially in low-income countries. This replication study attempts to understand the impact of an unconditional cash transfer on asset holdings, consumption, income, education, food security and female empowerment, as well as health and psychological well-being, by reexamining a 2016 study conducted in rural Kenya by Johannes Haushofer and Jeremy Shapiro.

Haushofer and Shapiro's study examined the short-term impacts of unconditional cash transfer – and the differential impacts by transfer recipient's gender (female versus male), timing (monthly versus lump sum) and magnitude (large versus small) – using data collected in a randomized controlled trial from 2011 to 2012. The researchers found the unconditional cash transfer to increase assets, consumption, revenue, food security and psychological well-being indices but to have no overall effects on health, education or female empowerment indices. Compared to lump-sum transfers, monthly transfers improved food security but reduced non-land asset holdings. Large transfers, when compared to small transfers, increased non-land asset holdings and improved the psychological well-being index.

Our pure replication results are consistent with the findings published in the original study on the overall effects of the unconditional cash transfer and the comparison across different treatment arms. These findings are sustained in our robustness checks, including different variance and covariance structure assumptions, different model specifications, multilevel modeling, and multivariate analysis considering correlation among multiple outcome measures. However, the principal component analysis results suggest a need for further examination of the method of measuring food security, health and psychological well-being.

## Acknowledgements

This replication study was funded and facilitated by the International Initiative for Impact Evaluation (3ie) as part of its replication series. We are thankful to the 3ie reviewers, Dr Benjamin Wood, Dr John Creamer and Dr Eric Djimeu, as well as several anonymous reviewers, for their helpful comments. We also thank the program manager, Scott Neilitz, for his help and support at each stage of the replication. We thank Nicholas Hein, Harlan Sayles and Morshed Alam from the Department of Statistics at the University of Nebraska Medical Center for their help with coding in SAS and Stata.

We give special thanks to the original authors, Dr Johannes Haushofer and Dr Jeremy Shapiro, for kindly sharing their codes, data set and methodological documents for the replication study.

## Contents

## List of tables



## Appendix figure and tables

## Abbreviations and acronyms

| | |
|---|---|
| ANCOVA | Analysis of covariance |
| CES-D | Center for Epidemiology Studies – Depression Scale |
| DID | Difference in difference |
| FWER | Family-wise error rate |
| HH | Household |
| OLS | Ordinary least squares |
| PCA | Principal component analysis |
| PPP | Purchasing power parity |
| SD | Standard deviation |
| SE | Standard error |
| SUR | Seemingly unrelated regression |
| UCT | Unconditional cash transfer |
| WVS | World Values Survey |

# 1. Introduction

Cash transfers have become one of the most widespread poverty interventions because of their well-documented benefits. The benefits of unconditional cash transfer (UCT) include increasing consumption in general and increasing food consumption and diversity in particular, as well as improving health and education outcomes and potentially increasing income and capital investments (Baird et al. 2011; Cesarini et al. 2015; Merttens et al. 2013; Attah et al. 2016).

The current debate has focused on whether the impact of UCT is short term or can sustain and lead to long-term transformative impact in economy. Compared to in-kind transfers, cash transfers are not distortionary, can meet heterogeneous needs for welfare improvement, have psychological benefits by empowering recipients and have lower delivery costs (Baird et al. 2011). Compared to conditional cash transfer, UCT is cheaper to implement and may be inferior in improving outcomes related to the set conditions, but is superior in improving other outcomes (Baird et al. 2011).

However, UCT might be spent on temptation goods and decrease welfare in the long run, and its income effects could reduce labor supply and lead to inflation over time (Cesarini et al. 2015). In reviewing the literature on cash transfer, Wydick (2018) suggests that although increased spending on temptation goods and reduction in labor supply as a result of cash transfers are observed in some advanced economies, such as the United States, these effects are not present in many studies carried out in low-income economies. More evidence regarding the impact of UCT in low-income countries is needed to support poverty alleviation efforts.

Impact evaluations are helpful for studying the influence of UCTs on various economic and welfare measures in order to inform policy and program design in low-income countries. Haushofer and Shapiro's study (2016) used a randomized controlled trial to examine the effects of a large UCT on economic outcomes and psychological well-being of poor households in rural Kenya. The study adopted a two-level, cluster-randomized controlled trial strategy by randomizing the UCT to households in villages where the treatment status was also randomly assigned. In the treatment households, the study randomly assigned cash transfers by recipient's gender, transfer size and transfer frequency. The experimental design of this study helped establish the causal effects of the UCT on cash transfer recipients in Kenya. The design also allowed the examination of differential impacts between the treatment arms.

Haushofer and Shapiro (2016) used an analysis of covariance (ANCOVA) model and reported statistically significant and economically meaningful impacts of cash transfers on economic outcomes and psychological well-being in the poor Kenyan households. Households that received UCT had significantly higher household consumption, asset holdings and monthly income. They also scored higher on the food security index and psychological well-being index. The same households did not report significant improvement in health, education or female empowerment.

The results comparing different treatment arms suggested that monthly payments were more likely than lump-sum transfers to increase food security, while lump-sum transfers led to higher levels of asset holdings. Compared to small cash transfers, larger transfers increased asset holdings and improved the psychological well-being of household members. The study found little evidence that providing cash transfers to women versus men differentially affected outcomes.

The findings from Haushofer and Shapiro's study (2016) are consistent with other studies in Kenya, which find that UCTs have positive effects on economic outcomes and psychological well-being for recipients (Merttens et al. 2013; Attah et al. 2016). The insights gained from the study also shed light on the specific mobile money technology used and the transfer design – in terms of recipients' gender, magnitude and frequencies – that fit the developing world. Their study supports the findings from another recent study on various cash transfer methods, which concluded mobile money technology was particularly effective in Kenya (Barca et al. 2013). Haushofer and Shapiro's study also makes a unique contribution to the health literature by examining the impact of UCT on recipients' health and psychological well-being.

This replication report independently reproduces the findings reported in *The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya* (Haushofer and Shapiro 2016). The pure replication analysis reproduces the main results of the original study and helps verify these important findings on the short-term impact of UCT. Through a pure replication of the Haushofer and Shapiro 2016 study, we contribute to the field by adding evidence on the short-term impact of UCT on a variety of outcome measures in a low-income economy. These findings provide policymakers with solid evidence to help them select the most effective and efficient policies to fight poverty, one of the key Sustainable Development Goals from United Nations.

In addition to a pure replication of the original study, we conducted extensive robustness checks on the measures and analytical methods used in the original study, interrogating specification choices and inference procedures. We scrutinized the methodology used in the original study and examined the model assumptions, verifying the regression methods chosen – ANCOVA versus difference in difference (DID) – examining different ways of treating missing values, testing on the treatment village and treatment gender interaction effects, and conducting multilevel analysis for psychological well-being measures for which both household and village-level correlations may exist. We also conducted principal component analysis (PCA) to examine index measures created to measure food security, health and psychological well-being and adopted multivariate analysis to consider the correlation among multiple outcome measures. The results of these robustness checks, and of multivariate analysis, support Haushofer and Shapiro's findings on the short-term impact of UCT in rural Kenya.

We also found village treatment interaction effects, which suggests that the treatment effects can vary in different villages and that a context analysis at the village level could be informative for understanding intervention implementation and the channels for differential impact. The PCA results suggest that the items used to create index measures in food security, health and psychological well-being may describe more than one feature in each domain. Additional studies in these three domains will be helpful for identifying appropriate measures in food security, health and psychological well-being and to accurately measure the impact of UCT in these domains.

The replication study is valuable not only to the contestability of findings particular to impact of UCT, but also to the applications of statistical methodologies in this field. The main results of the original study were replicated nearly exactly, with a few minor exceptions around particular inference procedures. Such results provide strong support for the empirical findings of the original study, on the one hand, and create opportunities for dialogue around the authors' specific methodologies, on the other. Researchers rarely have the chance to examine a research question using the same empirical data. The replication study helps address some of the recent controversy around replication in the social sciences and helps build scientific evidence for policy action and academic significance (Freese and Peterson 2017).

## 2. Replication methodology

### 2.1 Review of the original study

The Haushofer and Shapiro study used cluster random sampling method for treatment assignment to eligible households in poor villages in Kenya. Roofing material (thatched versus metal) was the eligibility criterion to target poor households and villages. From Rarieda, Kenya, 120 villages with the highest proportion of thatched roofs were chosen for the study, of which 60 villages were randomly assigned to receive cash transfers (treatment villages) and the other 60 villages were designated as control villages. All households from the treatment villages were assessed for whether they had a thatched roof at baseline. Half of the eligible households in the treatment villages were randomized into the treatment group to receive cash transfers (503 treatment households) and the other half were in the control group (505 spillover households).

Baseline and endline surveys were collected in the treatment households and spillover households. A total of 432 households with thatched roofs were randomly chosen from control villages at endline as pure control households, and only endline surveys were collected in these households. Baseline information was collected from April to November 2011. Endline data were collected between August and December 2012. Among the treatment households, the study randomly assigned cash transfers by recipient's gender (husband versus wife), monthly transfers versus lump-sum transfers, and large transfers (USD 1,525 purchasing power parity [PPP]) versus small transfers (USD 404 PPP). The researchers used the spillover households as the control group when examining the impact of UCT and treatment arms and conducting analyses based on their pre-analysis plan (Haushofer and Shapiro 2013).

The main analyses included examination of overall effects of cash transfers and the differential effects of treatment arms on indices in eight domains: assets, revenue, expenditure, food security, health, education, psychological well-being and female empowerment. When estimating the main impacts of the cash transfer, Haushofer and Shapiro conducted ANCOVA to examine the association between the index variable measures at endline and cash transfers, controlling for village-level fixed effects, index variable measures at baseline and household-level correlation of the error terms (McKenzie 2012).

To address the multiple inference issues related to the multiple outcomes, the authors used the family-wise error rate (FWER) to compute corrected p-values (Anderson 2008).

Additionally, they estimated equations jointly, using seemingly unrelated regression (SUR), and reported the joint significance of the treatment coefficient using Wald tests. They also examined the impacts of cash transfers and the recipient's gender, transfer frequency and transfer magnitude on individual component measures of psychological well-being, consumption, assets and income indices. The spillover households were used as a control group when examining the treatment effects, because the pure control group did not have baseline measures collected.

The researchers examined the validity of the main analyses by checking (1) baseline differences in index variables between treatment and spillover groups and (2) the spillover effects of the index variables by comparing the spillover group to the pure control group. Because pure control households were identified at the endline using the thatched roof criterion, they could differ from the spillover group, as some spillover households might have upgraded their roofs from thatch to metal by endline. The researchers estimated the group difference at five spillover households with metal-roof upgrades. They also conducted robustness checks on the spillover effects by estimating Lee bounds and Horowitz–Manski bounds using imputed data (Lee 2009; Horowitz and Manski 1995). Results of these analyses suggest that the spillover effects were small and unlikely to distort the identified treatment effects.

## 2.2 Pure replication

We designed our pure replication to independently test the main study findings by recreating all tables (I–VI) in the main body of the original study. We did not replicate the three tables (A.1–A.3) included in the appendix of the original study, of which Table A.1 reports ex post minimum detectable effect sizes, Table A.2 describes outcome variables and Table A.3 lists discrepancies in the analyses from the pre-analysis plan.

Our pure replication focused on analyses of the impact of cash transfers, as shown in Tables I–VI. Specifically, Table I examines the validity of the main analyses by checking baseline differences in index variables between treatment and spillover groups. Table II reports the overall effects of UCT and the differential effects of treatment arms on indices in eight domains (assets, revenue, expenditure, food security, health, education, psychological well-being and female empowerment). In Table III, the researchers examine and report the spillover effects of the index variables by comparing spillover households and pure control households. The researchers also report Lee bounds and Horowitz–Manski bounds for the spillover effects in Table III.

In addition, the authors examine the impacts of cash transfers and the recipient's gender, transfer frequency and transfer magnitude on individual component measures of psychological well-being (Table IV), consumption (Table V) and asset and income indices (Table VI). FWERs are reported only in Tables I and II, and joint significance by SUR estimates are reported in Tables I–VI.

By reproducing the analyses in Tables I–VI, we were able to validate the main findings reported in the original paper on the short-term impacts of UCT in poor households in Kenya.

## 2.3 Measurement and estimation analysis

We examined the robustness of the findings through additional measurement and estimation analysis. As shown in the final replication plan (Wang and Luo 2017), the measurement and estimation analysis included analyses on (1) model validation, (2) alternative model specification and (3) alternative measures of health, food security and psychological well-being indices.

### *2.3.1 Model validation*

For model validation, we examined the normal distribution and equal variance assumptions of the models for the main treatment effects of the UCT. In the original paper, the main outcomes of interest are assets, consumption, revenue, food security index, health index, education index, psychological well-being index and female empowerment index. The original paper does not discuss descriptive statistics related to economic measures, such as assets, consumption and revenue, which were likely to be right-skewed and deviate from the normal distribution assumption for inaccurate inference conclusions (Metcalf 1972). We used a graphical approach, histogram plots and residual plots to examine the normality and unequal variance assumption (Kutner et al. 2004; Kuehl 1999).[1]

In addition, we examined the autocorrelation and variability of the eight outcome indices to see if DID models should be used for analysis. The authors chose the ANCOVA model with village-level fixed effects, based on the McKenzie (2012) study which argues that ANCOVA estimates are preferable to DID estimates for outcome measures with high variability and low autocorrelation. Although economic outcomes such as income and expenditure are subject to high variability and low autocorrelation, the DID model is suitable for studying highly autocorrelated and relatively precisely measured outcomes in health and education domains (McKenzie 2012; Bertrand and Duflo 2004). McKenzie (2012) shows that the DID estimator is preferred in terms of power if the autocorrelation is greater than 0.5 when outcome measures were measured at baseline and with a single follow-up.

We examined the autocorrelation of the outcome indices to validate the adoption of ANCOVA using a cut-off point of $p > 0.5$. It is hypothesized that health, food security and psychological well-being measures in these poor households in Kenya have relatively large autocorrelation ($p > 0.5$). We conducted DID analysis for outcome variables with a high autocorrelation ($p > 0.5$) and compared DID estimates to the authors' to examine the robustness of the treatment effects.

We also conducted robustness checks on how missing values were treated. The authors of the original study used a missing-indicator method to handle missing baseline data by adding a dummy variable to indicate missing data and an interaction term of baseline data and missing data dummy. This practice is equivalent to the imputation of zeros for missing baseline measures. Jones (1996) investigates the possible bias in the estimators of the regression coefficients and residual variance derived from the missing-indicator method of handling missing data. The article suggests that this method may produce

[1] There are several tests for normality, such as the Shapiro–Wilk test, Kolmogorov–Smirnov test, Lilliefors test and Anderson–Darling test. However, these do not provide good power.

biased estimates if missing is not completely at random and may have "no advantage over the complete-case analysis" (Jones 1996). Although randomized controlled trials ensure the randomization of treatment, they do not guarantee that missing baseline measures are completely at random. We examined the robustness of this method by conducting complete case analysis on outcomes with missing baseline values.

For the psychological well-being index measure, we conducted multilevel modeling to consider both household- and village-level correlations when examining the treatment effects. Psychological outcomes were measured at the individual level, whereas other outcomes were measured at the household level. For psychological outcomes, there are potentially two levels of correlation that need to be considered – the village-level correlation and the household-level correlation. Failing to recognize the nested structure may result in underestimated standard errors of regression coefficients, leading to an overstatement of statistical significance. The original paper modeled village-level fixed effects and controlled household-level correlation for individual-level outcome measures.

