Stefan K Lhachimi
Till Seuring

Thou shalt be given…but how?

A replication study of a randomized experiment on food assistance in northern Ecuador

November 2018

Replication Paper 17

Multisector

International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) promotes evidence-informed equitable, inclusive and sustainable development. We support the generation and effective use of high-quality evidence to inform decision-making and improve the lives of people living in poverty in low- and middle-income countries. We provide guidance and support to produce, synthesise and quality assure evidence of what works, for whom, how, why and at what cost.

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 is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or the 3ie Board of Commissioners. Any errors and omissions are the sole responsibility of the authors. Please direct any comments or queries to the corresponding author, Stefan K Lhachimi, at Stefan.Lhachimi@uni-bremen.de.

Suggested citation: Lhachimi, SK and Seuring, 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. Washington, DC: International Initiative for Impact Evaluation (3ie). Available at: http://doi.org/10.23846/RPS0017

3ie Replication Paper Series executive editor: Marie Gaarder
Co-managing editor: Scott Neilitz and John Creamer
Production manager: Brigid Monaghan
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: Akarsh Gupta

Thou shalt be given... but how? A replication study of a randomized experiment on food assistance in northern Ecuador

Stefan K Lhachimi
Research Group for Evidence-Based Public Health

Till Seuring
Research Group for Evidence-Based Public Health

3ie Replication Paper 17

November 2018

Acknowledgments

Thanks go to the three anonymous reviewers, participants of the 3ie workshop in Washington, DC, the external project adviser and the authors of the original paper for their valuable comments.

Summary

The 2014 paper by Hidrobo and colleagues analyzes a cluster-randomized controlled trial comparing the effectiveness of different modes of food assistance (cash, food and vouchers) with a control mode (no assistance). The statistical analysis was done using an analysis of covariance model and included several food-related outcome measures (food consumption, several indices of food security and diet). Robustness checks were made and effect estimates reported, adjusted for co-variates. Finally, the costs associated with each mode of assistance are calculated and cost-effectiveness measures presented. Comparing these three modes directly against each other is of vital importance for policymakers, as these are thought to have distinct advantages and/or disadvantages in terms of efficacy, public acceptance and cost. The original paper is one of very few studies that report the results of a head-to-head comparison of all three modes.

A main objective of this replication research is to conduct a pure replication of Hidrobo and colleagues' study, i.e. analyzing whether findings can be reproduced using the original study's own data and methods. The second objective is to investigate the robustness of the findings through additional analysis of the data, in particular investigation of possible contamination during the sampling process and explicitly modeling the hierarchical structure of the sampling frame. In the theory of change analysis, we extend the paper's cost-effectiveness analysis and present additional measures, explicitly accounting for uncertainty.

We find that the results of the original paper are fully replicable. We do not find any meaningful differences in the reported results or the interpretation of the results. However, we identify indications of a contamination occurring across clusters and that some intervention effects may vary in magnitude by province. Our theory of change analysis suggests that further research should aim to reduce the uncertainty around whether cash or vouchers are more cost-effective.

Contents

List of figures and tables

Appendix tables

Abbreviations and acronyms

ANCOVA	Analysis of covariance
DDI	Dietary diversity index
EVPI	Expected value of perfect information
FCS	Food consumption score
GLMM	Generalized linear mixed models
HDDS	Household dietary diversity score
ICC	Intra-cluster correlation
ICER	Incremental cost-effectiveness ratio
Log	Logarithm
PFC	Poor food consumption
SD	Standard deviation
SDG	Sustainable Development Goal
WTP	Willingness to pay

1. Overview

This report documents the pure replication step of the replication of Hidrobo and colleagues' 2014 paper, *Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador*. We have already conducted a push-button replication (documented in Appendix A). To review the original analysis step by step and carefully check for possible errors and robustness, we reprogrammed the original analysis in R (Appendix B). For the main results of the paper, we were fully able to replicate all results satisfactory. However, we were not able to fully replicate all steps of the variable construction. Nevertheless, we deem the substantive result of the original paper replicable. The measurement and estimation analysis revealed some indication of contamination across clusters. Nevertheless, we did not find any meaningful differences in the reported results or in the interpretation of the results.

This paper is organized in eight sections. In Section 2, we give an overview of this paper and the broader research topic. In Section 3, we briefly outline the push-button replication. In Section 4, we describe our main analysis, including an overview of the data set and the issues we encountered in recreating some variable used in the analysis, in particular, the food consumption data (Section 4.1). Section 4.2 documents the replication of the main analysis, including a table on cost-effectiveness analysis for which no code was provided (Section 4.2.1). We proceed in replicating the robustness checks and additional analysis, which are mainly documented in the appendix of the original paper. We are fully able to replicate all tables. In Section 4.2.4 and following, we try to replicate the constructed variables (indices) and assess whether using these reconstructed variables changes the results of the analysis. Where we are able to replicate the variables, we do not find any meaningful differences in the results.

Section 5 reports the results of the measurement and estimation analysis, which explores alternative measurement and estimation techniques by investigating selected assumptions, i.e. no contamination of treatment arms, the specification of the employed an analysis of covariance (ANCOVA) model and no heterogeneous effects across provinces. We further test the robustness of results to the exclusion of *barrios* (neighborhoods).[1] In Section 6, we conduct a theory of change analysis for the cost-effectiveness analysis presented in the original paper. We discuss limitations in Section 7 and present our conclusion in Section 8.

2. Background

The paper by Hidrobo and colleagues (2014) analyzes a cluster-randomized controlled trial comparing the effectiveness of different modes of food assistance (cash, food and vouchers) with a control mode (no assistance). The intervention took place in two provinces of Ecuador, Carchi and Sucumbíos, that share a border with Colombia and experienced a strong influx of Colombian refugees before implementation. The provinces have important differences. Their locations – Carchi in the northern highlands and Sucumbíos in the Amazonian lowlands – means they have different cultural, socio-economic and geographical characteristics. Distinct *barrios* in urban areas were chosen

[1] *Barrios* are administrative units within the urban centers, with oversight over social services and other administrative functions.

as geographic units where the interventions should take place, with the possibility that a *barrio* could contain several smaller geographical units (clusters), where each cluster was used to implement one of the three treatment modes.

The statistical analysis uses an ANCOVA model and includes several food-related outcome measures: food consumption, several indices of food security, and diet. Robustness checks are made and the effect estimates, adjusted for co-variates, reported. Finally, the costs associated with each mode of assistance are calculated and cost-effectiveness measures presented. Food assistance to counter malnutrition or undernutrition among vulnerable segments of the population is an ongoing concern for many countries, most notably for low- and middle-income countries, where social protection mechanisms are often underdeveloped.

Malnutrition or undernutrition can have severe long-term consequence on human capital, particularly on the ability to study and work (Marmot et al. 2012; Hidrobo et al. 2014). General food subsidies – e.g. fixing the price of certain food commodities – have been shown to be highly inefficient in targeting the neediest part of the population while being prohibitively expensive for many governments in the long run. Recently, different types of social assistance interventions targeting the most vulnerable have become more common and gained prominence on development agendas in what is often considered a “quiet revolution” (Barrientos and Hulme 2008).

Social assistance interventions are usually defined as “noncontributory transfer programs targeted in some manner to the poor and those vulnerable to poverty and shocks” (World Bank 2011) to ensure an adequate standard of living and ensure long-term health. Social assistance interventions are often differentiated into cash transfers, in-kind transfers, fee waivers/vouchers, subsidies and public works programs (Pega et al. 2015b). All three modes investigated by the original paper (cash, in-kind transfers and vouchers) are under policy discussion or used in several countries, and it is widely agreed that all three modalities work, in absolute terms, in increasing caloric intake (Gentilini 2016).

However, recent systematic reviews show that evidence based on high-quality studies – i.e. randomized controlled trials that conduct a head-to-head comparison between these modes – is exceedingly rare (Pega et al. 2015b; Pega et al. 2017; Lagarde et al. 2009). Many studies are either observational or compare one of the three interventions against a non-intervention only. However, comparing these three modes directly against each other is of vital importance for policymakers, as the modes are thought to have distinct advantages and/or disadvantages in terms of efficacy, public acceptance and cost. Some see direct cash transfers as a more efficient way of providing help, due to their lower disbursement costs and ability to allow recipients to buy goods that truly increase their utility (Fiszbein et al. 2009). However, a potential drawback of cash is that recipients could spend cash not solely on beneficial goods, e.g. on tobacco (Pega et al. 2015a). In-kind transfers, which are not exchangeable, may have more beneficial health effects if the quality and quantity of food provided exceeds that bought from a cash transfer. Yet in-kind transfers are costly to administer and reduce the agency of the recipients. Hence, some argue that vouchers occupy a middle ground between these two modes of assistance.

Even less frequent in the literature on transfers is the provision of cost-effectiveness estimates. These estimates would allow policymakers not only to judge an intervention by its effects, but also to get an idea of how much they would need to invest to achieve a certain goal. Hidrobo and colleagues provide such an analysis, suggesting that the provision of food is the least cost-effective method, while vouchers are, in most cases, the cost-effective strategy. Based on these results, it appears that for policymakers, the more attractive option in a low-income setting is the use of vouchers to improve dietary outcomes. To add to this potentially important finding, we extend the existing cost-effectiveness analysis using a probabilistic sensitivity analysis that will help policymakers judge the robustness of this finding with respect to the uncertainty in the underlying costs and effect estimates.

These interventions relate to several Sustainable Development Goals (SDGs), as defined by a United Nations summit in 2015 – most clearly, of course, to SDG 1 (no poverty) and SDG 2 (zero hunger), but also with conceivable positive impact on SDG 3 (good health and well-being), SDG 4 (quality of education) and SDG 10 (reduced inequalities). Moreover, transfers targeting women may be able to play role in ensuring a higher degree of gender equality (SDG 5). The paper by Hidrobo and colleagues is one of the few studies that report the results of a head-to-head comparison of all three modes.

3. Push-button replication

Hidrobo and colleagues used the Stata software package. The authors provided us access to the relevant data and the replication code. We were able to conduct the push-button replication to our satisfaction (see Appendix A for results). The data were obtained via email from Melissa Hidrobo. The path names for the files indicated in the .do file were modified in order to run the code on our local machines. The code was modified to additionally report p-values, as suggested by the 3ie replication manual (Brown et al. 2014). Nothing else was changed in the code provided to us to carry out the push-button replication. After obtaining the results, we identified which results belonged to which tables in the paper based on the commented code, the paper's description of each table and the generated output file. Once we identified the results, we replicated the tables as they appear in the paper. Finally, we compared the coefficients and significance level of each table and reported the results.