An alternative method for estimation of this type of data is generalized linear mixed models, also known as hierarchical or multilevel models (Cohen et al. 2003; Gelman and Hill 2007). This model allows for estimation of error terms that correlated in two levels of clusters and increase efficiency of estimation. We conducted multilevel modeling for the psychological well-being index, as it considers the nested structure of the data collected by controlling correlation at both household and village levels.

#### *2.3.2 Modeling interaction effects*

The authors modeled village as a fixed effect, but they did not consider potential interaction effects between village and treatment or between recipient's gender and treatment. Some underlying village features may moderate the impacts of cash transfers; adding the interaction terms in the model allows estimation of such effects. In addition, the recipient's gender may be related to the effects of UCT, as wife and husband may play different roles in household decision-making. We examined two interaction effects – gender with treatment and village with treatment – and tested whether these interaction terms were significantly related to the eight outcomes of interest.

#### *2.3.3 Examining index measures of food security, health and psychological well-being*

The authors used indices in eight domains to reduce the number of outcome variables for estimation, following the weighted, standardized average approach (Anderson 2008). Although this practice is reasonable for conventional economic measures, such as consumption and income, it may not capture the underlying structure for complex measures of health, food security and psychological well-being. We conducted PCA on variables used in creating health, food security and psychological well-being indices to identify the underlying factors in each domain. The PCA results helped verify whether all variables used to create the index actually described one underlying factor (or more than one relevant but different factors) and suggested appropriate weights (factor loadings) to create an index based on these variables (Vyas and Kumaranayake 2006; Johnson and Wichern 2007).

## 2.4 Theory of change analysis

We extended the original study by conducting multivariate analysis of variance on the eight indices and on the individual measures in asset holdings, consumption and psychological well-being. Different from adjusting the p-values using a predetermined method to account for multiple inference issues in the FWER method, the multivariate analysis fits a simultaneous regression to multiple outcomes that are correlated. This methodology allows for the complete modeling of all data in one analysis, testing of correlations between multiple outcomes and direct estimation of the difference in the association between treatment effects on multiple outcomes. By employing a multivariate model, it is possible to gain precision, compared to estimating separate models for each outcome (Anderson 2003). We conducted multivariate analysis to examine the differential treatment effects on the correlated multiple outcomes, including the eight indices and individual measures composed of asset holdings, consumption and psychological well-being indices.

## 2.5 Data

The Haushofer and Shapiro study included four data sets, one of which was the final, clean data set used to generate the findings included in the published paper. The authors provided us with the final, cleaned data set in Stata. The data set contained data for 1,440 participating households, comprising 503 treatment households, 505 spillover households and 432 pure control households.[2] The pure replication used this de-identified, final, cleaned data set to reassess the impacts of the UCT in poor households in Kenya. Although there was no codebook accompanying the data, the variables included in the final data set were labeled with a general description of the variables, which we used as a codebook.

We also downloaded a Microsoft Excel file from the author's website,[3] which included all questions used in the endline survey for reference. The data covered a wide range of topics related to household measures on economic, education, food security, health outcomes and individual measures of psychological well-being. The data set also included information on treatment status and demographic information of the study participants.

## 2.6 Measures

The outcome measures used in the analyses included eight index variables, calculated as a standardized total or standardized average of the individual components in the eight areas of interest: asset holdings, consumption in specific categories, income and business activities, food security measures, education, health and psychological well-being measures. A number of measures on psychological well-being, consumption, assets and business activities were also used as outcome variables. For each outcome variable, there were both baseline and endline measures.

In addition, an indicator variable was included to show whether the baseline measure was missing for each outcome variable. Most of these outcome variables were collected

[2] The data set is available at http://www.princeton.edu/haushofer/.

[3] http://www.princeton.edu/haushofer/

through survey methods, except the salivary cortisol levels. Cortisol level was a measure of psychological well-being, obtained by collecting saliva samples from all respondents twice, at the beginning and at the end of the individual survey. A detailed explanation of all outcome variables used in the analyses is presented in Table A.2 of the original paper (Appendix Table A1 in this paper).

### 2.7 Analyses

The original authors acknowledged a few errors in the tables in their original paper and submitted an erratum to *The Quarterly Journal of Economics* (Haushofer and Shapiro 2017). The code they supplied to us was the corrected code. In our replication, we compared the reproduced tables to the final tables reported in the erratum if they were different from those in the original paper.

The authors also provided us with Stata codes used for the analyses. From these codes, we were able to obtain detailed information on the exact regression models, the sample selection and the exact variables used for each model to allow for a pure replication of the original analyses. We conducted the pure replication using the same regression models and conducting the same analyses to regenerate Tables I–VI.

The main statistical software used for our pure replication was SAS version 9.4, supplemented by Stata 14.1. We created SAS codes independently to run the main analyses to produce all six tables, including generation of FWERs. We used the SAS SURVEYREG procedure for all main analyses, which allowed for clustering by household or by village when examining the association between outcome variables and key predictor variables.

For SUR results and results related to Lee bounds and Horowitz–Manski bounds reported in Table III, we used Stata to rerun the analyses as standardized commands; procedures specific to these analyses were available in Stata 14.1 but not in SAS 9.4. We used SAS 9.4 for all measurement and estimation analyses and theory of change analyses. The replication study used entirely different software for most of the analyses, which adds to the robustness of the study results.

## 3. Pure replication results

### 3.1 Baseline differences: index variables

We began our pure replication by reproducing Table I, the baseline differences in index variables. These results compared the baseline summary statistics (mean and standard errors) between treatment households and spillover households. To ensure the integrity of the experimental design, it was important to examine whether the randomization led to similar baseline characteristics between groups. Regression models were run using the baseline eight index variables as outcome variables and treatment status as main explanatory variables, controlling for village-level fixed effects and clustered errors in households for the psychological well-being index.

Table 1 shows the replicated results, comparing eight index variables – value of non-land assets, non-durable expenditure, total monthly revenue, food security index, health index, education index, psychological well-being index and female empowerment index –

between treatment groups or treatment arms. Using the treatment status from the data set, we reproduced the means and standard errors of the group characteristics (Table 1).

We produced exactly the same results for means, coefficients and standard errors as published in Table I for all eight indices in all five columns. However, there are some differences in the FWER-corrected p-values in Table 1. We have shaded the cells in Table 1 where the difference between the reported FWERs and those reported in the original table was greater than one-hundredth. We include Table I from the original paper as Appendix B for comparison purposes.

**Table 1: Baseline differences on index variables, a reproduction of Table I in the original paper**

| | Control mean (SD) | Treatment effect | Female recipient | Monthly transfer | Large transfer | N |
|---|---|---|---|---|---|---|
| Value of non-land assets (USD) | 383.36 | –1.15 | 15.53 | 25.16 | 13.76 | 1,008 |
| | (374.15) | (24.74) | (43.62) | (39.33) | (42.77) | |
| | | [1.00] | [0.80] | [0.71] | [0.93] | |
| Non-durable expenditure (USD) | 181.99 | –6.16 | –28.05* | –8.01 | –5.56 | 1,008 |
| | (127.16) | (8.31) | (15.14) | (13.28) | (14.36) | |
| | | [0.96] | [0.33] | [0.71] | [0.93] | |
| Total revenue, monthly (USD) | 84.92 | –33.19* | –31.77** | –7.59 | –10.77 | 1,008 |
| | (402.59) | (18.54) | (14.34) | (14.99) | (12.38) | |
| | | [0.44] | [0.19] | [0.71] | [0.93] | |
| Food security index | 0.00 | 0.00 | 0.05 | 0.25** | –0.01 | 1,008 |
| | (1.00) | (0.06) | (0.09) | (0.10) | (0.09) | |
| | | [1.00] | [0.80] | [0.12] | [0.93] | |
| Health index | 0.01 | 0.03 | 0.26*** | 0.14 | –0.14 | 1,008 |
| | (1.02) | (0.06) | (0.09) | (0.10) | (0.10) | |
| | | [0.97] | [0.06]* | [0.56] | [0.72] | |
| Education index | 0.00 | –0.07 | 0.14 | 0.16* | –0.05 | 853 |
| | (1.00) | (0.06) | (0.09) | (0.09) | (0.09) | |
| | | [0.90] | [0.42] | [0.46] | [0.93] | |
| Psychological well-being index | 0.00 | 0.03 | 0.02 | 0.19** | 0.18** | 1,569 |
| | (1.00) | (0.05) | (0.08) | (0.08) | (0.08) | |
| | | [0.97] | [0.80] | [0.18] | [0.23] | |
| Female empowerment index | 0.00 | –0.05 | 0.08 | 0.18 | 0.03 | 751 |
| | (1.00) | (0.07) | (0.11) | (0.12) | (0.12) | |
| | | [0.96] | [0.80] | [0.56] | [0.93] | |
| Joint test (p-value) | | 0.64 | 0.00*** | 0.02** | 0.36 | |

Note: Ordinary least squares estimates of baseline differences in treatment arms. Outcome variables are listed on the left and described in detail in Table A1. For each outcome variable, we report the coefficients of interest and their standard errors in parentheses. FWER-corrected p-values are in brackets. Column (1) reports the mean and standard deviation of the control group for a given outcome variable. Column (2) reports the basic treatment effect, i.e. comparing treatment households to control households within villages. Column (3) reports the relative treatment effect of transferring to the female compared to the male; column (4) the relative effect of monthly compared to lump-sum transfers; and column (5) that of large compared to small transfers. The unit of observation is the household for all outcome variables except for the psychological variables index, where it is the individual. The sample is restricted to cohabitating couples for the female empowerment index and households with school-age children for the education index. The comparison of monthly to lump-sum transfers excludes large-transfer recipient households, and that for male versus female recipients excludes single-headed households. All columns include village-level fixed effects and cluster standard errors at the household level. The last row shows joint significance of the coefficients in the corresponding column from SUR estimation. * denotes significance at 10 percent, ** at 5 percent and *** at 1 percent level.

As we produced the FWER-adjusted p-values using bootstrap iterations, we expected to see some differences in these p-values due to randomness. However, as we used the same method, following Anderson (2008), the same seed to simulate data, and the same number of bootstrap iterations (10,000 times) as the authors of the original study to generate these FWER, the differences due to randomness alone should be very small. A close comparison between our SAS code and the original Stata code suggested that the observed differences in the FWER-adjusted p-values were due mainly to the differential practices of ranking FWER-adjusted p-values according to the order of the eight p-values using the actual data.[4]

We ranked the FWER-adjust p-values to enforce monotonicity, following the algorithm in the Anderson (2008) paper, whereas the authors used pairwise comparison when determining the final FWER-adjusted p-values. These differences in FWER-adjusted p-values did not change the study findings regarding the baseline differences between groups, except in two places: the difference in food security index between monthly transfer and lump-sum transfer households (p-value reported in the original paper is 0.08, but is 0.12 in our table), and the difference in health index between female recipient and male recipient households (p-value 0.04 in the original paper, but 0.06 in our table).

Using conventional p-values, our replication results are consistent with the original findings. There was no difference between treatment and spillover groups in all eight index variables at the 0.05 significance level. Compared to male recipient households, female recipient households had a higher health index score (0.26) and a lower income (USD 32 PPP) at baseline. Compared to lump-sum transfer households, monthly transfer households had a higher food security index score (0.25) and higher psychological well-being index score (0.19) at baseline (at a significance level of 0.05 or lower). Compared to small-transfer households, households receiving large transfers had a higher psychological well-being index (0.18) at baseline.

Using FWER-adjusted p-values, no difference was found between treatment households and spillover households or between large-transfer households and small-transfer households. SUR results supported these findings. The authors of the original paper found a higher health index score in female recipient households (0.26) than in male recipient households at baseline, with a FWER p-value of 0.04. Based on the reproduced FWER p-value (p = 0.06), the difference in health index between these two groups was only significant at the 0.10 level.

SUR results supported some baseline differences between female recipient households and male recipient households (p = 0.00). The authors of the original paper found a

---

[4] In our pure replication, we re-ranked the adjusted p-values according to Anderson (2008) and using the algorithm $p_r fwer$ = (min {$p_r fwer^*$, $p_{r+1} fwer^*$, …$p_m fwer^*$}, whereas the authors of the original study used equation $p_r fwer$ = max {$p_r fwer^*$, $p_{r-1} fwer^*$}. Here, $p_r fwer^*$ is the p-values based on permutated data and *r* is the order of p-values for the eight outcome variables, from most significant (smallest p-value) to least significant (highest p-value), ranging from 1 to *m* (*m* = 8 in Table 1). In addition, for analyses to generate column 3 and column 4 in the table, we permutated both the main treatment variable and the related interaction terms to ensure the consistency of treatment status and the interaction terms, whereas the authors of the original study permutated only the main treatment variable of interest. The differences in the FWER-adjusted p-values resulted mainly from the different practice of re-ranking to enforce monotonicity according to the order of the actual p-values.

higher food security index score in monthly transfer households (0.25) than that in lump-sum households with a FWER p-value of 0.08. Based on the reproduced FWER p-values (p = 0.12), the difference in food security index between these two groups was no longer statistically significant. SUR results support some baseline differences between monthly transfer and lump-sum transfer households (p = 0.02).

Although there were some minor differences in the FWERs we reproduced, compared to those reported in the original paper, these differences strengthened the original authors' conclusion regarding the integrity of the experiment, as baseline differences were not identified between treatment groups using FWER-adjusted p-values at the 0.05 level of significance.

### 3.2 Treatment effects: index variables

Table 2 reproduces the main results reported in Table II of the original paper, the impacts of the cash transfer program on the eight index variables in Kenya. To capture the treatment effects, village-level fixed effects models were run on treatment status by controlling baseline index variable measures, and on a dummy variable indicating whether the baseline outcome variable was missing. For psychological well-being index variable, standard errors are clustered at the household level.

Column 1 shows the mean and standard errors of the endline measures of the eight index variables in spillover households. Column 2 shows the differences in index variables between the treatment households and spillover households. Column 3 shows the differences in the eight index variables between the female recipient two-headed households and male recipient two-headed households. Column 4 shows the differences between the monthly and lump-sum recipient households. Column 5 shows the differences between households receiving large transfers and those receiving small transfers.

**Table 2: Treatment effects on index variables, a reproduction of Table II in the original paper**

| | Control mean (SD) | Treatment effect | Female recipient | Monthly transfer | Large transfer | N |
|---|---|---|---|---|---|---|
| Value of non-land assets (USD) | 494.80 | 301.51*** | –79.46 | –91.85** | 279.18*** | 940 |
| | (415.32) | (27.25) | (50.38) | (45.92) | (49.09) | |
| | | [0.00]*** | [0.53] | [0.30] | [0.00]*** | |
| Non-durable expenditure (USD) | 157.61 | 35.66*** | –2.00 | –4.20 | 21.25** | 940 |
| | (82.18) | (5.85) | (10.28) | (10.71) | (10.49) | |
| | | [0.00]*** | [0.85] | [0.90] | [0.21] | |
| Total revenue, monthly (USD) | 48.98 | 16.15*** | 5.41 | 16.33 | –2.44 | 940 |
| | (90.52) | (5.88) | (10.61) | (11.07) | (8.87) | |
| | | [0.02]** | [0.85] | [0.62] | [0.79] | |
| Food security index | 0.00 | 0.26*** | 0.06 | 0.26** | 0.18* | 940 |
| | (1.00) | (0.06) | (0.09) | (0.11) | (0.10) | |
| | | [0.00]*** | [0.85] | [0.15] | [0.26] | |
| Health index | 0.00 | –0.03 | 0.10 | 0.01 | –0.09 | 940 |
| | (1.00) | (0.06) | (0.09) | (0.10) | (0.09) | |
| | | [0.82] | [0.72] | [0.90] | [0.72] | |
| Education index | 0.00 | 0.08 | 0.06 | –0.05 | 0.05 | 823 |
| | (1.00) | (0.06) | (0.09) | (0.10) | (0.09) | |
| | | [0.43] | [0.85] | [0.90] | [0.79] | |
| Psychological well-being index | 0.00 | 0.26*** | 0.14* | 0.01 | 0.26*** | 1474 |
| | (1.00) | (0.05) | (0.08) | (0.08) | (0.08) | |
| | | [0.00]*** | [0.45] | [0.90] | [0.01]*** | |
| Female empowerment index | 0.00 | –0.01 | 0.17* | 0.05 | 0.22** | 698 |
| | (1.00) | (0.07) | (0.10) | (0.12) | (0.11) | |
| | | [0.88] | [0.51] | [0.90] | [0.21] | |
| Joint test (p-value) | | 0.00*** | 0.11 | 0.04** | 0.00*** | |

Note: Ordinary least squares estimates of treatment effects. Outcome variables are listed on the left and described in detail in Table A1. Higher values correspond to "positive" outcomes. For each outcome variable, we report the coefficients of interest and their standard errors in parentheses. FWER-corrected p-values are in brackets. Column (1) reports the mean and standard deviation of the spillover group. Column (2) reports the basic treatment effect, i.e. comparing treatment households to control households within villages. Column (3) reports the relative treatment effect of transferring to the female compared to the male; column (4) the relative effect of monthly compared to lump-sum transfers; and column (5) that of large compared to small transfers. The unit of observation is the household for all outcome variables except for the psychological variables index, where it is the individual. The sample is restricted to cohabitating couples for the female empowerment index and to households with school-age children for the education index. The comparison of monthly to lump-sum transfers excludes large-transfer recipient households, and that for male versus female recipients excludes single-headed households. All columns include village-level fixed effects, control for baseline outcomes and cluster standard errors at the household level. The last row shows joint significance of the coefficients in the corresponding column from SUR estimation. * denotes significance at 10 percent, ** at 5 percent and *** at 1 percent level.

We produced the exact same results for means, coefficients and standard errors as reported in Table II for all eight indices in all five columns. Again, as expected, we found some differences in the FWER-adjusted p-values. In Table 2, we shade the cells where the difference between the reported FWERs and the one reported in the original table was greater than one-hundredth. We include Table II from the original paper as Appendix C for comparison.

We used the same method as the Anderson (2008) article – same seed to generate data and same number of bootstrap iterations (10,000 times) as the authors of the original study – to generate these FWERs. We used the algorithm from Anderson (2008) to rank p-values and permutated all relevant treatment interactions terms when generating FWER-adjusted p-values in column 3 and column 4. The differences in FWER-adjusted p-values were mainly due to the different practices of ranking FWER-adjusted p-values. These differences in FWER-adjusted p-values did not change the study findings regarding the treatment effects between the treatment group and between treatment arms.

Main treatment effects are as follows. Treatment households reported higher values of non-land assets (301.51 USD PPP), higher non-durable expenditure (35.66 USD PPP), higher monthly revenue (16.55 USD PPP), higher food security index (0.26) and higher psychological well-being index (0.26) than those from the spillover group. No improvement in health, education or female empowerment was found in treatment households when compared to spillover households. SUR results supported treatment effects (p = 0.00) at a significance level of 0.01.

Between treatment arms, no statistically significant differences were found between female recipient households and male recipient households at the 0.05 significance level. Compared to lump-sum transfer households, households receiving monthly transfers reported a lower value of non-land assets (91.85 USD PPP) and a higher food security index (0.26), but neither effect survived FWER correction. SUR results supported treatment effects between monthly versus lump-sum transfers (p = 0.04). Compared to small-transfer households, households receiving large transfers reported a higher value of non-land assets (279.18 USD PPP) and a higher psychological well-being index (0.26), and both effects maintained significance using FWER-adjusted p-values. The group differences in non-durable expenditure (21.25 USD PPP) and female empowerment index (0.22) were significant at the 0.05 level, using conventional p-values, but no longer significant when using FWER-adjusted p-values. SUR results supported treatment effects between large and small cash transfers (p = 0.00).

Consistent with the original findings, our replication results suggest treatment effects on non-land assets, non-durable expenditure, monthly revenue, food security index and psychological well-being but not on health, education or female empowerment. Between treatment arms, households receiving large transfers reported higher value of non-land assets (279.18 USD PPP) and higher psychological well-being index (0.26) than households receiving small transfers, and these effects remained significant when using FWER-adjusted p-values.

### 3.3 Spillover effects: index variables

Table 3 reproduces the results reported in Table III of the original paper, the spillover effects on the eight index variables. The spillover effects were estimated by comparing the eight index variables between spillover households (control households in treatment villages) and pure control households (control households in control villages). The main treatment effects reported in Table 2 are unbiased if there were no spillover effects. The spillover effects were estimated by running models using index variables as outcome variables and a spillover household indicator as an explanatory variable, clustering standard errors at the village level.

**Table 3: Spillover effects of index variables, a reproduction of Table III in the original paper**

| | Spillover effects | | | | | | Lee Bounds | | Horowitz–Manski Bounds | |
|---|---|---|---|---|---|---|---|---|---|---|
| | All HH estimate | All HH estimate | Thatched estimate | Thatched estimate | Test (1)=(3) p-value | Test (2)=(4) p-value | Lower | Upper | Lower | Upper |
| Includes controls | No | Yes | No | Yes | No | Yes | No | No | No | No |
| Value of non-land assets (USD) | 1.00 | –11.99 | –18.73 | –32.61 | 0.01*** | 0.00*** | –3.38 | 12.84 | –2.38 | 7.39 |
| | (21.44) | (19.98) | (21.14) | (19.76) | | | (20.14) | (20.64) | (19.92) | (20.06) |
| Non-durable expenditure (USD) | –7.77 | –11.89* | –7.31 | –12.21* | 0.82 | 0.88 | –9.47 | –4.08 | –8.93 | –5.86 |
| | (7.20) | (6.50) | (7.27) | (6.67) | | | (5.84) | (5.86) | (5.79) | (5.83) |
| Total revenue, monthly (USD) | –3.68 | –3.64 | –5.23 | –5.78 | 0.56 | 0.45 | –4.29 | 2.32 | –4.18 | –1.91 |
| | (6.18) | (6.35) | (5.84) | (6.01) | | | (6.25) | (6.41) | (6.18) | (6.22) |
| Food security index | 0.06 | 0.05 | 0.06 | 0.05 | 0.93 | 0.90 | –0.01 | 0.08 | 0.03 | 0.07 |
| | (0.09) | (0.09) | (0.10) | (0.10) | | | (0.08) | (0.08) | (0.08) | (0.08) |
| Health index | –0.06 | –0.07 | –0.06 | –0.08 | 0.80 | 0.66 | –0.10 | –0.03 | –0.07 | –0.04 |
| | (0.08) | (0.08) | (0.08) | (0.08) | | | (0.07) | (0.07) | (0.07) | (0.07) |
| Education index | 0.01 | –0.01 | 0.00 | –0.03 | 0.36 | 0.29 | 0.10 | 0.10 | –0.01 | 0.03 |
| | (0.07) | (0.06) | (0.08) | (0.07) | | | (0.09) | (0.08) | (0.07) | (0.07) |
| Psychological well-being index | 0.03 | 0.03 | 0.03 | 0.02 | 0.77 | 0.71 | 0.03 | 0.04 | 0.01 | 0.05 |
| | (0.07) | (0.07) | (0.07) | (0.07) | | | (0.06) | (0.06) | (0.05) | (0.05) |
| Female empowerment index | 0.21** | 0.21** | 0.23** | 0.22** | 0.50 | 0.50 | 0.20** | 0.28*** | 0.18** | 0.23*** |
| | (0.09) | (0.08) | (0.09) | (0.09) | | | (0.09) | (0.11) | (0.08) | (0.08) |
| Joint test (p-value) | 0.38 | 0.25 | 0.23 | 0.11 | | | | | | |

Note: Ordinary least squares estimates of spillover effects. Outcome variables are listed on the left and described in detail in Table A1. The unit of observation is the household for all variables except psychological well-being, where it is the individual. The sample includes all households and individuals, except for the intrahousehold index, where it is restricted to cohabitating couples, and for the education index, where it is restricted to households with school-age children. Columns (1) and (2) report the “naïve” estimate of spillover effects, including spillover households that upgraded to metal roofs between baseline and endline. Columns (3) and (4) report estimates of the spillover effect excluding metal-roof households. Columns (1) and (3) exclude baseline covariates. Columns (2) and (4) include baseline covariates. Column (5) reports the p-value of the equality for the coefficient estimates in (1) and (3) after joint estimation of the two models using SUR. Column (6) reports the p-value of the equality of the coefficient estimates in (2) and (4) after joint estimation of the two models using SUR. The last row reports p-values on the joint significance of all coefficients in a given column after joint estimation using SUR. Columns (7) and (8) report the lower and upper Lee bounds adjusting for differential “attrition” of five spillover households into metal-roof upgrade. Columns (9) and (10) report lower and upper Horowitz–Manski bounds, imputing outcomes for the five attritting households using the 95th and 5th percentiles of observed outcomes, respectively. The sample is restricted, as in the previous tables. In columns (1) through (4), standard errors clustered at the village level are reported in parentheses. In columns (7) through (10), bootstrapped standard errors are reported in parentheses. * denotes significance at 10 percent, ** at 5 percent and *** at 1 percent level.

All coefficients, standard errors and p-values reported in columns 1–6 are reproduced exactly in Table 3. We were also able to produce an exact match for columns 9 and 10. In columns 7 and 8, there are some minor differences in the bootstrapped standard errors of Lee bounds for value of non-land assets (USD) (upper and lower); non-durable expenditure (USD) (upper and lower); total revenue, monthly (USD) (upper and lower); food security index (lower); education index (lower); and female empower index (lower). We include Table III from the original paper as Appendix 4 for comparison. The cells for standard errors that differ from those reported in the original table are shaded in Table 3. However, these differences are minor and do not change the conclusions of the original study.

The replication results for spillover effects are consistent with those reported in the original paper. No significant spillover effects were found in seven of the eight index variables examined. The only statistically significant difference is that spillover households reported a higher level of the female empowerment index (0.21) than pure control households. The spillover effects on female empowerment index did not change when restricting the comparison in thatched roof households. The Lee bounds and Horowitz–Manski bounds for spillover effects on female empowerment index were significant. SUR results did not support spillover effects at the 0.05 significance level.

The examination of spillover effects in these index variables suggested that a positive spillover effect existed in the female empowerment index but not in the other seven index variables. These results support the use of the spillover households as the control group when examining the treatment effects, except that the treatment effect on the female empowerment index might be underestimated.

### 3.4 Treatment effects: psychological well-being

In the original paper, treatment effects on psychological well-being were examined not only by looking at the psychological well-being index, but also by looking at the individual psychological well-being measures. The study collected a rich set of psychological well-being measures, including cortisol levels, depression (Center for Epidemiology Studies – Depression Scale [CES-D]), worries, stress (measured by the four-item version of Cohen's Stress Scale), happiness, life satisfaction, trust, locus of control (Rotter's Locus of Control scale), optimism (revised Scheier and Cerver's Life Orientation Test) and self-esteem (Rosenberg's self-esteem scale). Cortisol level was used as an indicator of both acute stress and more permanent stress-related conditions. The authors used two measures of cortisol levels: (1) log-transformed value of the average of two raw cortisol levels and (2) residuals of an ordinal linear regression of the log-transformed cortisol levels on several confounding factors.

We examined the overall treatment effects of UCT on these psychological well-being measures and the differential treatment effects of treatment arms, and report the results in Table 4. The treatment effects were estimated by running regression models on treatment status controlling for village-level fixed effects, baseline outcomes, an indicator variable for missing baseline outcomes and cluster standard errors at the household level. Column 1 in Table 4 reports the mean and standard errors of all psychological well-being measures in spillover households. Column 2 reports the treatment effects. Column 3 reports treatment effect differences between female recipient households and male recipient households. Column 4 reports treatment effect differences between monthly transfer households and lump-sum transfer households. Column 5 reports treatment effect differences between large-transfer households and small-transfer households.

Our Table 4 exactly reproduces Table IV in the original paper. Our replication results support the original authors' conclusion. These results suggest that the UCT improved some aspects of psychological well-being, but not others. Specifically, UCT reduced depression (–1.16), worries (–0.13) and stress (–0.26) and increased happiness (0.16) and life satisfaction (0.17) at a significance level of 0.01. However, no treatment effects were found on trust, locus of control, optimism, self-esteem or cortisol levels (an objective measure of stress levels) at a significance level of 0.05. SUR results supported the overall treatment effects ($p = 0.00$).

Compared to participants from male recipient households, participants from female recipient households had lower levels of cortisol (–0.17) and higher self-esteem (0.19), at a significance level of 0.05. Compared to lump-sum transfer recipients, monthly transfer recipients seemed to have a higher level of stress, as measured by log cortisol level (0.17), at a significance level of 0.05. Compared to participants receiving small transfers, large-transfer recipients reported lower stress levels (–0.24) and higher life satisfaction (0.19), at a significance level of 0.05. SUR results supported the differential treatment effects between large-transfer and small-transfer households ($p = 0.00$).