The original replication plan (Lhachimi 2017) incorporated comments from (i) several reviewers, (ii) the external adviser and (iii) the authors of the original paper. When drafting the replication plan, the replication team was already in possession of the data and the replication code, but had not yet interacted with the material provided. The push-button replication was the first step in interacting with the data provided by the authors.

4. Pure replication

Hidrobo and colleagues use baseline and endline data collected as part of a cluster-randomized controlled trial to assess and compare the impact of the provision of cash, food vouchers and food on food consumption in northern Ecuador, using different measures of food consumption. In detail, the investigated outcomes in the paper are dietary diversity index (DDI), household dietary diversity score (HDDS), food consumption score (FCS), caloric intake per capita, total monthly per capita consumption, monthly non-food per capita consumption and monthly food per capita consumption.

The statistical analysis in the paper can be broadly grouped in three segments. First, the authors discuss the problem of balancing and attrition, i.e. observations are not randomly distributed and/or lost to follow-up. Hence, they test the success of randomization into treatment and control groups. Second, the authors conduct the main analysis, i.e. the effectiveness of the different interventions for a set of outcome variables. Finally, the authors carry out a number of robustness checks, including the inclusion of baseline characteristics as additional covariates and a further assessment of the potential bias introduced by attrition using Lee bounds.

Hence, for the pure replication we proceed as follows: First, we review the data set and the construction of the several variables, in particular the food indices. We then replicate the tables of the paper using the code provided by the authors. Second, as an additional check, we rebuild the provided Stata code to estimate the main regression results using the statistic software R, i.e. translating the original code into a different statistical language. This approach assists us in understanding each decision made in the original code. We then aim to replicate all the original tables in the paper using the new code. Additionally, we aim to replicate the robustness checks, as we report in Appendix B. Third, we replicate OP-Table 6[1] on cost-effectiveness analysis, for which code was not proved by the authors. In a final step, we conduct several robustness checks by reconstructing the indices used in the analysis and investigated potential changes in outcomes and interpretation.

4.1 Original data and variable construction

The authors provided us with a data set that apparently did not contain the actual raw data. The accompanying .do file – i.e. the programming commands ("code") in Stata-readable format – did not show the steps taken to create the final data set.[2] Several variables based on a combination of underlying variables – e.g. food or asset indices – were already present. Hence, we are unable to investigate any potential steps taken for data cleaning, variable construction, recoding or labeling during the creation of the provided data set.

Nevertheless, we are able to reconstruct the HDDS and FCS. We find minor discrepancies for the latter, potentially due to small differences in the underlying data used to calculate the FCS. We are not able to reconstruct the DDI, as the provided data do not contain all 40 food items used to construct this index. We partly succeed in reconstructing the asset index used for robustness checks.

For the outcomes measuring caloric intake and the value of food and non-food consumption, we are unable to completely trace their creation, as the assignment of calories and prices to the different food items is not shown in the provided .do file, nor is the calculation of the aggregate food and non-food consumption based on the underlying food and non-food groups. Hence, we are unable to replicate the food consumption

[1] We refer to tables in the original paper as "OP-Table" – i.e. "Original Paper Table."

[2] The authors confirmed this in a personal communication.

aggregate and caloric intake, as the provided data do not contain the underlying 40 food items, the median prices assigned to the food groups or their respective caloric values. We are able to exactly replicate the non-food consumption aggregate by summing the value of the 17 non-food groups existent in the data. Furthermore, we are able to trace the provided .do file for further transformation of the aggregates from daily into monthly values and their logarithmic transformation, which are finally used in the analysis.

The data set provided to us is in Stata format. It contains the baseline and endline data, with the variables distinguished by the prefixes *bl* for baseline and *el* for endline data. The baseline data contain 2,357 households, of which 2,087 were both re-interviewed at endline and had complete food consumption data. Hence, those could be included in the analysis. The households were sampled from two provinces, Sucumbíos and Carchi in Northern Ecuador, which consisted of 80 *barrios,* further divided into 145 clusters.

4.2 Replication results

4.2.1 Attrition and balance

We are able to fully replicate the steps taken by the original authors to investigate attrition and balance.

The authors' data set drops all attritted cases, as well as observations where information on food consumption is missing in either or both baseline and endline. The authors check the potential bias arising due to attrition in two ways: (i) by investigating whether attrition rates differed between intervention arms and (ii) by investigating whether attrition in intervention or treatment arms was statistically related to specific variables. To do so, they compare baseline characteristics of households who left the study in each arm, finding that only three characteristics show a significant difference between those who left the control arm and those who left the treatment arm. However, the authors conclude that because these characteristics show no significant differences between treatment and control arms in those who remained in the study, any bias due to attrition is likely to be very small. Table 1 shows that only 26 households (1.1%) are not included in the analysis due to missing food consumption data. We therefore believe a further investigation of the effect of the missing data on the analysis is not warranted.

Then, the authors assess the balance of treatment and control arms by testing for mean differences in household head and household characteristics, as well as all outcome variables between control and treatment arms. Overall, the test does not indicate great imbalances and only four tests indicate statistically significant differences; households in the cash treatment arm are less likely to have a Colombian household head and have a somewhat smaller household size; households in the voucher treatment arm also have a smaller household size and fewer children 6–15 years old.

We are able to replicate these findings using the provided code and do not find issues in the provided code.

Table 1: Number of cases that attritted or had missing food consumption data

	Number of cases	Percent
Included in analysis	2,087	88.88
Attrited	235	10.01
Missing food consumption	26	1.11
Total	2,348	100.00

4.2.2 Main results and robustness checks by the original authors

Main results

With the .do file provided by the authors, we are able to replicate all results presented in the results section of the paper (OP-Tables 2–5 and OP-Table 7), with the exception of OP-Table 6, for which no code was provided. As an additional robustness check, we rebuild the code used to create the estimation sample and to carry out the analysis presented in OP-Tables 1 through 5 and OP-Table 7 in the statistical software R. Again, we are able to exactly reproduce the results.

In order to replicate OP-Table 6, which describes the cost-effectiveness analysis carried out to determine the most cost-effective strategy to increase outcome measures by 15 percent, we use the explanations provided on page 153 of the paper, as shown in Section 4.2.3.

Robustness checks

In Appendix B of the original paper, the authors report several robustness checks. We are able to fully replicate OP-Tables B.1–B.7.

The original paper uses an ANCOVA to identify the effect of the intervention. Because the authors deem randomization to be successful based on a comparison of household and household head characteristics between treatment and control arms, ANCOVA is carried out without controlling for any other baseline characteristics, using a linear regression model. In their robustness analysis, baseline covariates are added back in. The authors conduct two types of robustness checks: (i) Lee bounds and (ii) extend controls (with and without Winsorization). The authors check for heterogeneity of effects when controlling for timing of treatment and for impact of treatments on other transfers.

Lee bounds

Lee bounds are a well-known strategy to deal with non-random attrition (Lee 2009). The main assumption of Lee bounds is monotonicity, i.e. that treatment assignment affects attrition only in one direction. To calculate the Lee bounds, the sample of either the treated or the non-treated observations is trimmed to ensure the share of observations with observed outcome is equal for both groups. Trimming is done to either the upper or lower tails of the outcome distribution. Hence, two extreme scenarios are calculated: in the group that suffers less from attrition, either the largest or the smallest values of the outcome are regarded as “excess observations” and excluded from the analysis. In this application, the outcome distribution in each treatment arm is trimmed by those excess observed once from above and once from below. The assumption is that there are households that would have attritted had they not been assigned to treatment, but no households that attrit because they have received treatment. This seems a plausible assumption, as it would be counterintuitive to expect that households attritted because they were provided with cash, food or vouchers.

We report only our replication of OP-Table B.3, where we find (very minor) discrepancies for the tests of significant differences of the effects between treatment arms (Table 2). However, these differences are marginal and did not lead to any changes in interpretation of the results.

Table 2: Results for OP-Table B.3 using Lee bounds

Original results

	Log caloric intake			HDDS			DDI			FCS		
	Beta	Upper	Lower	Beta	Upper	Lower	Beta	Upper	Lower	Beta	Upper	Lower
Food treatment	0.21	0.26	0.15	0.61	0.85	0.55	2.36	2.93	1.90	6.96	8.86	4.84
	(0.04)***	(0.04)***	(0.04)***	(0.12)***	(0.12)***	(0.12)***	(0.44)***	(0.46)***	(0.42)***	(1.22)***	(1.23)***	(1.16)***
Cash treatment	0.12	0.17	0.08	0.47	0.65	0.43	2.64	3.08	2.32	6.57	7.99	5.06
	(0.04)***	(0.04)***	(0.04)**	(0.11)***	(0.11)***	(0.11)***	(0.42)***	(0.43)***	(0.42)***	(1.29)***	(1.33)***	(1.22)***
Voucher treatment	0.18	0.21	0.15	0.60	0.72	0.57	3.13	3.43	2.89	9.56	10.56	8.65
	(0.04)***	(0.04)***	(0.04)***	(0.12)***	(0.11)***	(0.12)***	(0.45)***	(0.44)***	(0.43)***	(1.39)***	(1.40)***	(1.30)***
R2	0.17	0.17	0.15	0.16	0.17	0.16	0.27	0.27	0.25	0.16	0.16	0.17
N	2,087	2,032	2,030	2,087	2,032	2,029	2,087	2,032	2,029	2,087	2,034	2,029
Baseline Mean	1895.43	1913.43	1855.16	9.18	9.24	9.17	17.27	17.43	17.11	59.86	60.35	59.69
P-value: food = voucher	0.40	0.13	0.95	0.86	0.15	0.85	0.07	0.24	0.01	0.07	0.23	0.00
P-value: cash = voucher	0.15	0.31	0.04	0.16	0.32	0.12	0.22	0.38	0.13	0.05	0.09	0.01
P-value: food = cash	0.03	0.01	0.05	0.12	0.02	0.18	0.48	0.73	0.26	0.77	0.53	0.86