**Table 4: Treatment effects on psychological well-being, a reproduction of Table IV in the original paper**

| | Control mean (SD) | Treatment effect | Female recipient | Monthly transfer | Large transfer | N |
|---|---|---|---|---|---|---|
| Log cortisol (no controls) | 2.46 | 0.00 | –0.17** | 0.16* | –0.09 | 1456 |
| | (0.89) | (0.05) | (0.07) | (0.08) | (0.07) | |
| Log cortisol (with controls) | –0.04 | 0.01 | –0.17** | 0.17** | –0.12* | 1456 |
| | (0.88) | (0.05) | (0.07) | (0.08) | (0.07) | |
| Depression (CES-D) | 26.48 | –1.16*** | –0.77 | –1.40* | –1.22* | 1474 |
| | (9.31) | (0.44) | (0.67) | (0.73) | (0.68) | |
| Worries | 0.00 | –0.13*** | –0.04 | –0.11 | –0.07 | 1474 |
| | (1.00) | (0.05) | (0.07) | (0.08) | (0.08) | |
| Stress (Cohen) | 0.00 | –0.26*** | –0.02 | –0.02 | –0.24*** | 1474 |
| | (1.00) | (0.05) | (0.08) | (0.09) | (0.08) | |
| Happiness (WVS) | 0.00 | 0.16*** | 0.07 | 0.03 | 0.07 | 1474 |
| | (1.00) | (0.05) | (0.08) | (0.09) | (0.08) | |
| Life satisfaction (WVS) | 0.00 | 0.17*** | –0.07 | 0.12 | 0.19** | 1474 |
| | (1.00) | (0.05) | (0.07) | (0.08) | (0.08) | |
| Trust (WVS) | 0.00 | 0.04 | 0.08 | –0.08 | –0.04 | 1474 |
| | (1.00) | (0.05) | (0.08) | (0.08) | (0.08) | |
| Locus of control | 0.00 | 0.03 | 0.04 | –0.03 | 0.08 | 1474 |
| | (1.00) | (0.05) | (0.08) | (0.09) | (0.08) | |
| Optimism (Scheier) | 0.00 | 0.10* | 0.07 | 0.02 | 0.16* | 1474 |
| | (1.00) | (0.05) | (0.08) | (0.09) | (0.09) | |
| Self-esteem (Rosenberg) | 0.00 | 0.00 | 0.19** | 0.09 | –0.15 | 1474 |
| | (1.00) | (0.05) | (0.09) | (0.09) | (0.10) | |
| Psychological well-being index | 0.00 | 0.26*** | 0.14* | 0.01 | 0.26*** | 1474 |
| | (1.00) | (0.05) | (0.08) | (0.08) | (0.08) | |
| Joint test (p-value) | | 0.00*** | 0.21 | 0.21 | 0.00*** | |

Note: Ordinary least squares estimates of treatment and spillover effects. Outcome variables are listed on the left and described in detail in Table A1. All variables are coded in z-score units, except raw cortisol, which is coded in nmol/l, and depression, which is coded in points. For each outcome variable, we report the coefficients of interest and their standard errors in parentheses. Column (1) reports the mean and standard deviation of the control group for a given outcome variable. Column (2) reports the basic treatment effect, i.e. comparing treatment households to control households within villages. Column (3) reports the relative treatment effect of transferring to the female compared to the male; column (4) the relative effect of monthly compared to lump-sum transfers; and column (5) that of large compared to small transfers. The unit of observation is the individual. The comparison of monthly to lump-sum transfers excludes large-transfer recipient households, and that for male versus female recipients excludes single-headed households. All columns include village-level fixed effects, control for baseline outcomes and cluster standard errors at the household level. The last row shows joint significance of the coefficients in the corresponding column from SUR estimation. * denotes significance at 10 percent, ** at 5 percent and *** at 1 percent level.

### 3.5 Treatment effects: consumption

Table 5 shows detailed results for different categories of consumption, including total food expenditure, alcohol expenditure, tobacco expenditure, social expenditure, medical expenditure and education expenditure. The treatment effects were estimated by running regression models on treatment status, controlling for village-level fixed effects, baseline outcomes and an indicator variable for missing baseline outcomes. Column 1 shows mean and standard errors in the spillover households. Columns 2–5 show overall treatment effects and treatment effect differences between treatment arms.

**Table 5: Treatment effects on consumption, a reproduction of Table V in the original paper**

| | Control mean (SD) | Treatment effect | Female recipient | Monthly transfer | Large transfer | N |
|---|---|---|---|---|---|---|
| Food total (USD) | 104.46 | 19.46*** | –1.81 | 1.79 | 8.28 | 940 |
| | (58.50) | (4.19) | (7.37) | (7.42) | (7.59) | |
| Cereals (USD) | 22.55 | 2.23** | 0.37 | –1.06 | 2.68 | 940 |
| | (17.18) | (1.13) | (1.87) | (1.86) | (2.07) | |
| Meat and fish (USD) | 12.97 | 5.05*** | 0.87 | –2.93 | 2.52 | 940 |
| | (13.75) | (1.01) | (1.82) | (1.92) | (1.63) | |
| Alcohol (USD) | 6.38 | –0.93 | 1.56 | 1.03 | –1.42 | 940 |
| | (16.56) | (0.99) | (1.62) | (1.64) | (1.33) | |
| Tobacco (USD) | 1.52 | –0.15 | 0.12 | 0.42 | –0.29 | 940 |
| | (4.13) | (0.22) | (0.34) | (0.33) | (0.30) | |
| Social expenditure (USD) | 4.36 | 2.43*** | –2.06** | –0.52 | 0.62 | 940 |
| | (5.38) | (0.48) | (0.97) | (0.99) | (0.90) | |
| Medical expenditure past month (USD) | 6.78 | 2.58*** | 2.06 | –1.34 | –0.29 | 940 |
| | (13.53) | (0.99) | (1.86) | (1.86) | (1.74) | |
| Education expenditure (USD) | 4.71 | 1.08** | 0.48 | –0.02 | 1.15 | 940 |
| | (8.68) | (0.51) | (0.88) | (0.87) | (0.91) | |
| Non-durable expenditure (USD) | 157.61 | 35.66*** | –2.00 | –4.20 | 21.25** | 940 |
| | (82.18) | (5.85) | (10.28) | (10.71) | (10.49) | |
| Joint test (p-value) | | 0.00*** | 0.47 | 0.13 | 0.01*** | |

Note: Ordinary least squares estimates of treatment and spillover effects. Outcome variables are listed on the left and described in detail in Table A1. All variables are reported in PPP-adjusted USD and are top-coded for the highest 1 percent of observations. For each outcome variable, we report the coefficients of interest and their standard errors in parentheses. Column (1) reports the mean and standard deviation of the control group for a given outcome variable. Column (2) reports the basic treatment effect, i.e. comparing treatment households to control households within villages. Column (3) reports the relative treatment effect of transferring to the female compared to the male; column (4) the relative effect of monthly compared to lump-sum transfers; and column (5) that of large compared to small transfers. The unit of observation is the household. The comparison of monthly to lump-sum transfers excludes large-transfer recipient households, and that for male versus female recipients excludes single-headed households. All columns include village-level fixed effects and control for baseline outcomes. The last row shows joint significance of the coefficients in the corresponding column from SUR estimation. * denotes significance at 10 percent, ** at 5 percent and *** at 1 percent level.

Our Table 5 exactly reproduces Table V in original paper. Our results support the authors' findings regarding the impact of UCT on consumption. Overall, the UCT increased expenditure on food (19.46 USD PPP), cereals (2.23 USD PPP) and meat and fish (5.05 USD PPP) in particular, as well as social expenditure (2.43 USD PPP), medical expenditure (2.58 USD PPP) and education expenditure (1.08 USD PPP), but not expenditure on temptation goods, such as alcohol and tobacco. SUR results supported the existence of overall treatment effects (p = 0.00).

Compared to male recipient households, female recipient households spent less on social activities (–2.06 USD PPP). No differences were found between monthly and lump-sum transfer households. Compared to small-transfer households, large-transfer households had higher total non-durable expenditure, but no other significant differences were found in the expenditure categories examined. SUR results supported the differential treatment effects between large-transfer and small-transfer households (p = 0.01).

### 3.6 Treatment effects: assets and agricultural and business activities

Table 6 shows detailed results for assets and agricultural and business activities. Specific asset measures included value of livestock, value of durable goods, value of savings, land owned and a dummy variable indicating whether the household had a non-thatched roof. Specific business activities the original authors accounted for in their analyses included indicators of having wage labor as primary income, own-farm primary income, non-agricultural business ownership, total monthly revenue, total monthly expenses and total monthly profit. The treatment effects were estimated by running a regression model on treatment status controlling for village-level fixed effects, baseline outcomes and indicator variable for missing baseline outcomes. Column 1 shows the mean and standard deviations of the outcome variables in the spillover households. Columns 2–5 show the overall treatment effects and treatment effect differences between treatment arms.

Table 6 exactly matches Table VI from the original paper and supports the authors' findings regarding the impacts of UCT on asset holdings and business activities. Overall, the UCT increased the recipient households' value of livestock (83.18 USD PPP), value of durable goods (52.99 USD PPP), savings (10.10 USD PPP) and likelihood of having a better-quality (non-thatched) roof (0.24). SUR results supported the overall treatment effects on assets outcomes (p = 0.00).

Compared to male recipient households, female recipient households were less likely to have non-thatched roofs (–0.11). Compared to lump-sum transfer households, monthly transfer households were less likely to have non-thatched roofs (–0.12). Compared to small-transfer households, large-transfer households had a higher value of livestock (63.45 USD PPP), value of durable goods (64.90 USD PPP) and savings (10.26 USD PPP) and were more likely to have a non-thatched roof (0.23). SUR results supported the group differences in treatment effects (p = 0.00).

Households that received the UCT had higher total monthly revenue (16.15 USD PPP) and total monthly expenses (12.53 USD PPP). No difference was reported on monthly profit, the likelihood of having wage labor primary income, having own-farm primary income or being a non-agricultural business owner. SUR results supported treatment effects on outcomes related to business activities (p = 0.00). No significant differences were found between treatment arms in the variables related to business activities.

**Table 6: Treatment effects: assets and agricultural and business activities, a reproduction of Table VI in the original paper**

| | Control mean (SD) | Treatment effect | Female recipient | Monthly transfer | Large transfer | N |
|---|---|---|---|---|---|---|
| **Panel A: Assets** | | | | | | |
| Value of non-land assets (USD) | 494.80 | 301.51*** | –79.46 | –91.85** | 279.18*** | 940 |
| | (415.32) | (27.25) | (50.38) | (45.92) | (49.09) | |
| Value of livestock (USD) | 166.82 | 83.18*** | 4.84 | 0.08 | 63.45** | 940 |
| | (240.59) | (15.22) | (29.32) | (27.36) | (28.51) | |
| Value of durable goods (USD) | 207.30 | 52.59*** | –0.24 | –7.31 | 64.90*** | 940 |
| | (130.60) | (8.61) | (14.40) | (14.16) | (15.70) | |
| Value of savings (USD) | 10.93 | 10.10*** | –3.31 | 1.86 | 10.26** | 940 |
| | (29.09) | (2.46) | (5.03) | (4.57) | (5.04) | |
| Land owned (acres) | 1.31 | 0.04 | –0.08 | 0.04 | 0.35 | 940 |
| | (1.88) | (0.14) | (0.18) | (0.17) | (0.32) | |
| Has non-thatched roof (dummy) | 0.16 | 0.24*** | –0.11** | –0.12** | 0.23*** | 940 |
| | (0.37) | (0.03) | (0.05) | (0.05) | (0.05) | |
| Joint test (p-value) | | 0.00*** | 0.29 | 0.22 | 0.00*** | |
| **Panel B: Business activities** | | | | | | |
| Wage labor primary income (dummy) | 0.16 | 0.00 | 0.02 | 0.02 | 0.01 | 940 |
| | (0.37) | (0.02) | (0.04) | (0.04) | (0.04) | |
| Own farm primary income (dummy) | 0.56 | –0.01 | 0.00 | 0.00 | 0.01 | 940 |
| | (0.50) | (0.03) | (0.05) | (0.05) | (0.05) | |
| Non-agricultural business owner (dummy) | 0.32 | 0.01 | –0.02 | 0.08 | 0.01 | 940 |
| | (0.47) | (0.03) | (0.05) | (0.05) | (0.05) | |
| Total revenue, monthly (USD) | 48.98 | 16.15*** | 5.41 | 16.33 | –2.44 | 940 |
| | (90.52) | (5.88) | (10.61) | (11.07) | (8.87) | |
| Total expenses, monthly (USD) | 23.95 | 12.53*** | 5.42 | 9.41 | –0.35 | 940 |
| | (61.71) | (4.21) | (7.45) | (7.75) | (6.23) | |
| Total profit, monthly (USD) | 20.78 | –0.21 | 1.41 | 7.29 | –2.02 | 940 |
| | (46.22) | (3.68) | (6.68) | (7.92) | (5.32) | |
| Joint test (p-value) | | 0.00*** | 0.90 | 0.59 | 1.00 | |

Note: Ordinary least squares estimates of treatment and spillover effects. Outcome variables are listed on the left and described in detail in Table A1. Variables are in PPP-adjusted 2012 USD and are top-coded for the highest 1 percent of observations. For each outcome variable, we report the coefficients of interest and their standard errors in parentheses. Column (1) reports the mean and standard deviation of the control group for a given outcome variable. Column (2) reports the basic treatment effect, i.e. comparing treatment households to control households within villages. Column (3) reports the relative treatment effect of transferring to the female compared to the male; column (4) the relative effect of monthly compared to lump-sum transfers; and column (5) that of large compared to small transfers. The unit of observation is the household. The comparison of monthly to lump-sum transfers excludes large-transfer recipient households and that for male versus female recipients excludes single-headed households. All columns except the spillover regressions include village-level fixed effects and control for baseline outcomes. The last row shows joint significance of the coefficients in the corresponding column from SUR estimation. * denotes significance at 10 percent, ** at 5 percent and *** at 1 percent level.

# 4. Measurement and estimation analysis results

Results from the measurement and estimation analysis include findings from analysis on (1) model validation, (2) alternative model specification and (3) alternative measures of health, food security and psychological well-being indices.

## 4.1 Model validation

For model validation, we examined the normal distribution and equal variance assumptions of the models for the main treatment effects of the UCT, examined the autocorrelation and variability of the eight outcome indices to see whether DID models should be used for analysis, conducted robustness checks on how missing values were treated for the education index, psychological well-being index and female empowerment index, and conducted multilevel modeling to consider both household- and village-level correlations when examining the treatment effects on psychological well-being index.

### *4.1.1 Normal distribution and equal variance assumption*

We examined normal distribution and equal variance assumptions by graphing the histogram plots and residual plots of the eight outcome indices. In general, the histogram plots of the residuals of the eight outcome indices showed that the distributions were close to the bell-shaped symmetric normal distribution, with slight skewness, except for revenue. The histogram plot for revenue showed a relatively heavy tail to the right (kurtosis = 14.76, skewness = 3.56). The plots of residuals to the outcome indices showed that the values of residuals varied according to the values of the outcome indices.

Potentially, data with positive values with right-skewed distribution can be log transformed to bring them closer to normal distribution and equal variance distribution (Metcalf 1972; Kutner et al. 2004). We conducted log transformation on total revenue and ran the regression analysis using log-transformed total revenue. Regression results showed that UCT had a positive impact on log revenue (coefficient = 0.27, standard error = 0.09, $p < 0.01$) after controlling for village-level fixed effects and log baseline revenue. The regression results supported the original study's conclusion that UCT had a positive impact on total revenue.

Figure A1 in the Appendix of this paper shows the plots for total revenue and log-transformed revenue. The figure shows a more symmetric distribution in the histogram plot and more spread-out distribution of residuals for log total revenue, as compared to total revenue. The kurtosis/skewness statistics of the log revenue are 1.24/–0.28, closer to zero than those of untransformed revenue (kurtosis = 14.76, skewness = 3.56). Regardless, the estimation based on the log-transformed outcome variable did not change the original study's conclusion as to the positive treatment effect on revenue.

### *4.1.2 Autocorrelation and variability of outcome variables and DID models*

We examined the autocorrelation and variability of the outcome measures of interest. We hypothesized that economic outcome measures such as value of non-land assets, non-durable expenditure and total revenue would have a low autocorrelation between baseline and endline measures, while health-related measures such as health index and psychological well-being index would have relatively large autocorrelation.

Table 7 reports correlations between baseline and endline outcome measures using (1) all study sample data and (2) control group–only data to avoid the potential treatment effects. As the table shows, the results were similar using the two samples. The autocorrelation of the eight indices was relatively small. Of the eight indices examined, only the education index had an autocorrelation coefficient greater than 0.5.