Replicated results

	Log caloric intake			HDDS			DDI			FCS		
	Beta	Upper	Lower	Beta	Upper	Lower	Beta	Upper	Lower	Beta	Upper	Lower
Food treatment	0.21	0.26	0.15	0.61	0.84	0.55	2.36	2.93	**1.88**	6.96	**8.85**	**4.85**
	(0.04)***	(0.04)***	(0.04)***	(0.12)***	(0.12)***	(0.12)***	(0.44)***	(**0.**46)***	**(0.41)*** **	(1.22)***	(1.23)***	(1.16)***
Cash treatment	0.12	0.17	0.08	0.47	0.65	0.43	2.64	3.08	2.32	6.57	7.99	**5.05**
	(0.04)***	(0.04)***	(0.04)**	(0.11)***	(0.11)***	(0.11)***	(0.42)***	(0.43)***	(0.42)***	(1.29)***	**(1.34)*** **	(1.23)***
Voucher treatment	0.18	0.21	0.15	0.60	0.72	**0.58**	3.13	3.43	**2.90**	9.56	**10.55**	8.65
	(0.04)***	(0.04)***	(0.04)***	(0.12)***	(0.11)***	(0.12)***	(0.45)***	(0.44)***	(0.43)***	(1.39)***	(1.40)***	**(1.29)*** **
R2	0.17	0.17	0.15	0.16	**0.18**	0.16	0.27	0.27	0.25	0.16	0.16	0.17
N	2,087	2,032	2,030	2,087	2,032	2,029	2,087	2,032	2,029	2,087	2,034	2,029
Baseline Mean	1895.43	1913.43	1855.16	9.18	9.24	9.17	17.27	17.43	**17.12**	59.86	60.35	**59.70**
P-value: food = voucher	0.40	0.13	0.95	0.86	0.15	**0.81**	0.07	0.24	0.01	0.07	0.23	0.00
P-value: cash = voucher	0.15	0.31	0.04	0.16	**0.34**	**0.10**	0.22	0.38	0.13	0.05	0.09	0.01
P-value: food = cash	0.03	0.01	0.05	0.12	0.02	**0.17**	0.48	**0.72**	**0.23**	0.77	0.53	**0.87**

Note: Standard errors in parenthesis clustered at the cluster level. All estimations control for baseline outcome variable and province. Differing values in bold.
* p < 0.1 ** p < 0.05; *** p < 0.01.

Extended controls and Winsorizations
We are able to fully replicate OP-Table B.4 and B.5, which report additional robustness checks. To investigate any potential biases arising from imbalances at baseline, the authors re-estimate their regressions, including baseline control variables for age, gender, nationality, any secondary education of the household head, the number of children 0–5 years and 6–15 years living in the household, and household wealth quintiles. Hidrobo and colleagues pointed out to us that the editor of the journal explicitly advised against the inclusion of additional covariates in the main analysis; hence, this analysis is part of the robustness checks in the appendix[1]. Moreover, the authors conducted a 1 percent-Winsorization,that is, converting the bottom and top 1 percent observation to the value of the 1st and 99th percentile, respectively. This is a well-established method to investigate and mitigate the effect of outliers – i.e. unusually large observations.

Heterogeneity with respect to timing and impact of treatment on transfers
Because they find a consistently smaller impact of cash on the value of consumption than the other treatments (OP-Table B.4), the authors investigate two potential explanations for this finding. First, they investigate whether this could be explained by differences in how quickly cash was used for consumption, compared to food or vouchers. The worry is, if cash is consumed more quickly, its effects may not be captured in the 7-day food consumption recall, which took place 3 to 17 days after the last transfer in Sucumbíos and 18 to 30 days after the last transfer in Carchi. Creating indicators for a relatively later recall (after 1 week in Sucumbíos and after 3 weeks in Carchi), they investigate differences in the estimated effects for those being surveyed before and after the later recall cut-off. They do not find that timing of the survey affected the results; thus, recall timing would not explain the smaller effects of cash transfers on consumption (OP-Table B.6). We are able to exactly replicate this table using the .do file provided to us by the authors.

Second, the paper investigates whether cash transfers could have crowded out other transfers, so that overall the available budget used for consumption increased by less than the full amount of the cash transfer. The results shown in OP-Table B.7 support this, finding that households receiving cash were significantly less likely to receive loans or cash, which is not the case for the food and voucher households.

4.2.3 Replication of cost-effectiveness analysis
In OP-Table 6, the authors report a cost-effectiveness analysis that is not included in the code they provided to us. However, we are able to replicate it using the information provided in the paper.

An integral part of the paper is the cost-effectiveness analysis, simulating how much additional cost would occur to achieve a 15 percent increase in food security outcomes – log caloric intake (per capita), HDDS, DDI, FCS and poor food consumption (PFC) using each treatment. In the original paper, using vouchers appears to be the most cost-effective option for all outcomes, followed closely by cash, while direct food provision appears to be by far the least cost-effective.

[1] The authors confirmed this in a personal communication

For the replication, we first calculated the percentage increase in each outcome for each treatment, using the coefficients shown in OP-Tables 2 and 3. For log estimates (log caloric intake), the actual coefficient could be interpreted as a percentage change; for the other estimates (HDDI, DDI, FCS, PFC), we divide them by the baseline mean also presented for each outcome in OP-Table 3 and then multiply by 100 to arrive at the percentage change. We then round the percentage to the closest full number. We then divide the 15 percent by the percentage change achieved with the current treatments, and multiply this by the cost per transfer, shown in OP-Table B.8 of the original paper's supplementary material. For example, to calculate the costs for a 15 percent increase in the food consumption score due to the delivery of cash, we first calculate the percentage change due to the current treatment: (6.57 / 59.86) * 100 = 10.98, which we round to 11. We then calculate (15 / 11) * 2.99 = 4.08, giving us exactly the number stated in OP-Table 6.

4.2.4 Replication of the HDDS

The HDDS differs from DDI, which sums the number of distinct food items consumed by the household in the previous seven days, in that frequency is measured across standardized food groups instead of individual food items. The HDDS is calculated by summing the number of food groups consumed in the previous seven days from the following 12 groups: cereals; roots and tubers; vegetables; fruits; meat, poultry and offal; eggs; fish and seafood; pulses, legumes and nuts; milk and milk products; oils and fats; sugar and honey; and miscellaneous (Kennedy et al. 2011).

Using the information from the original paper on the food groups used, we are able to replicate the HDDS index exactly for the endline values and find only minor deviations for the baseline values (Table 3). We therefore do not replicate the original results using the reconstructed index.

Table 3: Original and reconstructed HDDS index

	N	Mean	SD	Min.	Max.
Original index					
HDDS (baseline)	2,087	9.18	1.81	1	12
HDDS (endline)	2,087	10.69	1.56	1	12
Reconstructed index					
HDDS (baseline)	2,087	9.18	1.82	1	12
HDDS (endline)	2,087	10.69	1.56	1	12

4.2.5 Replication of the FCS

According to the original paper, the FCS is calculated by summing the number of days the household consumed each of the following food groups (staples, pulses, vegetables, fruit, meat and fish, milk and dairy, sugar and honey, oils and fats), multiplying the number of days by the food group's weighted frequencies and summing across categories to obtain a single proxy indicator. The authors then categorize households as having poor to borderline consumption if their FCS score is less than or equal to 35 (WFP 2008).

Although we are able to calculate an FCS, we are unable to exactly replicate the index as in the original paper (Table 4). Our FCS diverges for 1,143 observations, although it remains the same for 944 observations. For observations that diverge, the index is always somewhat higher. We are unable to find an explanation for these differences, looking for potential patterns that would emerge when comparing those with divergent and non-divergent scores. In particular, we compared both groups in terms of their descriptive statistics, looked for missing values in the calculation of the FCS and tried to find systematic differences in the underlying components used to calculate the FCS, but did not discover any obvious patterns. However, given the number of operations required to calculate the FCS and the, even slight, divergences in how the score was calculated in the original analysis could have led to relatively large differences in the FCS and be the reason for the discrepancies we found.

When we used our FCS to create a binary variable indicating PFC and using the threshold of 35, our created index has a lower prevalence of PFC than the FCS in the original table. Overall, 68 observations that previously were below the threshold of 35 now exceeded it at baseline. On average when using the original index, the FCS for these 68 observations is 32, whereas with the reconstructed index it is 40. Because of these differences in scores, we re-estimated all tables, investigating the effects of the treatments on the FCS and PFC.

Table 4: Original and reconstructed FCS index

	N	Mean	SD	Min.	Max.
Original index					
FCS (baseline)	2,087	59.86	20.11	1	112
FCS (endline)	2,087	67.28	20.07	4.5	112
PFCS (baseline)	2,087	0.11	0.32	0	1
PFCS (endline)	2,087	0.05	0.22	0	1
Reconstructed index					
FCS (baseline)	2,087	64.58	20.05	1	112
FCS (endline)	2,087	73.09	19.75	4.5	112
PFCS (baseline)	2,087	0.08	0.27	0	1
PFCS (endline)	2,087	0.04	0.19	0	1

Results

We start with the results reported in OP-Table 3, showing the effects of the treatments on food security outcomes (Table 5). We limit it to displaying the results for the FCS, comparing the original results and the results using the reconstructed index. Overall, the results using the reconstructed index support the findings from the original analysis. The main differences are that now, cash treatment also has a statistically significant effect on PFC, similar in size to the two other treatments, whereas the paper found that “the size of the decrease is significantly larger for the food arm when compared with the cash arm” (Hidrobo et al. 2014, p.150).

Table 5: Results for OP-Table 3 of the impact of treatment arms on food consumption score using reconstructed FCS

	Original results		Results using reconstructed FCS	
	FCS	PFC	FCS	PFC
Food treatment	6.96	–0.05	7.80	–0.04
	(1.22)***	(0.02)***	(1.29)***	(0.01)***
Cash treatment	6.57	–0.02	6.58	–0.03
	(1.29)***	(0.02)	(1.31)***	(0.01)**
Voucher treatment	9.56	–0.04	9.23	–0.03
	(1.39)***	(0.02)***	(1.40)***	(0.01)**
R^2	0.16	0.08	0.19	0.07
N	2,087	2,087	2,087	2,087
Baseline mean	59.86	0.11	64.58	0.08
P-value: food = voucher	0.07	0.73	0.30	0.45
P-value: cash = voucher	0.05	0.13	0.06	0.79
P-value: food = cash	0.77	0.09	0.36	0.30

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province.

Robustness

The next table in the paper to use the FCS is OP-Table B.3, showing the use of Lee bounds to investigate the robustness of the results to attrition. Our Table 6 shows that estimates using the reconstructed FCS diverge somewhat from the original results; nonetheless, they still support the original conclusions that even the lower bounds suggest large and significant impacts across all treatment arms.

Table 6: Results for OP-Table B.3 using Lee bounds and reconstructed FCS

	FCS (original)			FCS (reconstructed)		
	Beta	Upper	Lower	Beta	Upper	Lower
Food treatment	6.96	8.86	4.85	7.80	9.87	6.00
	(1.22)***	(1.23)***	(1.16)***	(1.29)***	(1.28)***	(1.31)***
Cash treatment	6.57	7.99	5.06	6.58	8.10	5.17
	(1.29)***	(1.33)***	(1.22)***	(1.31)***	(1.34)***	(1.28)***
Voucher treatment	9.56	10.56	8.65	9.23	10.32	8.56
	(1.39)***	(1.40)***	(1.30)***	(1.40)***	(1.39)***	(1.36)***
R^2	0.16	0.16	0.17	0.19	0.19	0.19
N	2,087	2,034	2,029	2,087	2,034	2,029
Baseline mean	59.86	60.35	59.69	64.58	65.12	64.33
P-value: food = voucher	0.07	0.23	0.00	0.30	0.73	0.06
P-value: cash = voucher	0.05	0.09	0.01	0.06	0.11	0.01
P-value: food = cash	0.77	0.53	0.86	0.36	0.18	0.53

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province.

We also re-estimate OP-Table B.5 using the reconstructed FCS, which shows the results of adding additional control variables and of Winsorizing the tails (Table 7). Again, overall the use of the reconstructed FCS leads to only slight changes and generally supports the findings from the original paper, although delivering less evidence that vouchers lead to higher FCS than the other treatments, food in particular.

Table 7: Results for OP-Table B.5 using reconstructed FCS

	Original results					Results using reconstructed FCS				
	FCS			PFC		FCS			PFC	
	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls
Food treatment	6.96	6.80	6.80	–0.05	–0.05	7.80	7.60	7.61	–0.04	–0.04
	(1.22)***	(1.24)***	(1.23)***	(0.02)***	(0.02)***	(1.29)***	(1.29)***	(1.29)***	(0.01)***	(0.01)***
Cash treatment	6.57	6.58	6.56	–0.02	–0.03	6.58	6.54	6.53	–0.03	–0.03
	(1.29)***	(1.25)***	(1.24)***	(0.02)	(0.02)*	(1.31)***	(1.23)***	(1.23)***	(0.01)**	(0.01)**
Voucher treatment	9.56	9.40	9.42	–0.04	–0.05	9.23	9.02	9.04	–0.03	–0.03
	(1.39)***	(1.43)***	(1.42)***	(0.02)***	(0.02)***	(1.40)***	(1.42)***	(1.40)***	(0.01)**	(0.01)**
R^2	0.16	0.18	0.18	0.08	0.10	0.19	0.21	0.21	0.07	0.09
N	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087
Baseline mean	59.86	59.86	59.88	0.11	0.11	64.58	64.58	64.60	0.08	0.08
P-value: food = voucher	0.07	0.08	0.08	0.73	0.77	0.30	0.33	0.32	0.45	0.51
P-value: cash = voucher	0.05	0.07	0.07	0.13	0.16	0.06	0.09	0.08	0.79	0.86
P-value: food = cash	0.77	0.87	0.86	0.09	0.12	0.36	0.43	0.42	0.30	0.39

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province. Extended controls include household head's characteristics (age, gender, nationality, education), number of children and household asset index quintiles.

OP-Table B.6 investigates whether differences across treatments in how quickly (or slowly) transfers were consumed could have led to an underestimation of the impact of cash. Re-estimating this using the reconstructed FCS, the results support the original conclusion of no differential effects for the FCS (Table 8). Table 9 replicates the results of the second part of OP-Table B.6, showing coefficients for those with more than 1 or 3 weeks since receiving the transfer. The replication using the reconstructed FCS leads to similar conclusions overall; however, food transfers now tend to have a somewhat bigger effect compared to the original FCS, which also causes the difference between food transfers and vouchers in Sucumbíos, found in the original paper, to decrease.[1]

Table 8: Results for OP-Table B.6 on impact of treatment arms on consumption outcomes, by days of intervention

	FCS (original)		FCS (reconstructed)	
	Sucumbíos	Carchi	Sucumbíos	Carchi
Food treatment	7.23	8.67	7.48	10.08
	(2.15)***	(2.04)***	(1.81)***	(1.59)***
Cash treatment	4.89	6.21	5.82	6.33
	(1.97)**	(2.64)**	(2.31)**	(2.61)**
Voucher treatment	7.17	9.26	7.16	8.68
	(2.78)**	(2.40)***	(2.90)**	(2.55)***
Food X More than 1 week since treatment	–1.26		–0.83	
	(2.67)		(2.42)	
Cash X More than 1 week since treatment	1.53		0.20	
	(2.13)		(2.19)	
Voucher X More than 1 week since treatment	4.08		3.33	
	(2.61)		(2.62)	
Food X More than 3 weeks since treatment		–1.82		–1.98
		(2.27)		(2.02)
Cash X More than 3 weeks since treatment		1.83		1.76
		(2.65)		(2.54)
Voucher X More than 3 weeks since treatment		–1.52		–0.66
		(2.89)		(3.08)
N	1,277	810	1,277	810
P-value: food = voucher	0.99	0.81	0.91	0.55
P-value: cash = voucher	0.46	0.31	0.69	0.45
P-value: food = cash	0.36	0.36	0.49	0.12

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province.

[1] Using the reconstructed FCS, the p-value for comparing the effects of food and voucher indicates no significant difference anymore (p = 0.11), in contrast with the original analysis, where the p-value was 0.02, i.e. indicating a statistically significant difference (Table 8). However, comparing these coefficients across models, i.e. FCS (original) versus FCS (reconstructed) – statistical testing indicates no significantly different coefficients for food treatment or for voucher treatment.

Table 9: Results for OP-Table B.6 on impact of treatment arms on consumption outcomes, coefficient for those with more than 1 or 3 weeks since treatment

	FCS (original)		FCS (reconstructed)	
	Sucumbíos	Carchi	Sucumbíos	Carchi
Food treatment	5.97	6.85	6.65	8.10
	(2.01)***	(2.13)***	(2.28)***	(2.25)***
Cash treatment	6.42	8.04	6.02	8.09
	(1.71)***	(2.46)***	(1.63)***	(2.33)***
Voucher treatment	11.25	7.74	10.49	8.02
	(1.72)***	(2.47)***	(1.75)***	(2.45)***
N	1,277	810	1,277	810
P-value: food = voucher	0.02	0.73	0.11	0.98
P-value: cash = voucher	0.02	0.92	0.01	0.98
P-value: food = cash	0.84	0.65	0.78	1.00

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province.

Finally, we also replicate the cost-effectiveness analysis, using the reconstructed FCS, and the coefficients and baseline estimates presented in OP-Table 6. Table 10 displays the results of the cost-effectiveness analysis, showing that cost-effectiveness does not change for food, although costs increase somewhat for cash and vouchers. Vouchers, however, remain the most cost-effective, followed closely by cash, while food is the least cost-effective, supporting the conclusions of the original paper.

Table 10: Results for OP-Table 6 on cost-effectiveness, using reconstructed FCS

	Food ($)	Cash ($)	Voucher ($)
FCS (original)	14.33	4.08	3.07
FCS (reconstructed)	14.33	4.49	3.50

Note: Modality-specific costs per transfer are used to calculate the cost of increasing each outcome by 15 percent.

4.2.6 Replication of non-food consumption outcomes

We are not able to replicate the creation of the food consumption outcomes, as the data set does not provide the necessary information. We can, however, replicate the non-food consumption outcome. According to the original paper, this is calculated from the value of items purchased or acquired in the last month or 3 months for the following 17 items: personal care, home and kitchenware, communication (telephone and Internet), electricity and gas, transportation, water, housing (rent and repairs), entertainment, beauty services, clothes and shoes for adult males, clothes and shoes for adult females, clothes and shoes for children, furniture and electronics, jewelry, toys, education and tobacco. The data set contains the necessary information on the value of purchased goods in each non-food group and, by summing all 17 groups, we are able to exactly replicate the food consumption aggregate used for the study.

4.2.7 Replication of asset index

We then construct the asset index using the *pca* command in Stata. The resulting index is well-correlated (0.8281) with the original index but does not fully replicate it (Table 11).

The asset index (also sometimes referred to in the paper as the wealth index) is used as an additional baseline control variable as part of the robustness checks in the paper. The original paper states (Hidrobo et al. 2014, p.149),

> The household wealth quintiles are constructed from a wealth index that is created using the first principal from a principal components analysis (PCA). Variables used to construct the index are housing infrastructure indicators (e.g., type of floor, roof, toilet, light, fuel, and water source) and 11 asset indicators (e.g., refrigerator, mobile phone, TV, car, and computer).

Because only examples of included indicators are given, it is unclear which additional indicators were used to create the index. Because the exact number of household indicators is not stated, we choose to limit them to the ones mentioned: type of floor, roof, toilet, light, fuel and water source. For the asset indicators, we use those mentioned in the paper: refrigerator, mobile phone, TV, car and computer, plus 6 additional indicators from the 15 indicators available in the data: owning land, washing machine, agricultural tools, big animals, a separate kitchen and electric fan. Other potential indicators would have been owning small animals, a microwave, a bicycle. Because we do not know the actual assets used, the choice can be regarded as arbitrary.[1]

Table 11: Original and reconstructed asset index

	N	Mean	SD	Min	Max
Original asset index (baseline)	2,087	0.10	1.92	–5.16	4.98
Reconstructed asset index (endline)	2,087	0.07	1.50	–4.75	3.21

Robustness

To test whether the results presented in OP-Tables B.4 and B.5 are robust to the calculation of the asset index, we replicate the results using the reconstructed asset index. Using this indicator, results change only marginally, not leading to different interpretations (Table 12, Table 13 and Table 14).

Conclusion

In this pure replication, we review the original code provided to us by the original authors and reconstruct the code in the R language. Neither exercise suggests any concerns regarding coding decisions and both lead to similar results as the original paper. Because some of the variables in the data set had been constructed beforehand, we aimed to reconstruct several of the outcome indices and the wealth index based on

[1] As an additional robustness check, we create several additional indices, varying the type and number of additional asset items. Including additional items only increases the correlation coefficient of the indices with the original index up to 0.94 (when including several additional indicators on the number of rooms, if household has a bedroom, the general state of the home and if the house is shared with other families). We also check whether replacing three assets (having a big animal, a kitchen, electric fan) with the three assets that had been initially left out by us (small animals, a microwave, a bicycle) would change the correlation of the indices. This leads to a slight increase in the correlation with the original index. Overall, we conclude that it is very unlikely that using a different combination of assets would lead to substantially different results, due to the high correlation of the reconstructed indices with the original index.

information from the original paper, provided that the needed underlying data were available in the provided data set.