**Table 7: Correlation coefficients between baseline and endline outcome measures**

| Variable | Control group | | | All study sample | | |
|---|---|---|---|---|---|---|
| | N | Correlation | P-value | N | Correlation | P-value |
| Value of non-land assets (USD) | 505 | 0.44 | < .01 | 1,440 | 0.39 | < .01 |
| Non-durable expenditure (USD) | 505 | 0.22 | < .01 | 1,440 | 0.13 | < .01 |
| Total revenue, monthly (USD) | 505 | 0.36 | < .01 | 1,440 | 0.22 | < .01 |
| Food security index | 469 | 0.19 | < .01 | 940 | 0.18 | < .01 |
| Health index | 469 | 0.08 | 0.08 | 940 | 0.10 | < .01 |
| Education index | 375 | 0.56 | < .01 | 769 | 0.52 | < .01 |
| Psychological well-being index* | 656 | 0.19 | < .01 | 1,313 | 0.18 | < .01 |
| Female empowerment index | 309 | 0.29 | < .01 | 625 | 0.37 | < .01 |

Note: * Individual observations were used for the psychological well-being index.

We fitted DID models on the education index, as its autocorrelation coefficient was greater than 0.5. The comparison between DID model estimates and ANCOVA model estimates in the original paper helped us understand the robustness of the treatment effects on the education index. We fitted DID models using (1) difference in education index for households with non-missing baseline data, and (2) difference in education index with imputed missing baseline measures, controlling for a dummy variable indicating missing baseline education index (as in the original paper). Table 8 shows the regression results for these two DID models.

The results from the DID model using complete records showed that there were no treatment effects on the education index (coefficient = 0.12, $p > 0.05$). The results did not change when using imputed data on the education index (coefficient = 0.11, $p = 0.10$). The DID model results were not different from those reported in the original paper (coefficient = 0.08, $p = 0.17$), which showed that UCT did not have any impact on the education index.

**Table 8: Regression results for DID model on education index**

| Variable | Difference in education index (n = 769) | | | Difference in education index with imputed baseline measures (n = 823) | | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | P | Coefficient | SE | P |
| Intercept | –0.23 | 0.30 | 0.45 | –0.11 | 0.29 | < .01 |
| Treatment | 0.12 | 0.06 | 0.08 | 0.11 | 0.07 | 0.10 |
| Baseline education index missing | | | | 0.69 | 0.20 | < .01 |

Note: * Village-level fixed effects were controlled, but results are not reported in the table.

#### *4.1.3 Missing values*

For three indices that had missing values (education index, psychological well-being index and female empowerment index), the original authors used the missing-indicator method by adding a dummy variable, indicating whether the baseline index measure was missing, and an interaction term of the baseline index measure and the dummy variable for missing in the model. Jones (1996) suggests that this method may produce biased estimates when missing is not completely at random and may have no advantage over the complete-case analysis.

We reran the regression for the education index, psychological well-being index and female empowerment index using only observations with complete information on all variables. Table 9 shows the results for complete-case analysis. Using observations without missing values, the regression results showed positive effects on the psychological well-being index (coefficient = 0.28, $p < 0.01$), but no treatment effects on the education index (coefficient = 0.08, $p = 0.15$) or female empowerment index (coefficient = –0.01, $p = 0.86$). Although the coefficient estimates and p-values differed slightly, the main results for education index, psychological well-being index and female empowerment index, using full observations, were the same as those reported in the original study.

**Table 9: Regression results for education index, psychological well-being index and female empowerment index using full observations**

| Variable | Education index (n = 769) | | | Psychological well-being Index (n = 1,313) | | | Female empowerment index (n = 625) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Coefficient | SE | P | Coefficient | SE | P | Coefficient | SE | P |
| Intercept | –0.21 | 0.27 | 0.44 | 0.51 | 0.17 | < .01 | 0.39 | 0.29 | 0.18 |
| Treatment | 0.08 | 0.08 | 0.15 | 0.28 | 0.05 | < .01 | –0.01 | 0.07 | 0.86 |
| Baseline index | 0.55 | 0.04 | < .01 | 0.19 | 0.03 | < .01 | 0.38 | 0.04 | < .01 |

Note: * Village-level fixed effects were controlled, but results are not reported in the table.

#### *4.1.4 Multilevel modeling*

Psychological well-being index information was collected from the primary members of a household – husband and wife, for a two-headed household – from all villages participating in the study. People from the same households, and similarly from the same villages, may report similar psychological well-being, compared to people from different households or villages. We ran a multilevel analysis on the psychological well-being index to account for the potential correlations at both household and village levels when examining the treatment effects. Table 10 shows the results of the multilevel analysis.

**Table 10: Multilevel analysis results for psychological well-being index**

| Variable | Psychological well-being index (n = 2,140) | | |
|---|---|---|---|
| | Coefficient | SE | P-value |
| Intercept | –0.04 | 0.09 | 0.65 |
| Treatment | 0.25 | 0.05 | < .01 |
| Baseline psychological well-being index | 0.19 | 0.03 | < .01 |
| Baseline psychological well-being index missing | 0.04 | 0.08 | 0.60 |

The multilevel analysis results indicated correlation at the village level ($p < 0.01$) but not at the household-level ($p = 0.12$). This result suggests that there were no additional household-level correlations, beyond the correlations at the village level, in the psychological well-being index. The multilevel analysis results suggest that UCT had a positive impact on the psychological well-being index (coefficient = 0.25, $p < 0.01$), similar to the results reported in the original study (coefficient = 0.26, $p < 0.01$).

### 4.2 Modeling interaction effects

We added the treatment village and treatment gender interaction terms in the regression models, examining the treatment effects of all eight indices. The F-test results suggested significant effects of village treatment interaction for the eight indices, but not on gender treatment interaction. Table 11 shows the regression results, considering treatment village interaction and treatment gender interaction. The main treatment effects remained significant for assets, expenditure and revenue and remained insignificant for health index, education index and female empowerment index. The main treatment effects on the food security index and psychological well-being index were not significant after controlling for treatment village interactions, suggesting that the treatment effects may vary by village.

Contextual analysis could be conducted to determine whether differences between villages participating in the study were substantial, which might have had an impact on the effects of the UCT. The insignificant findings related to the interaction terms need to be interpreted with caution, as cell size may be too small to provide sufficient power for analysis.

**Table 11: Regression results for the eight indices, controlling for village effects, village treatment interaction and gender treatment interaction**

| | Value of non-land assets (USD) | | Non-durable expenditure (USD) | | Total revenue, monthly (USD) | | Food security index | | Health index | | Education index | | Psychological well-being index | | Female empowerment index | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Coeff.** | **P** | **Coeff.** | **P** | **Coeff.** | **P** | **Coeff.** | **P** | **Coeff.** | **P** | **Coeff.** | **P** | **Coeff.** | **P** | **Coeff.** | **P** |
| Intercept | 122.62 | 0.12 | 108.84 | < .01 | 42.65 | 0.01 | –0.26 | 0.20 | 0.04 | 0.93 | –1.02 | 0.00 | 0.65 | 0.01 | –0.22 | 0.68 |
| Treatment | 582.52 | 0.01 | 129.85 | < .01 | 95.09 | 0.03 | 0.16 | 0.67 | –0.05 | 0.92 | –0.05 | 0.92 | –0.11 | 0.73 | 0.55 | 0.31 |
| Baseline index measure | 0.48 | < .01 | 0.20 | < .01 | 0.08 | <.01 | 0.1865 | < .01 | 0.09 | 0.02 | 0.53 | < .01 | 0.20 | < .01 | 0.37 | < .01 |
| Baseline index missing | | | | | | | | | | | 0.68 | < .01 | 0.05 | 0.57 | –0.05 | 0.71 |
| Treatment gender interaction | 45.26 | 0.32 | –5.62 | 0.55 | –7.50 | 0.43 | 0.08 | 0.38 | –0.04 | 0.67 | –0.01 | 0.90 | –0.05 | 0.47 | –0.18 | 0.09 |

Note: * Village-level fixed effects and treatment village interaction were controlled in the model, but coefficient estimates are not reported for the large number of interaction terms.

## 4.3 Index measures of health, food security and psychological well-being

We conducted PCA based on variables used to create the food security index, health index and psychological well-being index. Our PCA differs from the weighted standardized approach used by the original authors, following Anderson (2008). The underlying assumption of using a weighted mean of several standardized variables to create a summary index is that all variables used to create the index describe the same concept or factor. PCA transforms the data of correlated variables to linearly uncorrelated variables or orthogonal vectors and determines the number of factors retained based on the size of their eigenvalues (Johnson and Wichern 2007; Vyas and Kumaranayake 2006).

PCA allows us to examine whether all variables used to create the index describe more than one latent variable (factor) and how they contribute to one factor differentially (through factor loading). If all variables used to create the index are loaded into one factor, then we can create a new index, using weights based on the factor loadings, for comparison. If more factors are obtained through PCA analysis, it questions the practice of creating one index based on all variables, using the weighted standardized approach, and suggests that the variables are describing two or more relevant but different concepts or factors.

### *4.3.1 Food security index*

In the original paper, the food security index was created based on 17 items, of which 12 were continuous variables: number of times in the last month that household adults skipped meals, children skipped meals, adults went whole days without food, children went whole days without food; number of times in the preceding month household members had to eat cheaper or less preferred food, had to rely on others for food, had to purchase food on credit, had to hunt for or gather food, had to beg for food; number of times household members ate meat or fish in the preceding week; proportion of household members who ate protein in the last 24 hours; and proportion of children who ate protein in the last 24 hours.

Five indicator variables were not incorporated in the PCA analysis: whether the respondent went to sleep hungry in the preceding week, whether household members usually ate at least two meals per day, usually ate until content, had enough food in the house for the next day, and whether the respondent had eaten protein in last 24 hours.

Table 12 reports PCA results on 12 continuous food security items. The first row (eigenvalue) shows the amount of variance in the data that was explained by each factor. Using eigenvalues of 1 or greater as criteria for explaining an adequate amount of the data variance, four factors were retained for further analysis with factor patterns (Vyas and Kumaranayake 2006; Johnson and Wichern 2007). The variance explained by each factor ranged from 3.10 to 1.01. These four factors explained 63 percent of the variation.

In Table 12, the factor loading section shows the unique variance each factor contributed to the variance of an observed variable. Ideally, if one variable has a moderately high to high loading in only one factor, it is easily interpretable. For the 12 items related to food security, although the first nine items loaded moderately high to high on factor 1 (0.43 to

0.66), some of these items also loaded moderately high to high on factors 3 and 4. The last three food security items loaded moderately high to high on factor 2 (0.60 to 0.87) and moderately high on factor 1 (–0.27 to –0.38).

Overall, considering the eigenvalue criteria and factor loading patterns together, the PCA results on the 12 food security items suggested that these items represented more than one underlying factor. Additional analyses are needed to explore the factors and determine whether each factor describes a particular dimension in food security that warrants composing four indices out of the 12 food security items.

**Table 12: Factor pattern on 12 food security items**

| | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| Eigenvalue | 3.10 | 2.20 | 1.28 | 1.01 |
| Proportion of variation | 0.26 | 0.18 | 0.11 | 0.08 |
| *Factor loading* | | | | |
| Meals skipped (adults, \# last month) | 0.66 | 0.27 | –0.31 | –0.40 |
| Meals skipped (children, \# last month) | 0.65 | 0.27 | –0.13 | –0.41 |
| Whole days without food (adults, \# last month) | 0.63 | 0.06 | 0.46 | –0.12 |
| Whole days without food (children, \# last month) | 0.53 | 0.12 | 0.59 | –0.12 |
| Eat less preferred/cheaper foods (\# last month) | 0.46 | 0.17 | –0.48 | –0.12 |
| Rely on help from others for food (\# last month) | 0.50 | 0.22 | –0.24 | 0.44 |
| Purchase food on credit (\# last month) | 0.43 | 0.27 | –0.32 | 0.44 |
| Hunt, gather wild food, harvest prematurely (\# last month) | 0.53 | –0.01 | 0.12 | 0.14 |
| Beg because not enough food in the house (\# last month) | 0.53 | 0.10 | 0.36 | 0.45 |
| Number of times ate meat or fish (last week) | –0.27 | 0.60 | –0.13 | 0.17 |
| Proportion of HH who ate protein (last 24h) | –0.38 | 0.87 | 0.16 | –0.06 |
| Proportion of children who ate protein (last 24h) | –0.37 | 0.87 | 0.17 | –0.06 |

#### *4.3.2 Health index*

In the original paper, the health index was created as a standardized weighted average of eight items: proportion of household adults who were sick or injured in the last month, proportion of household children who were sick or injured in the last month, proportion of sick or injured family members for whom the household could afford treatment, proportion of illnesses for which a doctor was consulted, proportion of newborns who were vaccinated, proportion of children under age 14 who had received a health checkup in the preceding six months, proportion of children under 5 who had died in the preceding year, and a children's anthropometrics index (consisting of body mass index, height-for-age, weight-for-age and upper arm circumference, relative to World Health Organization development benchmarks). There were many missing values on the proportion of vaccinated newborns (1,083 of 1,440 observations were missing, or 75.2%) and the proportion of children under 5 who died in the preceding year (481 of 1,440 observations were missing, or 33.4%). These two variables were not included in the PCA due to the large proportion of missing values that could bias the results.

Table 13 reports PCA results on the other six health items. Using eigenvalues of 1 or greater as criteria for explaining an adequate amount of the data variance, three factors were retained for further analysis with factor patterns. The variance explained by each of the factors ranged from 2.27 to 1.14. These four factors explained a total of 85 percent of the variation. The factor loading section shows that the eight health items loaded moderately high to high on the four factors retained. Specifically, the first two items loaded high on factor 1 (0.85, 0.91). The next four items loaded moderately high to high on at least two of the three factors and did not show an exclusive pattern.

Considering the eigenvalue criteria and factor loading patterns together, the PCA results on the six health items suggested that these items represented more than one underlying factor. Additional analyses are needed to explore the factors and determine whether each factor describes a particular dimension in health that warrants composing three health indices out of the six items.

**Table 13: Factor pattern on six health items**

| | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Eigenvalue | 2.27 | 1.67 | 1.14 |
| Proportion of variation | 0.38 | 0.28 | 0.19 |
| *Factor loading* | | | |
| Proportion of household sick/injured (1 month) | 0.85 | 0.21 | 0.30 |
| Proportion of children sick/injured (1 month) | 0.91 | 0.14 | 0.24 |
| Proportion of sick/injured who could afford treatment | 0.09 | 0.67 | –0.53 |
| Proportion of illnesses where doctor was consulted | 0.13 | 0.71 | –0.45 |
| Proportion of children < 14 getting checkup (6 months) | –0.18 | 0.70 | 0.65 |
| Children health index | –0.81 | 0.41 | 0.30 |

#### *4.3.3 Psychological well-being index*

In the original paper, the psychological well-being index was created as a standardized weighted average of an individual's scores on six well-being items, the CES-D scale, a custom worries questionnaire, Cohen's stress scale, their response to the World Values Survey (WVS) happiness and life satisfaction questions, and their log cortisol levels adjusted for confounders. Table 14 reports PCA results on the six psychological well-being items. Using eigenvalues of 1 or greater as criteria for explaining an adequate amount of the data variance, two factors were retained for further analysis with factor patterns. The variances explained by factor 1 and factor 2 were 1.74 and 1.02, respectively. These two factors explained a total of 46 percent of the variation. The factor loading section shows that the six psychological well-being items loaded moderately high to high on the two factors retained. Specifically, worries (0.64, 0.42) and happiness (–0.45, 0.58) loaded moderately high on both factors.

Considering the eigenvalue criteria and factor loading patterns together, the PCA results on the six psychological well-being items suggested that these items represented more than one underlying factor. Additional analyses are needed to explore the factors and determine whether each factor describes a particular dimension in psychological well-being that warrants composing two factors out of the six psychological well-being items.

**Table 14: Factor pattern for psychological well-being items**

| | Factor 1 | Factor 2 |
|---|---|---|
| Eigenvalue | 1.74 | 1.02 |
| Proportion of variation | 0.29 | 0.17 |
| *Factor loading* | | |
| Depression (CES-D) | 0.68 | 0.09 |
| Worries | 0.64 | 0.42 |
| Stress (Cohen) | 0.61 | 0.24 |
| Happiness (WVS) | –0.45 | 0.58 |
| Life satisfaction (WVS) | –0.53 | 0.23 |
| Log cortisol (with controls) | 0.15 | –0.62 |

# 5. Theory of change analysis results

## 5.1 Multivariate analysis on eight indices

We examined the treatment effects on the indices using multivariate analysis, as the index measures in each domain were expected to be correlated. The use of multivariate regression analyses allowed tests for differential impacts of cash transfers on the economic and health outcomes, after considering the correlation.

For the multivariate analysis of the eight indices, we considered treatment effects at the household level and converted the individual psychological well-being index into a household average. The household average psychological well-being index score was equal to the individual psychological well-being index score if the index score was only available for one person in a household.

Table 15 shows results of the multivariate analysis on the eight indices. The multivariate analysis results showed that the UCT increased value of non-land assets by 301.28 USD PPP, non-durable expenditure by 35.41 USD PPP, total monthly revenue by 15.97 USD PPP, food security index by 0.26 and psychological well-being index by 0.20. The UCT was not significantly associated with health index, education index or female empowerment index. The joint test on the overall treatment effects on the eight indices returned a p-value smaller than 0.01, supporting the conclusion of the original paper that the UCT had a significant impact on the outcome variables.

**Table 15: Multivariate analysis results on eight indices**

| | Treatment | SE | T-statistics | P-value |
|---|---|---|---|---|
| Value of non-land assets (USD) | 301.28 | 27.45 | 10.98 | <.01 |
| Non-durable expenditure (USD) | 35.41 | 5.82 | 6.09 | <.01 |
| Total revenue, monthly (USD) | 15.97 | 5.82 | 2.74 | <.01 |
| Food security index | 0.26 | 0.06 | 4.12 | <.01 |
| Health index | –0.03 | 0.06 | –0.55 | 0.58 |
| Education index | 0.07 | 0.06 | 1.26 | 0.21 |
| Psychological well-being index | 0.20 | 0.07 | 3.01 | <.01 |
| Female empowerment index | –0.01 | 0.07 | –0.18 | 0.85 |
| Joint test (p-value) | | | | <.01 |

Note: Baseline measures and an indicator of missing baseline measures and village-level fixed effects were controlled in the models.

## 5.2 Multivariate analysis on individual measures of psychological well-being, consumption and assets

We examined the treatment effects on the individual outcome measures that composed the indices of assets, consumption and psychological well-being. As the individual measures in each domain were expected to be correlated, the use of multivariate regression analyses allowed tests for differential impacts of cash transfers on the individual outcomes in each of the four domains.

### *5.2.1 Psychological well-being*

Table 16 shows results of multivariate analysis on the 11 individual psychological well-being measures. The multivariate analysis results showed that the UCT reduced CES-D depression scale by 1.16, worries scale by 0.13 and Cohen stress scale by 0.26, and increased happiness scale by 0.16 and life satisfaction scale by 0.17. The UCT was not found to be significantly related to cortisol levels, trust, locus of control, optimism or self-esteem measures. These findings are consistent with those reported in the original paper. The joint test on the overall treatment effects on the 11 psychological well-being measures returned a p-value smaller than 0.01, supporting the conclusion of the original paper that the UCT had a significant impact on the psychological well-being measures.

**Table 16: Multivariate analysis results on psychological well-being measures**

| | Coefficient | SE | T-statistics | P-value |
|---|---|---|---|---|
| Log cortisol (no controls) | –0.00 | 0.05 | –0.01 | 0.99 |
| Log cortisol (with controls) | 0.00 | 0.05 | 0.06 | 0.96 |
| Depression (CES-D) | –1.16 | 0.45 | –2.6 | <.01 |
| Worries | –0.13 | 0.05 | –2.67 | <.01 |
| Stress (Cohen) | –0.26 | 0.05 | –5.03 | <.01 |
| Happiness (WVS) | 0.16 | 0.05 | 3.09 | <.01 |
| Life satisfaction (WVS) | 0.17 | 0.05 | 3.41 | <.01 |
| Trust (WVS) | 0.04 | 0.05 | 0.81 | 0.42 |
| Locus of control | 0.04 | 0.05 | 0.68 | 0.50 |
| Optimism (Scheier) | 0.10 | 0.05 | 1.87 | 0.06 |
| Self-esteem (Rosenberg) | 0.00 | 0.06 | 0.04 | 0.97 |
| Joint test (p-value) | | | | <.01 |

Note: Baseline measures and an indicator of missing baseline measures and village-level fixed effects were controlled in the models.

### *5.2.2 Consumption*

Table 17 shows results of multivariate analysis on the eight individual consumption measures. The multivariate analysis results showed that the UCT increased total food consumption by 19.10 USD PPP, spending on cereals by 2.21 USD PPP, spending on meat and fish by 5.12 USD PPP, social expenditure 2.43 USD PPP, medical expenditure in the past month by 2.58 USD PPP and education expenditure by 1.08 USD PPP. The UCT was not significantly related to expenditure on alcohol or tobacco. These findings are consistent with those reported in the original paper. The joint test on the overall treatment effects on the eight individual consumption measures returned a p-value smaller than 0.01, supporting the conclusion of the original paper that the UCT had a significant impact on the individual consumption measures.

**Table 17: Multivariate analysis results on consumption measures**

| | Coefficient | SE | T-statistics | P-value |
|---|---|---|---|---|
| Food total (USD) | 19.10 | 4.14 | 4.61 | <.01 |
| Cereals (USD) | 2.21 | 1.13 | 1.96 | 0.05 |
| Meat and fish (USD) | 5.12 | 1.00 | 5.09 | <.01 |
| Alcohol (USD) | –0.93 | 0.99 | –0.94 | 0.35 |
| Tobacco (USD) | –0.15 | 0.22 | –0.69 | 0.49 |
| Social expenditure (USD) | 2.43 | 0.48 | 5.03 | <.01 |
| Medical expenditure past month (USD) | 2.58 | 1.00 | 2.59 | 0.01 |
| Education expenditure (USD) | 1.05 | 0.52 | 2.00 | 0.05 |
| Joint test (p-value) | | | | <.01 |

Note: Baseline measures and an indicator of missing baseline measures and village-level fixed effects were controlled in the models.

#### *5.2.3 Asset holdings*

Table 18 shows results of multivariate analysis on the individual asset measures. The multivariate analysis results showed that the UCT increased value of livestock by 83.07 USD PPP, value of durable goods by 52.63 USD PPP, savings by 10.10 USD PPP and non-thatched roof by 0.24. The UCT was not significantly related to the amount of land owned. These findings are consistent with those reported in the original paper. The joint test on the overall treatment effects on the five asset measures returned a p-value smaller than 0.01, supporting the conclusion of the original paper that the UCT had a significant impact on the individual asset measures.

**Table 18: Multivariate analysis results on asset measures**

| | Coefficient | SE | t-statistics | p-value |
|---|---|---|---|---|
| Value of livestock (USD) | 83.07 | 15.39 | 5.4 | <.01 |
| Value of durable goods (USD) | 52.63 | 8.59 | 6.13 | <.01 |
| Value of savings (USD) | 10.10 | 2.45 | 4.12 | <.01 |
| Land owned (acres) | 0.04 | 0.13 | 0.31 | 0.76 |
| Has non-thatched roof (dummy) | 0.24 | 0.03 | 8.38 | <.01 |
| Joint test (p-value) | | | | <.01 |

Note: Baseline measures and an indicator of missing baseline measures and village-level fixed effects were controlled in the models.

## 6. Discussion

Our pure replication findings are consistent with the findings reported in the original study. The baseline comparison and spillover effects results support the original authors' use of the spillover group as the control group to examine the treatment effects of receiving a UCT in Kenya. The treatment effects on eight index variables suggest that providing UCT in Kenya had positive effects on asset holdings, consumption, monthly revenue, food security and psychological well-being. The treatment effects on index variables are supported, in general, by the more detailed analyses conducted on individual measures in psychological well-being, consumption, assets and business activities.

As reported in the original study, we find no consistent pattern on treatment effects between treatment arms in psychological well-being, consumption, assets and business activities. However, we confirm that monthly transfers were more likely than lump-sum transfers to improve food security, while lump-sum transfers were more likely to be spent on durables. Unsurprisingly, large cash transfers were more likely to increase asset holdings, increase consumption and improve psychological well-being than small cash transfers. Our study replicates these main findings reported in the original paper, validating the short-term impact of UCT in these different areas in Kenya.

Our measurement and estimation analysis also supports the statistical analysis and results presented in the original paper. The robustness checks on model validation include examination of normal and unequal variance assumptions, examination of the autocorrelation and variability of outcome variables, methods of handling missing values and application of multilevel modeling for the psychological well-being index.

The robustness checks on normal distribution and equal variance assumptions of the outcome variables support the adoption of the model specified in the original study. Although log transformation was recommended for the total revenue outcome due to its skewed distribution, the regression based on the log-transformed revenue returned positive treatment effects. Of the eight indices examined, only education index had an autocorrelation coefficient greater than 0.5, which suggests the use of a DID model to examine the treatment effects on education index. The DID estimates on education index are consistent with the findings reported in the original paper.

The complete-case analysis of education index, psychological well-being index and female empowerment index, and the multilevel analysis of psychological well-being index, considering the nested data structure, also returns findings consistent with those reported in the original paper. These analyses confirm the robustness of the positive short-term effects of UCT on asset holdings, consumption and revenue, as well as food security and psychological well-being in Kenya. These findings are also consistent with findings from the cash transfer literature in Kenya (Merttens et al. 2013; Attah et al. 2016).

The modeling of interaction effects does suggest the existence of village treatment interaction effects, especially for the psychological well-being index. The main treatment effect on the psychological well-being index was not significant after controlling for treatment village interactions, suggesting that the treatment effects may vary by village and the interaction effects are relatively large. Results from the multilevel analysis of the psychological well-being index show that there is a statistically significant village-level correlation. These findings suggest a need for contextual analysis to be conducted at the village-level to examine whether village differences were substantial and what village characteristics might have an impact on the effects of the UCT.

The replication study results also suggest a need for further examination of methods composing index measures of food security, health and psychological well-being. The PCA results for all three indices returned more than one factor: 4 factors on the 12 food security items, 3 factors on the 6 health items and 2 factors on the 6 psychological well-being items. Additional analyses are needed to explore and analyze the factors to determine what particular dimensions or concepts in food security, health or psychological well-being these factors measure and how they could be differentially affected by the UCT.

These results suggest that the items used to create the food security index, health index and psychological well-being index may represent two or more relevant but different factors in each domain. Although creating an index from a group of items in the same domain has certain benefits, including reducing the number of outcome variables for estimation, the exploration of more accurate definitions and measurement of food security, health and psychological well-being in this context is necessary. Accurate measurement of these outcomes can help understand the mechanisms that link UCT to the change in these outcomes to inform future study design and policy interventions.

The multivariate analyses on the eight indices, and on individual measures of psychological well-being, consumption and assets, had findings consistent with the results reported in the original paper. Overall, the main conclusion of the study on the impact of the UCT in rural Kenya did not change after considering the correlation among the outcome measures in the multivariate analyses. As the differences between the multivariate analysis results and the ANCOVA results are very small, they helped validate the main findings regarding the short-term impact of UCT. We advocate for the adoption of multivariate analysis as a methodology for dealing with multiple outcome measures that are correlated in this study.

The original study measured outcomes in a relatively short period after the rollout of the intervention, with about nine months between the baseline and endline measures. There were short-term effects of UCT on a variety of outcomes, including consumption, income, asset holdings, food security and psychological well-being, but not on other measures, such as health and education. One could argue that the effects of UCT on health and education could take more time to manifest.

The authors of the original study examined the long-run effects of UCT on the same outcome indices three years after implementation (Haushofer and Shapiro 2018). The long-run study results show sustained impact on asset holdings, but are not conclusive in terms of the long-term impacts on other outcome measures, due to the existence of spillover effects. These new findings, together with our replication study results, warrant further investigation into the effects of the UCT using more rigorous methods and a disinterested third-party evaluation (Özler 2018).

## 7. Limitations

Our replication was based on the final cleaned data provided to us by the authors. The data set was first transferred to SAS format, and we then conducted all replication analyses based on the cleaned data. After the initial review of the data files and software, we determined that the Stata codes provided to us by the original authors did not include the detailed formulas required to reproduce the index variables. We subsequently emailed the authors to verify our assessment and ask for the Stata codes used to create the indices using raw data, but we failed to receive a timely response. For the replication analyses, we did not recreate the index variables from individual measures and used the index variables created by the authors. This practice allowed us to repeat the original authors' statistical analyses by focusing on replicating their regression models to validate the reported study findings. However, the PCA results on the health, food security and psychological well-being indices suggested considerations of more than one factor (index) based on the individual measures in each of the three domains.

## 8. Conclusion

The results of this replication study support the conclusion that UCT in Kenya had short-term positive effects on economic outcomes, such as assets, consumption and revenue, and on recipients' psychological well-being. These findings were supported by a pure replication using entirely different software to replicate the analyses. The short-term impact of UCT was sustained when we conducted a variety of robustness checks concerning model validation, model specification and variance-covariance structure. These results provide evidence for the short-term impact of UCT and support using cash transfers as a way to reduce poverty and stimulate the economy in low-income countries. However, additional studies are needed to examine the effects of UCT in the longer term, especially its impact on health, education and psychological well-being measures. A comprehensive understanding of the effects of cash transfer programs over time will provide strong evidence for policy interventions, especially in low-income settings.