The reconstructed HDDS shows very minor discrepancies, which did not lead to important changes in the regression results. Larger discrepancies are found for the FCS for about half of the observations, which also reduces the prevalence of PFC. As a result, the regression analysis using this reconstructed PFC indicator finds cash to be as effective as food in reducing PFC, whereas the original analysis suggests food provision is superior. Unfortunately, because we do not know how the original FCS was created and due to the relative complexity of its creation, the underlying reasons for these discrepancies are unclear to us. We therefore cannot make any further claims about the differences we have found.

In a final step, we are able to replicate the cost-effectiveness analysis by following the steps outlined in the original paper and the information provided in the supplementary material. The results of this pure replication indicate no significant concerns with the original analysis and support the conclusions of the original paper.

Table 12: Results for OP-Table B4 Panel A on impact of treatment arms on consumption outcomes, robustness checks (reconstructed asset index)

Original results

	Food consumption (per capita)			Non-food consumption (per capita)			Total consumption (per capita)		
	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.
Food treatment	9.22	9.11	9.18	9.22	8.75	7.85	18.50	17.65	16.86
	(2.79)***	(2.55)***	(2.47)***	(3.30)***	(3.19)***	(3.08)**	(5.02)***	(4.76)***	(4.65)***
Cash treatment	5.47	4.59	4.49	6.81	6.08	3.62	12.66	10.65	8.58
	(2.56)**	(2.23)**	(2.10)**	(3.93)*	(3.97)	(3.13)	(5.09)**	(4.74)**	(4.27)**
Voucher treatment	6.38	5.09	5.17	6.78	6.11	5.92	13.45	11.03	10.89
	(2.58)**	(2.32)**	(2.20)**	(2.82)**	(2.72)**	(2.67)**	(4.38)***	(4.07)***	(3.98)***
R^2	0.21	0.27	0.27	0.17	0.19	0.25	0.22	0.25	0.27
N	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087
Baseline mean	47.54	47.54	47.33	64.29	64.29	63.81	111.83	111.83	111.46
P-value: food = voucher	0.31	0.12	0.12	0.46	0.41	0.53	0.33	0.17	0.20
P-value: cash = voucher	0.73	0.83	0.76	0.99	0.99	0.45	0.88	0.94	0.59
P-value: food = cash	0.17	0.07	0.06	0.57	0.53	0.20	0.30	0.18	0.08

Replication results

	Food consumption (per capita)			Non-food consumption (per capita)			Total consumption (per capita)		
	Main	Extended controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.
Food treatment	9.22	**9.20**	**9.26**	9.22	**8.80**	**7.86**	18.50	**17.91**	**17.07**
	(2.79)***	**(2.57)*** **	**(2.49)*** **	(3.30)***	**(3.20)*** **	**(3.07)** **	(5.02)***	**(4.81)*** **	**(4.70)*** **
Cash treatment	5.47	**4.66**	**4.56**	6.81	**6.40**	**3.82**	12.66	**11.16**	**8.99**
	(2.56)**	**(2.27)** **	**(2.13)** **	(3.93)*	**(4.04)**	**(3.12)**	(5.09)**	**(4.83)** **	**(4.31)** **
Voucher treatment	6.38	**5.12**	**5.20**	6.78	**6.50**	**6.15**	13.45	**11.53**	**11.27**
	(2.58)**	**(2.31)** **	(2.20)**	(2.82)**	(2.72)**	**(2.66)** **	(4.38)***	(4.05)***	(3.97)***
R^2	0.21	0.27	0.27	0.17	0.19	**0.26**	0.22	0.25	0.27
N	2087	2087	2087	2087	2087	2087	2087	2087	2087
Baseline mean	47.54	47.54	47.33	64.29	64.29	63.81	111.83	111.83	111.46
P-value: food = voucher	0.31	0.12	**0.11**	0.46	**0.48**	**0.58**	0.33	**0.19**	**0.22**
P-value: cash = voucher	0.73	**0.85**	**0.78**	0.99	**0.98**	0.45	0.88	0.94	**0.60**
P-value: food = cash	0.17	**0.08**	0.06	0.57	**0.57**	**0.22**	0.30	**0.20**	**0.09**

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province. Extended controls include household head's characteristics (age, gender, nationality, education), number of children, and household asset index quintiles. Differing values in bold.

Table 13: Results for OP-Table B4 Panel B on impact of treatment arms on log consumption outcomes, robustness checks (reconstructed asset index)

Original results
Panel B

	Log food consumption (per capita)			Log non-food consumption (per capita)			Log total consumption (per capita)		
	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.
Food treatment	0.20	0.20	0.19	0.15	0.14	0.14	0.17	0.17	0.16
	(0.04)***	(0.04)***	(0.04)***	(0.07)**	(0.07)**	(0.07)**	(0.05)***	(0.05)***	(0.04)***
Cash treatment	0.14	0.13	0.12	0.07	0.06	0.06	0.11	0.10	0.09
	(0.04)***	(0.04)***	(0.04)***	(0.06)	(0.06)	(0.06)	(0.04)***	(0.04)**	(0.04)**
Voucher treatment	0.15	0.13	0.13	0.13	0.12	0.12	0.13	0.11	0.11
	(0.04)***	(0.04)***	(0.04)***	(0.06)**	(0.05)**	(0.05)**	(0.04)***	(0.04)***	(0.04)***
R^2	0.26	0.31	0.31	0.25	0.27	0.28	0.25	0.29	0.29
N	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087
Baseline mean	47.54	47.54	47.33	64.29	64.29	63.81	111.83	111.83	111.46
P-value: food = voucher	0.23	0.10	0.10	0.75	0.76	0.78	0.40	0.24	0.24
P-value: cash = voucher	0.80	0.94	0.95	0.35	0.29	0.27	0.63	0.65	0.62
P-value: food = cash	0.14	0.08	0.09	0.27	0.23	0.22	0.21	0.11	0.11

Replication results

	Log food consumption (per capita)			Log non-food consumption (per capita)			Log total consumption (per capita)		
	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.
Food treatment	0.20	0.20	0.19	0.15	0.14	0.14	0.17	0.17	0.16
	(0.04)***	(0.04)***	(0.04)***	(0.07)**	(0.07)**	(0.07)**	(0.05)***	(0.05)***	(0.04)***
Cash treatment	0.14	0.13	0.12	0.07	0.06	0.06	0.11	0.10	**0.10**
	(0.04)***	(0.04)***	(0.04)***	(0.06)	(0.06)	(0.06)	(0.04)***	(0.04)**	(0.04)**
Voucher treatment	0.15	0.13	**0.12**	0.13	**0.13**	**0.13**	0.13	**0.12**	0.11
	(0.04)***	(0.04)***	(0.04)***	(0.06)**	(0.05)**	(0.05)**	(0.04)***	(0.04)***	(0.04)***
R^2	0.26	0.31	0.31	0.25	0.27	0.28	0.25	0.29	0.29
N	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087
Baseline mean	47.54	47.54	47.33	64.29	64.29	63.81	111.83	111.83	111.46
P-value: food = voucher	0.23	**0.09**	0.10	0.75	**0.80**	**0.83**	0.40	**0.25**	**0.25**
P-value: cash = voucher	0.80	**0.95**	0.95	0.35	0.29	0.27	0.63	**0.66**	**0.63**
P-value: food = cash	0.14	0.08	0.09	0.27	**0.24**	**0.24**	0.21	**0.13**	**0.12**

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province. Extended controls include household head's characteristics (age, gender, nationality, education), number of children, and household asset index quintiles. Differing values in bold.

Table 14: Results for OP-Table B5 on impact of treatment arms on food security outcomes, robustness checks (reconstructed asset index)

Original results

	Log caloric intake (per capita)			DDI			FCS			HDDS		PFC	
	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Main	Ext. controls
Food treatment	0.21	0.20	0.19	2.36	2.25	2.24	6.96	6.80	6.80	0.61	0.59	–0.05	–0.05
	(0.04)***	(0.04)***	(0.03)***	(0.44)***	(0.44)***	(0.43)***	(1.22)***	(1.24)***	(1.23)***	(0.12)***	(0.12)***	(0.02)***	(0.02)***
Cash treatment	0.12	0.10	0.09	2.64	2.56	2.54	6.57	6.58	6.56	0.47	0.46	–0.02	–0.03
	(0.04)***	(0.03)***	(0.03)***	(0.42)***	(0.41)***	(0.40)***	(1.29)***	(1.25)***	(1.24)***	(0.11)***	(0.11)***	(0.02)	(0.02)*
Voucher treatment	0.18	0.14	0.13	3.13	3.02	2.98	9.56	9.40	9.42	0.60	0.58	–0.04	–0.05
	(0.04)***	(0.03)***	(0.03)***	(0.45)***	(0.45)***	(0.44)***	(1.39)***	(1.43)***	(1.42)***	(0.12)***	(0.12)***	(0.02)***	(0.02)***
R^2	0.17	0.24	0.27	0.27	0.29	0.29	0.16	0.18	0.18	0.16	0.18	0.08	0.10
N	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087
Baseline mean	1895.43	1895.43	1833.43	17.27	17.27	17.27	59.86	59.86	59.88	9.18	9.18	0.11	0.11
P-value: food = voucher	0.40	0.10	0.10	0.07	0.08	0.09	0.07	0.08	0.08	0.86	0.91	0.73	0.77
P-value: cash = voucher	0.15	0.18	0.18	0.22	0.27	0.27	0.05	0.07	0.07	0.16	0.18	0.13	0.16
P-value: food = cash	0.03	0.00	0.01	0.48	0.44	0.46	0.77	0.87	0.86	0.12	0.18	0.09	0.12

Replication results

	Log caloric intake (per capita)			DDI			FCS			HDDS		PFC	
	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Ext. controls + Winsoriz.	Main	Ext. controls	Main	Ext. controls
Food treatment	0.21	**0.21**	0.19	2.36	2.25	2.24	6.96	**6.73**	**6.73**	0.61	0.59	–0.05	–0.05
	(0.04)***	(0.04)***	(0.03)***	(0.44)***	**(0.43)*** **	(0.43)***	(1.22)***	(1.24)***	(1.23)***	(0.12)***	(0.12)***	(0.02)***	(0.02)***
Cash treatment	0.12	**0.11**	**0.10**	2.64	2.56	**2.53**	6.57	**6.60**	**6.58**	0.47	0.46	–0.02	–0.03
	(0.04)***	(0.03)***	(0.03)***	(0.42)***	(0.41)***	(0.40)***	(1.29)***	**(1.27)*** **	**(1.27)*** **	(0.11)***	(0.11)***	(0.02)	(0.02)*
Voucher treatment	0.18	**0.15**	**0.14**	3.13	**3.03**	**2.99**	9.56	**9.42**	**9.44**	0.60	**0.59**	–0.04	–0.05
	(0.04)***	(0.03)***	(0.03)***	(0.45)***	(0.45)***	(0.44)***	(1.39)***	(1.43)***	(1.42)***	(0.12)***	(0.12)***	(0.02)***	**(0.01)*** **
R^2	0.17	0.23	0.27	0.27	0.29	0.29	0.16	0.18	0.18	0.16	**0.19**	0.08	0.10
N	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087	2,087
Baseline mean	1895.43	1895.43	1833.43	17.27	17.27	17.27	59.86	59.86	59.88	9.18	9.18	0.11	0.11
P-value: food = voucher	0.40	0.10	**0.11**	0.07	0.08	**0.08**	0.07	**0.07**	**0.07**	0.86	**0.95**	0.73	0.77
P-value: cash = voucher	0.15	**0.20**	**0.21**	0.22	**0.26**	**0.26**	0.05	0.07	0.07	0.16	0.18	0.13	0.16
P-value: food = cash	0.03	**0.01**	0.01	0.48	**0.45**	**0.47**	0.77	**0.92**	**0.91**	0.12	**0.19**	0.09	**0.11**

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province. Extended controls include household head's characteristics (age, gender, nationality, education), number of children, and household asset index quintiles. Differing values in bold.

5. Measurement and estimation analysis

This measurement and estimation analysis (Brown et al. 2014) presents additional robustness checks of the original results by testing some of the original assumptions. The robustness checks are pre-specified in the replication plan (Lhachimi 2017). In Section 5.1, we investigate some potential consequences of the sampling frame—in particular, contamination between clusters (Chow et al. 2004; Keogh-Brown et al. 2007). Moreover, we check whether there are interaction effects between province and treatment. In Section 5.2, we employ a competing model design that explicitly accounts for the hierarchical nature of the data – a generalized linear mixed model (GLMM).

5.1 Sampling frame

5.1.1 Contamination of clusters

The sampling strategy of the original paper used geographical units (*barrios*). A *barrio* in the treatment arm, however, can contain up to five clusters (Table 16). On average, clusters had between 14 and 18 observations (Table 17). In most cases, not all clusters in the same *barrio* received the same treatment. Hence, most *barrios* included clusters receiving different treatments, i.e. cash, food or vouchers (Table 18). Therefore, in *barrios* with different treatments, the food consumption behavior of people receiving food or vouchers could have been influenced by observing other households receiving cash. However, we do not think this concern is warranted, as there is no obvious reason why food households would have been influenced in their food consumption by someone else's receiving cash or vouchers instead.

We think, however, that the composition of *barrios* with one or more treatment modes could have affected the treatment preferences investigated in OP-Table 7. The geographical proximity of other treatment modes – e.g. of a cluster receiving cash next to a cluster receiving food – could have affected the opinion about the preferred treatment. We therefore re-estimate OP-Table 7 for *barrios* where only one treatment mode was present (Table 19) and for *barrios* where two or more treatment modes were present (Table 20). Overall, results in both tables are very similar to OP-Table 7, suggesting that the contamination bias, if any, is very small.

A t-test of the mean preferences in *barrios* with only one versus *barrios* with more treatment modes shows that a significant difference exists only for vouchers, with significantly more households in *barrios* with two or more treatment modes preferring to no longer receive any share of their transfers in the form of vouchers (Table 21). Interestingly, when looking at the baseline means of those living in *barrios* with one treatment mode, compared to those living in *barrios* with either two or more treatment modes, we find that the latter had significantly higher levels of food and non-food consumption (Table 15). This could suggest that their preference for transfer modes other than vouchers is not due to contamination due to other treatments modes in the same *barrio*, but rather is based on their different baseline characteristics; e.g. the higher food consumption levels at baseline could indicate a lower need for additional food consumption, making it less attractive to receive treatments that could be used solely to consume food.

Table 15: Baseline means by the number of distinct treatments within one *barrio*

	N	Means: One treatment	Means: Two or more treatments	P-value of diff. One treatment - two or more treatments
Household head is female	1,525	0.28	0.28	0.96
Household head is Colombian	1,525	0.26	0.25	0.79
Sucumbíos province	1,525	1.61	1.58	0.76
Household head is married	1,525	0.26	0.30	0.18
Age of household head	1,525	41.20	41.92	0.48
Household head secondary edu. or higher	1,525	0.32	0.39	0.05
Number of children 0–5 years	1,525	0.72	0.56	0.00
Number of children 6–15 years	1,525	0.98	0.80	0.03
Household size	1,525	4.03	3.68	0.01
Floor type: dirt	1,525	0.05	0.03	0.22
Owns television	1,525	0.80	0.81	0.72
Owns computer	1,525	0.28	0.31	0.42
Owns mobile phone	1,525	0.85	0.80	0.08
Owns car/truck/motorcycle	1,525	0.23	0.24	0.83
Owns land	1,525	0.14	0.11	0.26
Dietary diversity index	1,525	16.99	17.62	0.11
Household dietary diversity score	1,525	9.13	9.26	0.31
Food consumption score	1,525	58.65	61.15	0.06
Total consumption per capita (monthly)	1,525	102.10	118.71	0.00
Non-food consumption per capita (monthly)	1,525	57.75	67.75	0.01
Food consumption per capita (monthly)	1,525	44.34	50.95	0.00
Caloric intake per capita (daily)	1,525	1,812.11	1,873.27	0.43
P-value from joint F-test			0.00	

Notes: P-values are reported from tests on the equality of means for each variable. Standard errors are clustered at the cluster level. F-tests of joint significance: test of joint significance in regression of respective treatment dummies on all 18 baseline variables.

Table 16: Number of clusters that comprise *barrios* (only treatment *barrios*)

Number of clusters per *barrio*	Overall number of observations	Number of clusters
1	477	34
2	357	26
3	306	21
4	320	24
5	65	5
Total	1525	110

Table 17: Number of observations per cluster

	Control			Food			Cash			Voucher		
	mean	min	max	mean	min	max	mean	min	max	mean	min	max
	17.83	4	26	18.36	3	26	15.23	4	24	14.07	5	20
N	562			413			539			573		

Table 18: Number of households receiving treatment in *barrio* with one, two or more treatment modes

	One treatment per *barrio*		Two treatments per *barrio*		Three treatments per *barrio*	
	Number of observations	Number of clusters	Number of observations	Number of clusters	Number of observations	Number of clusters
Food	207	12	92	6	114	8
Cash	185	15	233	15	121	9
Voucher	213	16	266	20	94	9

Table 19: Transfer preference by treatment status (1 treatment mode in *barrio*)

	Means			P-value of diff.		
	Food	Cash	Voucher	Food - Cash	Food - Voucher	Cash - Voucher
All	0.56	0.76	0.59	0.02	0.72	0.02
None	0.25	0.07	0.25	0.01	0.99	0.00
N	193	148	157			

Note: P-values are reported from Wald tests on the equality of means for each variable.

Table 20: Transfer preference by treatment status (two or more treatment modes in *barrio*)

	Means			P-value of diff.		
	Food	Cash	Voucher	Food - Cash	Food - Voucher	Cash - Voucher
All	0.55	0.78	0.53	0.00	0.88	0.00
None	0.29	0.10	0.35	0.00	0.38	0.00
N	143	273	277			

Note: P-values are reported from Wald tests on the equality of means for each variable. Standard errors are clustered at the cluster level.

Table 21: T-test of mean difference in preferences between *barrios* with one treatment mode and *barrios* with two or more treatment modes

	Food	Cash	Voucher
All	0.0193	–0.0235	0.0581
	(0.73)	(0.58)	(0.24)
None	–0.0398	–0.0282	–0.0918**
	(0.42)	(0.34)	(0.05)
N	336	421	434

Note: P-values in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Additionally, we carry out a leave-one-out analysis, i.e. we investigate the effect of a stepwise exclusion of *barrios* on the sensitivity of the estimates. We limit the *barrios* we leave out to those in the treatment group, as the question of contamination is relevant only for treatment groups. There are 61 *barrios* in the treatment group; accordingly, we ran 61 regressions, each time leaving out one of the *barrios*. In Figure 1, we plot the resulting coefficients and p-values for the regression results of OP-Table 2 and OP-Table 3. For all outcomes, the new coefficients of the sensitivity analysis regressions are within the confidence interval of the original estimate. This indicates that no single *barrio* sufficiently affects the results to lead to significantly different estimates.

We also test whether leaving out a cluster would lead to a change in statistical significance, i.e. if some p-values would now be equal to or above 0.05 when the original p-value was below 0.05, and vice versa. In Figure 2, we plot the p-values, using a log-scale for better visibility. As can be seen, for three outcomes the p-values cross the 0.05 threshold. For food consumption, for four *barrios*, its exclusion leads to p-values above 0.05 for the cash treatment, whereas the original analysis indicates a statistically significant effect. For non-food consumption, one *barrio*'s exclusion leads to a p-value above 0.05 for voucher treatment, whereas the original analysis indicates a statistically significant effect. Additionally, for the cash treatment, the exclusion of one other *barrio* leads to a p-value below 0.05, whereas the original analysis indicates a statistically non-significant effect. For the log of non-food consumption, the exclusion of one *barrio* leads to a p-value above 0.05 for voucher treatment, whereas the original analysis indicates a statistically significant effect.

Finally, two *barrios*' exclusions lead to a p-value above 0.05 for food treatment, whereas the original analysis indicated a statistically significant effect. We look for commonalities among the *barrios* that cause the estimates to cross the p = 0.05 threshold (Table 22). We specifically look at the province, number of clusters in these *barrios* and the number of different modes applied in these *barrios*. However, we could not identify a distinct pattern due to the low number of *barrios* that caused a crossing of the threshold. More of these *barrios* are in the Sucumbíos province, but this is not notable given that a larger part of the overall sample came from this province. The *barrios* also had more than one distinct treatment applied in the *barrio* and mostly had more than one cluster, which could suggest that leaving out *barrios* with a relatively larger sample size could cause the p-value to cross the threshold. Overall, however, it appears that the results are quite robust to variations in the included clusters.