# Appendix: Figure and tables

## Figure A1: Histogram and residual plots for revenue and log-transformed revenue

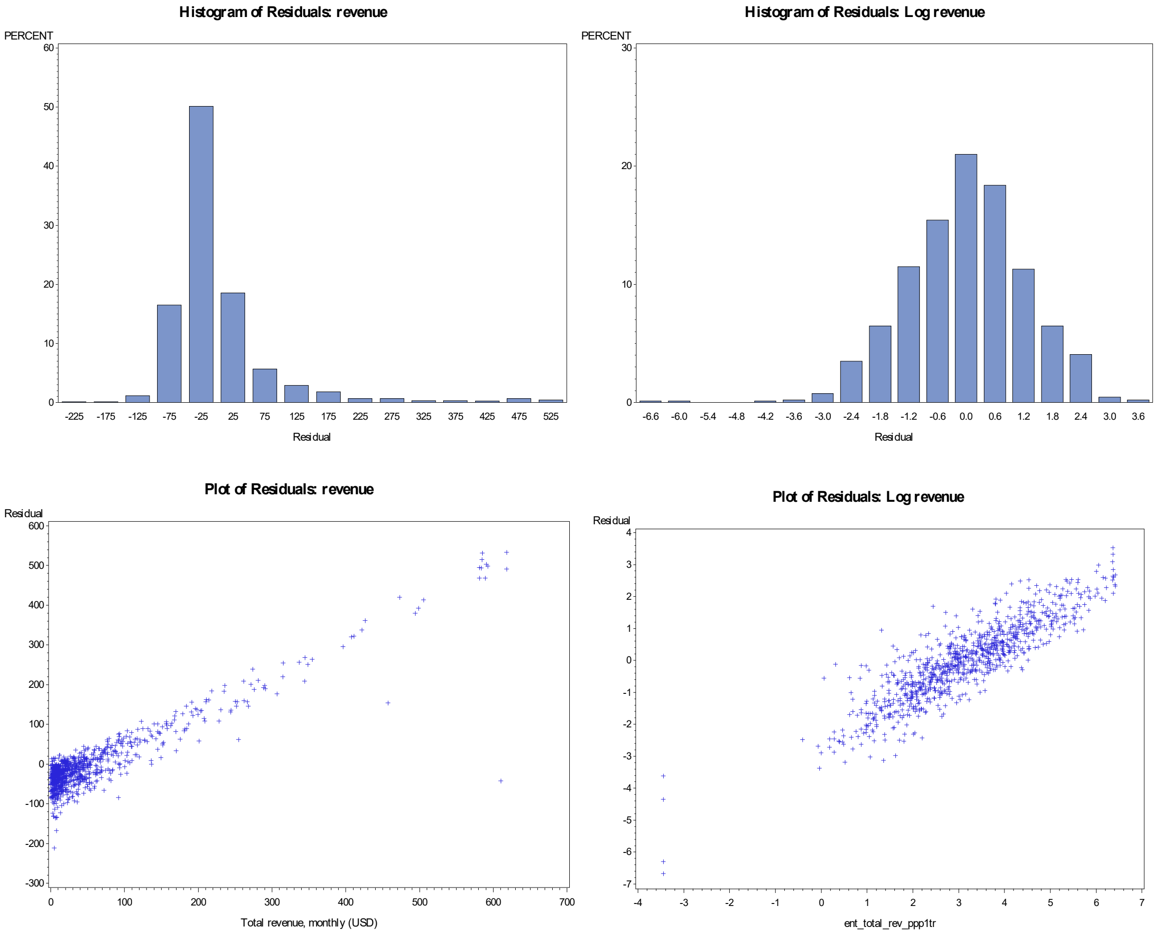

**Table A1: Outcome variable descriptions**

This table is from the original paper.

| **Index variables** | |
|---|---|
| Value of non-land assets (USD) | Total value (in 2012 USD PPP) of all non-land assets owned by the household, including savings, livestock, durable goods, and metal roofs. |
| Non-durable expenditure (USD) | Total monthly spending (in 2012 USD PPP) on non-durables, including food, temptation goods, medical care, education expenditure, and social expenditure. |
| Total revenue, monthly (USD) | Total monthly revenue (in 2012 USD PPP) from all household enterprises, including revenue from agriculture, stock and flow revenue from animals owned by the household, and revenue from all non-farm enterprises owned by any household member. |
| Food security index | A standardized weighted average of the (negatively coded) number of times household adults and children skipped meals, went whole days without food, had to eat cheaper or less preferred food, had to rely on others for food, had to purchase food on credit, had to hunt for or gather food, had to beg for food, or went to sleep hungry in the preceding week; a (negatively coded) indicator for whether the respondent went to sleep hungry in the preceding week; the (positively coded) number of times household members ate meat or fish in the preceding week; (positively coded) indicators for whether household members ate at least two meals per day, ate until content, had enough food for the next day, and whether the respondent ate protein in the last 24 hours; and the (positively coded) proportion of household members who ate protein in the last 24 hours, and proportion of children who ate protein in the last 24 hours. |
| Health index | A standardized weighted average of the (negatively coded) proportion of household adults who were sick or injured in the last month, the (negatively coded) proportion of household children who were sick or injured in the last month, the (positively coded) proportion of sick or injured family members for whom the household could afford treatment, the (positively coded) proportion of illnesses for which a doctor was consulted, the (positively coded) proportion of newborns who were vaccinated, the (positively coded) proportion of children below age 14 who received a health checkup in the preceding six months, the (negatively coded) proportion of children under 5 who died in the preceding year, and a children's anthropometrics index consisting of body mass index, height-for-age, weight-for-age, and upper arm circumference relative to World Health Organization development benchmarks. |
| Education Index | A standardized weighted average of the proportion of household children enrolled in school and the amount spent by the household on educational expenses per child. |
| Psychological well-being index | A standardized weighted average of their (negatively coded) scores on the CES-D scale,[i] a custom worries questionnaire (negatively coded), Cohen's stress scale (negatively coded),[ii] their response to the World Values Survey happiness and life satisfaction questions, and their log cortisol levels adjusted for confounders (negatively coded). |
| Female Empowerment Index | A standardized weighted average of a measure of two other indices, a violence and an attitude index. The violence index is a weighted standardized average of the frequency with which the respondent reports having been physically, sexually, or emotionally abused by her husband in the preceding six months; the attitude index is a weighted standardized average of a measure of the respondent's view of the justifiability of violence against women, and a scale of male-focused attitudes. |

| | |
|---|---|
| **Psychological well-being** | |
| Log cortisol (no controls) | Log of salivary cortisol levels in nm/l, taken as the average of level across two samples. |
| Log cortisol (with controls) | Residuals of an OLS regression of the log-transformed cortisol levels on dummies for having ingested food, drinks, alcohol, nicotine, or medications in the two hours preceding the interview, for having performed vigorous physical activity on the day of the interview, and for the time elapsed since waking, rounded to the next full hour |
| Depression (CES-D) | The standardized total of the score from the 20 elements of the CES-D questionnaire.[i] |
| Worries | The standardized total of the score from the 16 elements of a custom worries questionnaire. |
| Stress (Cohen) | The standardized total of score from 4 elements of Cohen’s stress scale. [ii] |
| Happiness (WVS) | The standardized numerical response to the World Values Survey happiness question: Taking all things together, would you say you are “very happy” (1), “quite happy” (2), “not very happy” (3), or “not at all happy” (4)? |
| Life satisfaction (WVS) | The standardized numerical response to the World Values Survey life satisfaction question: All things considered, how satisfied are you with your life as a whole these days on a scale of 1 to 10? (1= very dissatisfied…10= very satisfied) |
| Trust (WVS) | The standardized numerical response to the World Values Survey life satisfaction question: Generally speaking, would you say that most people can be trusted (1) or that you need to be very careful in dealing with people (2)? |
| Locus of control | The standardized weighted average of the total score on Rotter’s locus of control questionnaire and the numerical response the World Values Survey locus of control question (Some people believe that individuals can decide their own destiny, while others think that it is impossible to escape a predetermined fate. Please tell me which comes closest to your view on this scale on which 1 means “everything in life is determined by fate” and 10 means “people shape their fate themselves.”) Higher scores indicate a more internal locus of control. |
| Optimism (Scheier) | The standardized total score on the 6 question Scheier optimism questionnaire. |
| Self-esteem (Rosenberg) | The standardized total score on the 10 question Rosenberg optimism questionnaire. |
| **Consumption** | |
| Food total (USD) | The combined monthly total (in 2012 USD PPP) of all spending on food by the household and the value of all food produced from agriculture and livestock that was consumed by the household (calculated as the monthly average of total production in the last year). |
| Cereals (USD) | The monthly total spending (in 2012 USD PPP) by all members of the household on all cereal grains, flours, breads, pastas, cakes and biscuits. |
| Meat & fish (USD) | The monthly total spending (in 2012 USD PPP) by all members of the household on all meat and fish. |
| Alcohol (USD) | The monthly total spending (in 2012 USD PPP) by all members of the household on alcoholic beverages. |
| Tobacco (USD) | The monthly total spending (in 2012 USD PPP) by all members of the household on tobacco products, including cigarettes, cigars, tobacco, snuff, khatt, and miraa. |
| Social expenditure (USD) | The monthly total spending (in 2012 USD PPP) by all members of the household on ceremonies, weddings, funerals, dowries / bride prices, charitable donations, village elder fees, and recreation or entertainment. |
| Medical expenditure past month (USD) | The monthly total spending (in 2012 USD PPP) on medical care for all household members including consultation fees, medicines, hospital costs, lab test costs, ambulance costs, and related transport. |
| Education expenditure (USD) | The monthly total spending (in 2012 USD PPP) for household members on school/college fees, uniforms, books, and other supplies. |

| | |
|---|---|
| **Assets** | |
| Value of livestock (USD) | The total value (in 2012 USD PPP) of all cattle (cows, bulls, or calves), birds (chicken, turkeys, doves, quails, etc.), and small livestock (pigs, sheep, goats, etc.) owned by the household. |
| Value of durable goods (USD) | The total value (in 2012 USD PPP) of all durable goods, including transportation, furniture, agricultural equipment, appliances, radios and televisions, and phones owned by the household. |
| Value of savings (USD) | The total value (in 2012 USD PPP) of all savings held by all members of the household in any account (post bank, SACCO, village bank, M-Pesa, Zap, ROSCA, commercial bank, microfinance institution, etc.), at home, or held by friends, relatives or colleagues. |
| Land owned (acres) | The total value (in 2012 USD PPP) of all land owned by any household member. |
| Has non-thatched roof (dummy) | An indicator variable taking the value of one if the household's primary residence has a metal or concrete roof and zero otherwise. |
| **Business activities** | |
| Wage labor primary income (dummy) | An indicator variable taking the value of one if the household's primary source of income is wage labor and zero otherwise. |
| Own farm primary income (dummy) | An indicator variable taking the value of one if the household's primary source of income is a farm owned by the household and zero otherwise. |
| Non-agricultural business owner (dummy) | An indicator variable taking the value of one if any member of the household is full or part owner of a non-agricultural enterprise and zero otherwise. |
| Total revenue, monthly (USD) | Total monthly revenue (in 2012 USD PPP) from all household enterprises, including revenue from agriculture, stock and flow revenue from animals owned by the household, and revenue from all non-farm enterprises owned by any household member. |
| Total expenses, monthly (USD) | Total in 2012 USD PPP of all expenses on agricultural enterprises including seeds, fertilizers / herbicides / pesticides, hired machines, water, and labor, and all expenses on non-agricultural enterprises including machinery / durable goods, inputs / inventory, salaries / wage, transport, electricity, and water. |
| Total profit, monthly (USD) | Total imputed profit in 2012 USD PPP from all agricultural and non-agricultural enterprises owned by the household. |

Notes:

i. Radloff, LS, 1977. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, pp.385–401.

ii. Cohen, S, Kamarck, T and Mermelstein, R, 1983. A global measure of perceived stress. *Journal of Health and Social Behavior*, 24(4), pp.385–396.

### Table A2: Baseline differences in index variables

| | Control mean (SD) | Treatment effect | Female recipient | Monthly transfer | Large transfer | N |
|---|---|---|---|---|---|---|
| Value of non-land assets (USD) | 383.36 | –1.15 | 15.53 | 25.16 | 13.76 | 1,008 |
| | (374.15) | (24.74) | (43.62) | (39.33) | (42.77) | |
| | | [1.00] | [0.93] | [0.91] | [0.99] | |
| Non-durable expenditure (USD) | 181.99 | –6.16 | –28.05* | –8.01 | –5.56 | 1,008 |
| | (127.16) | (8.31) | (15.14) | (13.28) | (14.36) | |
| | | [0.97] | [0.33] | [0.91] | [0.99] | |
| Total revenue, monthly (USD) | 84.92 | –33.19* | –31.77** | –7.59 | –10.77 | 1,008 |
| | (402.59) | (18.54) | (14.34) | (14.99) | (12.38) | |
| | | [0.43] | [0.15] | [0.91] | [0.95] | |
| Food security index | 0.00 | 0.00 | 0.05 | 0.25** | –0.01 | 1,008 |
| | (1.00) | (0.06) | (0.09) | (0.10) | (0.09) | |
| | | [1.00] | [0.93] | [0.08]* | [0.99] | |
| Health index | 0.01 | 0.03 | 0.26*** | 0.14 | –0.14 | 1,008 |
| | (1.02) | (0.06) | (0.09) | (0.10) | (0.10) | |
| | | [0.98] | [0.04]** | [0.54] | [0.69] | |
| Education index | 0.00 | –0.07 | 0.14 | 0.16* | –0.05 | 853 |
| | (1.00) | (0.06) | (0.09) | (0.09) | (0.09) | |
| | | [0.89] | [0.41] | [0.42] | [0.98] | |
| Psychological well-being index | 0.00 | 0.03 | 0.02 | 0.19** | 0.18** | 1,569 |
| | (1.00) | (0.05) | (0.08) | (0.08) | (0.08) | |
| | | [0.98] | [0.93] | [0.14] | [0.17] | |
| Female empowerment index | 0.00 | –0.05 | 0.08 | 0.18 | 0.03 | 751 |
| | (1.00) | (0.07) | (0.11) | (0.12) | (0.12) | |
| | | [0.97] | [0.93] | [0.52] | [0.99] | |
| Joint test (p-value) | | 0.64 | 0.00*** | 0.02** | 0.36 | |

Note: Ordinary least squares estimates of baseline differences in treatment arms. Outcome variables are listed on the left and described in detail in Table A1. For each outcome variable, we report the coefficients of interest and their standard errors in parentheses. FWER-corrected p-values are in brackets. Column (1) reports the mean and standard deviation of the control group for a given outcome variable. Column (2) reports the basic treatment effect, i.e. comparing treatment households to control households within villages. Column (3) reports the relative treatment effect of transferring to the female compared to the male; column (4) the relative effect of monthly compared to lump-sum transfers; and column (5) that of large compared to small transfers. The unit of observation is the household for all outcome variables except for the psychological variables index, where it is the individual. The sample is restricted to cohabitating couples for the female empowerment index and households with school-age children for the education index. The comparison of monthly to lump-sum transfers excludes large-transfer recipient households, and that for male versus female recipients excludes single-headed households. All columns include village-level fixed effects and cluster standard errors at the household level. The last row shows joint significance of the coefficients in the corresponding column from SUR estimation. * denotes significance at 10 percent, ** at 5 percent and *** at 1 percent level.