Table 22: Characteristics of *barrios* whose exclusion caused a crossing of the threshold of p = 0.05

Barrio (ID)	Province	Number of distinct treatments	Clusters per *barrio*	Treatment
		Food consumption		
9	Carchi	3	4	Cash
22	Sucumbíos	2	2	Cash
39	Sucumbíos	2	2	Cash
43	Sucumbíos	2	2	Cash
		Log non-food consumption		
1	Carchi	1	3	Food
18	Carchi	2	3	Food
37	Sucumbíos	2	5	Voucher
		Non-food consumption		
36	Sucumbíos	1	1	Cash
37	Sucumbíos	2	5	Voucher

Figure 1: Sensitivity of coefficients of OP-Table 2 and OP-Table 3 to exclusion of *barrios*

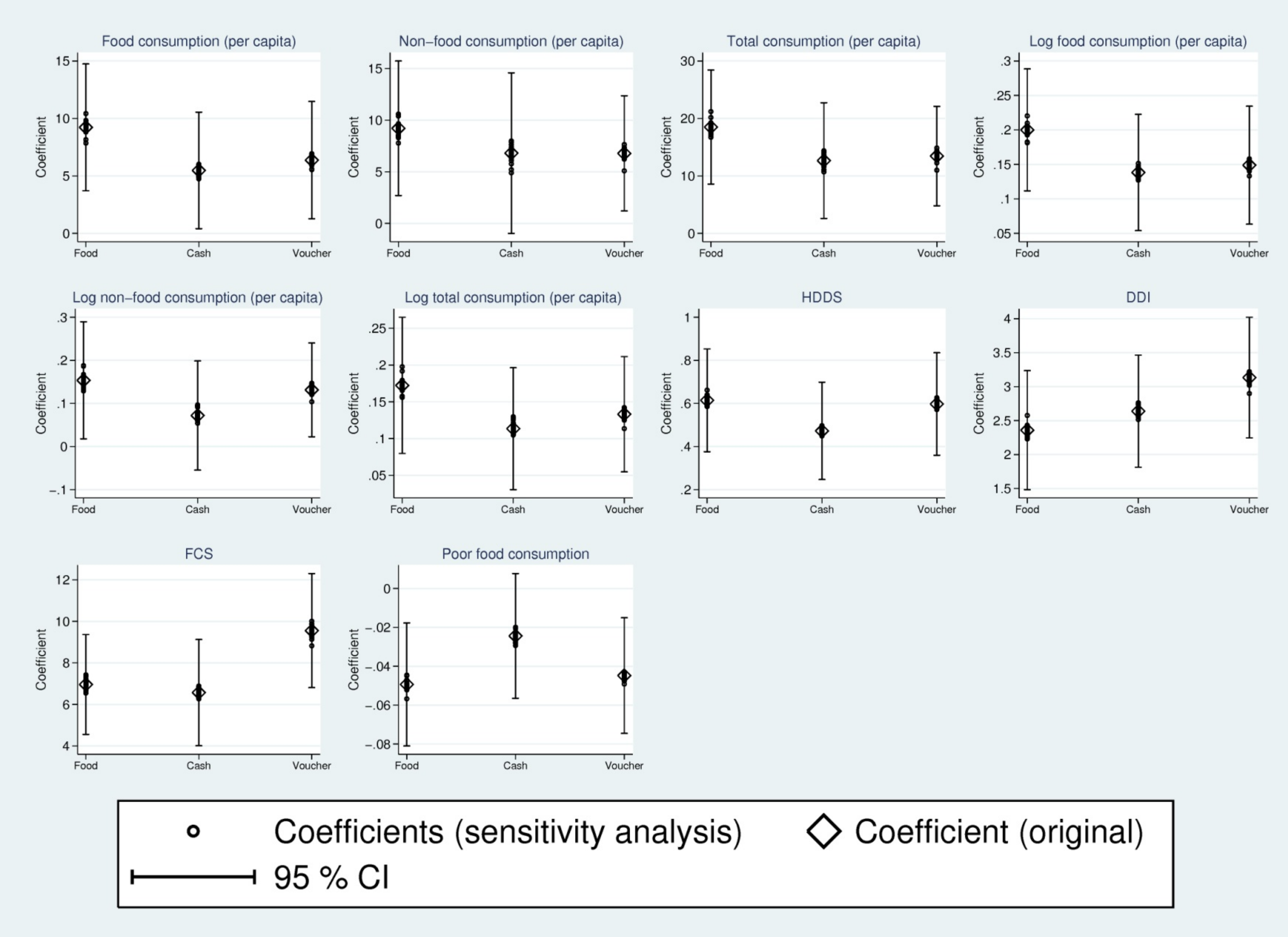
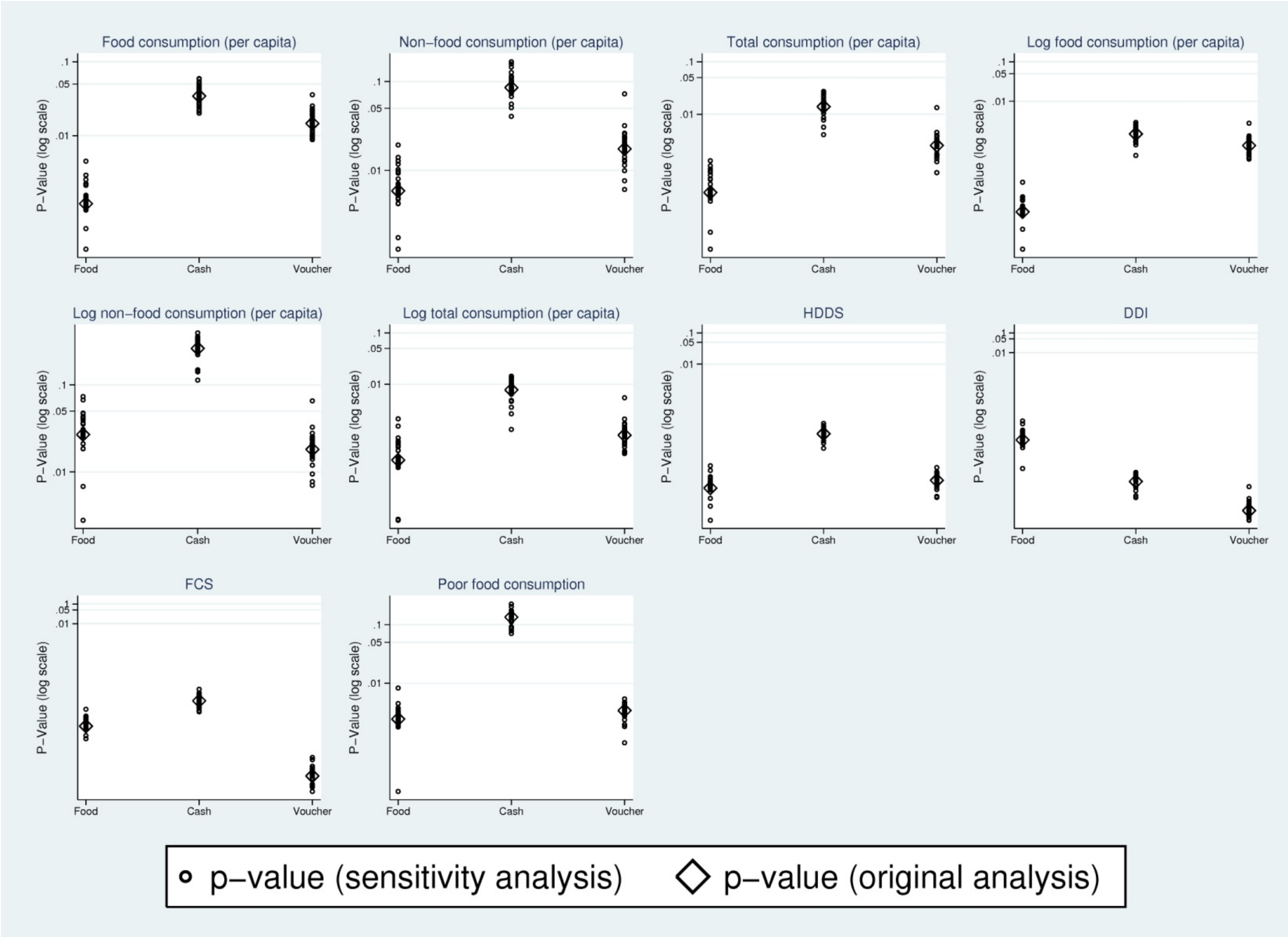


Figure 2: Sensitivity of p-values of OP-Table 2 and OP-Table 3 to exclusion of *barrios*



5.1.2 Province interactions

Province was a significant predictor of food consumption and diversity outcomes; hence, we were interested whether the intervention effects might differ by province. We therefore create an interaction variable (treatment mode x Sucumbíos province). The results indicate no significant differences in effects between provinces, with two exceptions: (i) HDDS, where food treatment has a significantly smaller impact in Sucumbíos, and (ii) PFC, which decreases significantly less in Sucumbíos than Carchi (Table 23 and Table 24).

As an additional check, we estimate models without interaction effects but stratified by province. This allows for a more intuitive interpretation of the effects in each province, in particular for those outcomes where the interaction effect was statistically significant. This analysis confirms the results of the interaction analysis above (Table 25, Table 26, Table 27 and Table 28). The likely reason for the smaller positive effects in Sucumbíos is that overall levels of HDDS and PFC were already at better levels than in Carchi, likely making it more difficult to improve them even further, while the poorer levels in Carchi made it easier to achieve larger improvements in these outcomes. This finding may indicate that effectiveness of an intervention also depends on the context of an area, such as the baseline situation.

Table 23: OP-Table 2 with interaction term defines as being a household in Sucumbíos province

	LEVELS			LOGS		
	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)
Food treatment	7.59	6.19	13.97	0.19	0.15	0.16
	(4.67)	(4.75)	(7.72)*	(0.07)***	(0.12)	(0.08)**
Cash treatment	7.59	6.19	13.97	0.19	0.15	0.16
	(4.67)	(4.75)	(7.72)*	(0.07)***	(0.12)	(0.08)**
Voucher treatment	3.55	2.19	5.80	0.09	0.06	0.06
	(5.17)	(4.53)	(7.97)	(0.07)	(0.11)	(0.07)
Food treatment	2.73	5.02	7.50	0.02	0.00	0.02
x Sucumbíos	(5.87)	(6.57)	(10.24)	(0.09)	(0.15)	(0.10)
Cash treatment	–1.00	–0.14	–1.43	0.01	–0.11	–0.03
x Sucumbíos	(5.58)	(7.87)	(9.97)	(0.09)	(0.13)	(0.08)
Voucher treatment	4.32	6.99	11.66	0.09	0.11	0.11
x Sucumbíos	(5.87)	(5.76)	(9.49)	(0.09)	(0.13)	(0.09)
Sucumbíos	1.35	9.49	9.20	0.05	0.23	0.12
	(4.24)	(4.08)**	(6.65)	(0.07)	(0.09)**	(0.06)*
R^2	0.21	0.17	0.22	0.26	0.25	0.25
N	2,087	2,087	2,087	2,087	2,087	2,087
Baseline mean	47.54	64.29	111.83	3.67	3.84	4.54

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province.