## Table A3: Treatment effects: index variables

| | Control mean (SD) | Treatment effect | Female recipient | Monthly transfer | Large transfer | N |
|---|---|---|---|---|---|---|
| Value of non-land assets (USD) | 494.80 | 301.51*** | –79.46 | –91.85** | 279.18*** | 940 |
| | (415.32) | (27.25) | (50.38) | (45.92) | (49.09) | |
| | | [0.00]*** | [0.51] | [0.27] | [0.00]*** | |
| Non-durable expenditure (USD) | 157.61 | 35.66*** | –2.00 | –4.20 | 21.25** | 940 |
| | (82.18) | (5.85) | (10.28) | (10.71) | (10.49) | |
| | | [0.00]*** | [0.93] | [0.99] | [0.22] | |
| Total revenue, monthly (USD) | 48.98 | 16.15*** | 5.41 | 16.33 | –2.44 | 940 |
| | (90.52) | (5.88) | (10.61) | (11.07) | (8.87) | |
| | | [0.03]** | [0.93] | [0.59] | [0.84] | |
| Food security index | 0.00 | 0.26*** | 0.06 | 0.26** | 0.18* | 940 |
| | (1.00) | (0.06) | (0.09) | (0.11) | (0.10) | |
| | | [0.00]*** | [0.93] | [0.12] | [0.24] | |
| Health index | 0.00 | –0.03 | 0.10 | 0.01 | –0.09 | 940 |
| | (1.00) | (0.06) | (0.09) | (0.10) | (0.09) | |
| | | [0.82] | [0.73] | [0.99] | [0.71] | |
| Education index | 0.00 | 0.08 | 0.06 | –0.05 | 0.05 | 823 |
| | (1.00) | (0.06) | (0.09) | (0.10) | (0.09) | |
| | | [0.43] | [0.93] | [0.99] | [0.84] | |
| Psychological well-being index | 0.00 | 0.26*** | 0.14* | 0.01 | 0.26*** | 1,474 |
| | (1.00) | (0.05) | (0.08) | (0.08) | (0.08) | |
| | | [0.00]*** | [0.43] | [0.99] | [0.01]*** | |
| Female empowerment index | 0.00 | –0.01 | 0.17* | 0.05 | 0.22** | 698 |
| | (1.00) | (0.07) | (0.10) | (0.12) | (0.11) | |
| | | [0.88] | [0.50] | [0.99] | [0.22] | |
| Joint test (p-value) | | 0.00*** | 0.11 | 0.04** | 0.00*** | |

Note: Ordinary least squares estimates of treatment effects. Outcome variables are listed on the left and described in detail in Table A1. Higher values correspond to “positive” outcomes. For each outcome variable, we report the coefficients of interest and their standard errors in parentheses. FWER-corrected p-values are in brackets. Column (1) reports the mean and standard deviation of the spillover group, column (2) the basic treatment effect, i.e. comparing treatment households to control households within villages. Column (3) reports the relative treatment effect of transferring to the female compared to the male; column (4) the relative effect of monthly compared to lump-sum transfers; and column (5) that of large compared to small transfers. The unit of observation is the household for all outcome variables except for the psychological variables index, where it is the individual. The sample is restricted to cohabitating couples for the female empowerment index, and households with school-age children for the education index. The comparison of monthly to lump-sum transfers excludes large-transfer recipient households, and that for male versus female recipients excludes single-headed households. All columns include village-level fixed effects, control for baseline outcomes, and cluster standard errors at the household level. The last row shows joint significance of the coefficients in the corresponding column from SUR estimation. * denotes significance at 10 percent, ** at 5 percent and *** at 1 percent level.

**Table A4: Spillover effects: index variables**

| | Spillover effects | | | | | | Lee Bounds | | Horowitz–Manski Bounds | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **All HH estimate** | **All HH estimate** | **Thatched estimate** | **Thatched estimate** | **Test (1)=(3) p-value** | **Test (2)=(4) p-value** | **Lower** | **Upper** | **Lower** | **Upper** |
| **Includes controls** | **No** | **Yes** | **No** | **Yes** | **No** | **Yes** | **No** | **No** | **No** | **No** |
| Value of non-land assets (USD) | 1.00 | –11.99 | –18.73 | –32.61 | 0.01*** | 0.00*** | –3.38 | 12.84 | –2.38 | 7.39 |
| | (21.44) | (19.98) | (21.14) | (19.76) | | | (20.07) | (20.94) | (19.92) | (20.06) |
| Non-durable expenditure (USD) | –7.77 | –11.89* | –7.31 | –12.21* | 0.82 | 0.88 | –9.47 | –4.08 | –8.93 | –5.86 |
| | (7.20) | (6.50) | (7.27) | (6.67) | | | (6.08) | (6.32) | (5.79) | (5.83) |
| Total revenue, monthly (USD) | –3.68 | –3.64 | –5.23 | –5.78 | 0.56 | 0.45 | –4.29 | 2.32 | –4.18 | –1.91 |
| | (6.18) | (6.35) | (5.84) | (6.01) | | | (5.88) | (5.85) | (6.18) | (6.22) |
| Food security index | 0.06 | 0.05 | 0.06 | 0.05 | 0.93 | 0.90 | –0.01 | 0.08 | 0.03 | 0.07 |
| | (0.09) | (0.09) | (0.10) | (0.10) | | | (0.09) | (0.08) | (0.08) | (0.08) |
| Health index | –0.06 | –0.07 | –0.06 | –0.08 | 0.80 | 0.66 | –0.10 | –0.03 | –0.07 | –0.04 |
| | (0.08) | (0.08) | (0.08) | (0.08) | | | (0.07) | (0.07) | (0.07) | (0.07) |
| Education index | 0.01 | –0.01 | 0.00 | –0.03 | 0.36 | 0.29 | –0.10 | 0.10 | –0.01 | 0.03 |
| | (0.07) | (0.06) | (0.08) | (0.07) | | | (0.10) | (0.08) | (0.07) | (0.07) |
| Psychological well-being index | 0.03 | 0.03 | 0.03 | 0.02 | 0.77 | 0.71 | 0.03 | 0.04 | 0.01 | 0.05 |
| | (0.07) | (0.07) | (0.07) | (0.07) | | | (0.06) | (0.06) | (0.05) | (0.05) |
| Female empowerment index | 0.21** | 0.21** | 0.23** | 0.22** | 0.50 | 0.50 | 0.20** | 0.28** | 0.18** | 0.23*** |
| | (0.09) | (0.08) | (0.09) | (0.09) | | | (0.10) | (0.11) | (0.08) | (0.08) |
| Joint test (p-value) | 0.38 | 0.25 | 0.23 | 0.11 | | | | | | |

Note: Ordinary least squares estimates of spillover effects. Outcome variables are listed on the left and described in detail in Table A1. The unit of observation is the household for all variables except psychological well-being, where it is the individual. The sample includes all households and individuals, except for the intrahousehold index, where it is restricted to cohabitating couples, and for the education index, where it is restricted to households with school-age children. Columns (1) and (2) report the "naïve" estimate of spillover effects, including spillover households that upgraded to metal roofs between baseline and endline. Columns (3) and (4) report estimates of the spillover effect excluding metal-roof households. Columns (1) and (3) exclude baseline covariates. Columns (2) and (4) include baseline covariates. Column (5) reports the p-value of the equality for the coefficient estimates in (1) and (3) after joint estimation of the two models using SUR. Column (6) reports the p-value of the equality for the coefficient estimates in (2) and (4) after joint estimation of the two models using SUR. The last row reports p-values on the joint significance of all coefficients in a given column after joint estimation using SUR. Columns (7) and (8) report the lower and upper Lee bounds adjusting for differential "attrition" of five spillover households into metal-roof upgrade. Columns (9) and (10) report lower and upper Horowitz-Manski bounds, imputing outcomes for the five attritting households using the 95th and 5th percentiles of observed outcomes, respectively. The sample is restricted as in the previous tables. In columns (1) through (4), standard errors clustered at the village level are reported in parentheses. In columns (7) through (10), bootstrapped standard errors are reported in parentheses. * denotes significance at 10 percent, ** at 5 percent and *** at 1 percent level.

## References

Anderson, M, 2008. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), pp.1,481–1,495.

Anderson, TW, 2003. *An Introduction to Multivariate Statistical Analysis*. 3rd edition. New York: John Wiley & Sons.

Attah, R, Barca, V, Kardan, A, MacAuslan, I, Merttens, F and Pellerano, L, 2016. Can Social Protection Affect Psychosocial Wellbeing and Why Does This Matter? Lessons from Cash Transfers in Sub-Saharan Africa. *The Journal of Development Studies*, 52(8), pp.1,115–1,131.

Baird, S, McIntosh, C and Özler B, 2011. Cash or condition? Evidence from a cash transfer experiment. *The Quarterly Journal of Economics*, 126(4), pp.1,709–1,753.

Barca, V, Hurrell, A, MacAuslan, I, Visram, A and Willis, J, 2013. Paying attention to detail: how to transfer cash in cash transfers. *Enterprise Development and Microfinance*, 24(1), pp.10–27.

Bertrand, M, Duflo, E and Mullainathan, S, 2004. How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1), pp.249–275.

Cesarini, D, Lindqvist, E, Notowidigdo, MJ and Östling, R, 2015. *The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries*. Working Paper 21762, National Bureau of Economic Research.

Cohen, J, Cohen, P, West, SG and Aiken, LS, 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3rd Edition. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Cohen, S, Kamarck, T and Mermelstein, R, 1983. A global measure of perceived stress. *Journal of Health and Social Behavior*, 24(4), pp.385–396.

Freese, J and Peterson, D, 2017. Replication in Social Science. *Annual Review of Sociology*, 43, pp.147-165.

Gelman, A and Hill J, 2007. *Data Analysis using Regression and Multilevel/Hierarchical models.* Cambridge: Cambridge University Press.

Haushofer, J and Shapiro, J, 2013. Welfare Effects of Unconditional Cash Transfers: Pre-analysis Plan. Available at: <http://www.princeton.edu/haushofer/publications/Haushofer_Shapiro_UCT_Analysis_Plan_2013.pdf> [accessed December 2016].

Haushofer, J and Shapiro, J, 2016. The Short-Term Impact of Unconditional Cash Transfers to the Poor: Evidence from Kenya. *Quarterly Journal of Economics*, 131(4), pp.1,973–2,042.

Haushofer, J and Shapiro, J, 2017. Erratum to “The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya”. *Quarterly Journal of Economics*, 132(4), pp.2,057-2,060.

Haushofer, J and Shapiro, J, 2018. *The Long-Term Impact of Unconditional Cash Transfers: Experimental Evidence from Kenya*. Working Paper, Princeton University. Available at: <http://jeremypshapiro.com/papers/Haushofer_Shapiro_UCT2_2018-01-30_paper_only.pdf> [Accessed July 2018].

Horowitz, JL and Manski, CF, 1995. Identification and robustness with contaminated and corrupted data. *Econometrica*, 63(2), pp.281–302.

Johnson, RA and Wichern, DW, 2007. *Applied Multivariate Statistical Analysis*. 6th edition. Upper Saddle River, NJ: Pearson Prentice Hall.

Jones, MP, 1996. Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression. *Journal of the American Statistical Association*, 91(433), pp.222–230.

Kuehl, RO, 1999. *Design of Experiments: Statistical Principles of Research Design and Analysis*. 2nd edition. Pacific Grove, CA: Duxbury Press.

Kutner, MH, Nachsheim, CJ, Neter, J and Li, W, 2004. *Applied Linear Statistical Models*. 5th edition. McGraw-Hill/Irwin Publishers.

Lee, DS, 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3), pp.1071–1102.

McKenzie, D, 2012. Beyond baseline and follow-up: the case for more T in experiments. *Journal of Development Economics*, 99(2), pp.210–221.

Merttens, F, Hurrell, A, Marzi, M, Attah, R, Farhat, M, Kardan, A and MacAuslan, I, 2013. *Kenya Hunger Safety Net Programme Monitoring and Evaluation Component*. Impact Evaluation Final Report: 2009 to 2012. Available at: <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/284251/Kenya-Hunger-Safety-Net-Programme-2009-2012.pdf> [accessed December 2016].

Metcalf, C, 1972. *An Econometric Model of the Income Distribution*. Chicago: Markham.

Özler, B, 2018. GiveDirectly Three-Year Impacts, Explained. Available at: <http://blogs.worldbank.org/impactevaluations/givedirectly-three-year-impacts-explained> [Accessed July 2018].

Radloff, LS, 1977. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, pp.385–401.

Vyas, S and Kumaranayake, L, 2006. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy and Planning*, 21(6), pp.459–468.

Wang, H and Luo, J, 2017. *A replication study proposal for 3ie’s replication window 4: financial services for the poor on “the short-term impact of unconditional cash transfers to the poor: evidence from Kenya” by Johannes Haushofer and Jeremy Shapiro*.

Available at: <http://www.3ieimpact.org/media/filer_public/2017/06/30/wang_revised_replication_plan.pdf> [Accessed December 2017].

Wydick, B, 2018. *When are Cash Transfers Transformative?* Working Paper Series No. WPS-069. Center for Effective Global Action, University of California, Berkeley. Available at: <https://escholarship.org/uc/item/6xb8f996#article_main> [Accessed July 2018].

## Other publications in the 3ie Replication Paper Series

The following papers are available from http://3ieimpact.org/evidence-hub/publications/replication-papers

*Mobile money and its impact on improving living conditions in Niger: a replication study*. 3ie Replication Paper 19. Meneses, JP, Ventura, E, Elorreaga, O, Huaroto, C, Aguilar, G, and Beteta, E, 2018.

*Savings revisited: a replication study of a savings intervention in Malawi,* 3ie Impact Evaluation Report 18. Stage, J and Thangavelu, T 2018.

*Thou shalt be given…but how? A replication study of a randomized experiment on food assistance in northern Ecuador*, 3ie Replication Paper 17. Lhachimi, SK and Seuring, T, 2018.

*Preventing HIV and HSV-2 through improving knowledge and attitudes: a replication study of a multicomponent intervention in Zimbabwe.* 3ie Replication Paper 16. Hein, NA, Bagenda, DS and Yu, F, 2018.

*PEPFAR and adult mortality: a replication study of HIV development assistance effects in Sub-Saharan African countries.* 3ie Replication Paper 15. Hein, NA, Bagenda, DS and Luo, J, 2018.

*When to start ART? A replication study of timing of antiretroviral therapy for HIV-1-associated Tuberculosis*. 3ie Replication Paper 14. Djimeu, EW, 2018.

*STRETCHing HIV treatment: a replication study of task shifting in South Africa*. 3ie Replication Paper 13. Chen, B and Alam, M, 2017.

*Cash transfers and HIV/HSV-2 prevalence: a replication of a cluster randomized trial in Malawi*. 3ie Replication Paper 12. Smith, LM, Hein, NA and Bagenda, DS, 2017.

*Power to the people?: a replication study of a community-based monitoring programme in Uganda,* 3ie Replication Paper 11. Donato, K and Garcia Mosqueira, A (2016)

*Fighting corruption does improve schooling: a replication study of a newspaper campaign in Uganda,* 3ie Replication Paper 10. Kuecken, M, and Valfort, MA (2016)

*The effects of land titling on the urban poor: a replication of property rights,* 3ie Replication Paper 9. Cameron, Drew B, Whitney, Edward M and Winters, Paul C (2015)

*Male circumcision and HIV acquisition reinvestigating the evidence from young men in Kisumu, Kenya*, 3ie Replication Paper 8. Djimeu, EW, Korte, JE and Calvo, FA (2015)

*Walking on solid ground: a replication study on Piso Firme's impact,* 3ie Replication Paper 7. Basurto, MP, Burga, R, Toro, JLF and Huaroto, C (2015)

*The impact of India's JSY conditional cash transfer programme: A replication study,* 3ie Replication Paper 6. Carvalho, N and Rokicki, S (2015)

*Recalling extra data: A replication study of finding missing markets*, 3ie Replication Paper 5. Wood, BDK and Dong, M (2015)

*The long and short of returns to public investments in fifteen Ethiopian villages*, 3ie Replication Paper 4. Bowser, WH (2015)

*Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 2: Alternative analyses,* 3ie Replication Paper 3, part 2. Aiken, AM, Davey, C, Hayes, RJ and Hargreaves, JR (2014)

*Reanalysis of health and educational impacts of a school-based deworming program in western Kenya Part 1: A pure replication,* 3ie Replication Paper 3, part 1. Aiken, AM, Davey, C, Hargreaves, JR and Hayes, RJ (2014)

*TV, female empowerment and demographic change in rural India,* 3ie Replication Paper 2. Iversen, V and Palmer-Jones, R (2014)

*Quality evidence for policymaking: I'll believe it when I see the replication,* 3ie Replication Paper 1. Brown, AN, Cameron, DB and Wood, BDK (2014)

**Replication Paper Series**

International Initiative for Impact Evaluation
1029 Vermont Avenue, NW
Suite 1000
Washington, DC 20005
USA

replication@3ieimpact.org
Tel: +1 202 629 3939

www.3ieimpact.org