Table 24: OP-Table 3 with interaction term to be a household in Sucumbíos province

	Log caloric intake (per capita)	HDDS	DDI	FCS	PFC
Food treatment	0.27	0.98	2.68	7.57	–0.09
	(0.07)***	(0.16)***	(0.58)***	(1.88)***	(0.03)***
Cash treatment	0.19	0.67	2.50	7.58	–0.06
	(0.07)***	(0.17)***	(0.69)***	(2.32)***	(0.03)*
Voucher treatment	0.23	0.84	2.40	8.07	–0.09
	(0.08)***	(0.17)***	(0.67)***	(2.27)***	(0.03)***
Food treatment x Sucumbíos	–0.09	–0.62	–0.69	–1.16	0.07
	(0.09)	(0.22)***	(0.86)	(2.47)	(0.03)*
Cash treatment x Sucumbíos	–0.11	–0.29	0.21	–1.70	0.05
	(0.08)	(0.22)	(0.86)	(2.76)	(0.04)
Voucher treatment x Sucumbíos	–0.08	–0.36	1.12	2.29	0.08
	(0.09)	(0.23)	(0.88)	(2.84)	(0.03)**
Sucumbíos	0.13	0.50	0.03	2.08	–0.08
	(0.07)*	(0.19)**	(0.67)	(1.79)	(0.03)***
R^2	0.17	0.17	0.27	0.16	0.08
N	2,087	2,087	2,087	2,087	2,087
Baseline mean	1895.43	9.18	17.27	59.86	0.11

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province

Table 25: OP-Table 2 using only observations from Carchi province

	LEVELS			LOGS		
	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)
Food treatment	7.59	6.03	13.90	0.19	0.15	0.16
	(4.71)	(4.97)	(7.88)*	(0.07)**	(0.13)	(0.08)**
Cash treatment	5.93	6.26	13.04	0.13	0.13	0.13
	(4.76)	(5.88)	(7.12)*	(0.07)*	(0.10)	(0.06)**
Voucher treatment	3.54	2.30	6.03	0.09	0.06	0.06
	(5.24)	(4.62)	(8.13)	(0.08)	(0.11)	(0.07)
R^2	0.16	0.05	0.10	0.22	0.17	0.17
N	810	810	810	810	810	810
Baseline mean	46.76	53.14	99.91	3.68	3.68	4.47
P-value: food = voucher	0.36	0.44	0.32	0.13	0.46	0.20
P-value: cash = voucher	0.59	0.49	0.33	0.52	0.42	0.27
P-value: food = cash	0.67	0.97	0.90	0.36	0.85	0.59

Note: Standard errors in parenthesis clustered at the cluster level. * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable.

Table 26: OP-Table 2 using only observations from Sucumbíos province

	LEVELS			LOGS		
	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)
Food treatment	10.33	11.12	21.35	0.20	0.15	0.17
	(3.60)***	(4.51)**	(6.72)***	(0.06)***	(0.08)*	(0.06)***
Cash treatment	4.95	6.51	11.63	0.14	0.02	0.10
	(3.02)	(5.12)	(6.90)*	(0.05)***	(0.08)	(0.06)*
Voucher treatment	7.87	9.21	17.37	0.18	0.17	0.17
	(2.82)***	(3.58)**	(5.16)***	(0.05)***	(0.06)***	(0.05)***
R^2	0.24	0.21	0.26	0.28	0.25	0.27
N	1,277	1,277	1,277	1,277	1,277	1,277
Baseline mean	48.03	71.36	119.39	3.65	3.95	4.59
P-value: food = voucher	0.51	0.67	0.57	0.71	0.83	0.93
P-value: cash = voucher	0.37	0.60	0.43	0.42	0.09	0.21
P-value: food = cash	0.17	0.43	0.25	0.28	0.19	0.26

Note: Standard errors in parenthesis clustered at the cluster level. * $p < 0.1$ ** $p < 0.05$; *** $p < 0.01$. All estimations control for baseline outcome variable.

Table 27: OP-Table 3 using only observations from Carchi province

	Log caloric intake (per capita)	HDDS	DDI	FCS	PFC
Food treatment	0.27	0.97	2.68	7.42	–0.09
	(0.07)***	(0.16)***	(0.58)***	(1.86)***	(0.03)***
Cash treatment	0.19	0.67	2.50	7.54	–0.06
	(0.07)**	(0.17)***	(0.69)***	(2.23)***	(0.03)*
Voucher treatment	0.23	0.83	2.41	8.00	–0.09
	(0.08)***	(0.16)***	(0.67)***	(2.20)***	(0.03)***
R^2	0.18	0.17	0.26	0.18	0.09
N	810	810	810	810	810
Baseline mean	1986.84	8.82	16.94	57.53	0.14
P-value: food = voucher	0.57	0.26	0.57	0.78	0.87
P-value: cash = voucher	0.52	0.21	0.89	0.85	0.11
P-value: food = cash	0.17	0.02	0.72	0.96	0.24

Note: Standard errors in parenthesis clustered at the cluster level. * $p < 0.1$ ** $p < 0.05$; *** $p < 0.01$. All estimations control for baseline outcome variable.

Table 28: OP-Table 3 using only observations from Sucumbíos province

	Log caloric intake (per capita)	HDDS	DDI	FCS	PFC
Food treatment	0.18	0.36	1.99	6.49	–0.02
	(0.05)***	(0.15)**	(0.64)***	(1.63)***	(0.01)*
Cash treatment	0.09	0.38	2.70	5.93	–0.01
	(0.04)**	(0.14)**	(0.52)***	(1.50)***	(0.02)
Voucher treatment	0.15	0.48	3.52	10.39	–0.02
	(0.04)***	(0.16)***	(0.58)***	(1.72)***	(0.02)
R^2	0.17	0.15	0.28	0.15	0.06
N	1,277	1,277	1,277	1,277	1,277
Baseline Mean	1837.44	9.42	17.49	61.34	0.09
P-value: Food=Voucher	0.59	0.37	0.02	0.04	0.66
P-value: Cash=Voucher	0.17	0.41	0.11	0.02	0.50
P-value: Food=Cash	0.09	0.88	0.22	0.75	0.24

5.2 Generalized linear mixed models as competing modeling approaches

GLMM account for the hierarchical nature of the data (Gelman and Hill 2007) – e.g. clusters nested within *barrios*. This model class, also known as a hierarchical or multilevel model, can address typical violations of assumptions of standard linear models, e.g. normality, homogeneity of variance or independence of data. The latter is particularly useful in the present application, as the data are hierarchical and the authors' statistical model choice relies on a successful randomization. If randomization is successful, then the original, statistically less elaborate model is appropriate and sufficient. If the GLMM estimates for the intervention effects differ significantly from the originally estimated effects, we can conclude that the results are sensitive to the modeling choice. However, a significant difference may not necessarily translate into a substantial or practically meaningful difference. This has to be judged on a qualitative basis – e.g. does the interpretation of the results change? Moreover, the GLMM analysis will yield the intra-class correlation among the clusters (Eldridge and Kerry 2012). This statistic should be reported, according to the CONSORT guidelines, to allow proper inclusion of the estimates into a meta-analysis (Campbell et al. 2012).

In Table 29 and Table 30, we present the re-estimated results for OP-Table 2 and OP-Table 3, respectively, using GLMM, but an otherwise identical model in terms of the included variables and so forth. We focus on OP-Tables 2 and 3 as those contain the main effect estimates that are of substantial interest for policymakers. In terms of statistical significance, the effect of cash treatment on non-food consumption is no longer statistically significant at the 10 percent level for only one outcome variable (non-food consumption per capita in levels), as the coefficient is reduced from 6.81 in the original paper to 5.83 using GLMM, while the standard error remains similar. Furthermore, the difference in the effects between food and voucher treatment is no longer statistically significant at the 10 percent level for the FCS.

The other results are very close to those of the original paper. Furthermore, while the magnitude of coefficients differs slightly, the qualitative interpretation of the results – hence, the central argument of the original paper – does not change. This suggest that the original, statistically less elaborate model specification, which requires successful randomization, sufficiently accounts for the hierarchical nature of the data.

Table 29: Replication of OP-Table 2 using GLMM model accounting for hierarchical structure of data, i.e. clusters and *barrios*

	LEVELS			LOGS		
	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)
Food treatment	9.33	8.88	19.13	0.20	0.15	0.18
	(3.04)***	(3.54)**	(5.61)***	(0.05)***	(0.07)**	(0.05)***
Cash treatment	5.47	5.83	12.23	0.14	0.06	0.11
	(2.68)**	(3.85)	(5.27)**	(0.05)***	(0.06)	(0.05)**
Voucher treatment	5.96	6.81	12.67	0.14	0.13	0.13
	(2.64)**	(3.01)**	(4.75)***	(0.04)***	(0.06)**	(0.04)***
N	2,087	2,087	2,087	2,087	2,087	2,087
Baseline mean	47.54	64.29	111.83	3.67	3.84	4.54
P-value: food = voucher	0.26	0.53	0.25	0.23	0.65	0.33
P-value: cash = voucher	0.82	0.74	0.91	0.88	0.17	0.58
P-value: food = cash	0.16	0.42	0.22	0.15	0.12	0.17
ICC: *barrio*	0.031	0.057	0.051	0.046	0.086	0.069
ICC: cluster | *barrio*	0.018	0.057	0.049	0.023	0.074	0.059

Notes: Standard errors in parenthesis clustered at the cluster level. ICC stands for intra-cluster correlation. p-values in brackets, * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province.

Table 30: Replication of OP-Table 3 using GLMM model accounting for hierarchical structure of data, i.e. clusters and *barrios*

	Log caloric intake (per capita)	HDDS	DDI	FCS	PFC
Food treatment	0.20	0.65	2.33	7.20	–0.05
	(0.04)***	(0.14)***	(0.51)***	(1.20)***	(0.02)***
Cash treatment	0.12	0.52	2.78	6.54	–0.02
	(0.03)***	(0.14)***	(0.47)***	(1.22)***	(0.02)
Voucher treatment	0.17	0.62	3.13	9.30	–0.04
	(0.04)***	(0.14)***	(0.48)***	(1.38)***	(0.02)***
N	2,087	2,087	2,087	2,087	2,087
Baseline mean	7.36	9.18	17.27	59.86	0.11
P-value: food = voucher	0.49	0.77	0.08	0.12	0.70
P-value: cash = voucher	0.10	0.31	0.34	0.04	0.14
P-value: food = cash	0.02	0.15	0.28	0.53	0.07
ICC: *barrio*	0.031	0.033	0.059	0.032	0.015
ICC: cluster | *barrio*	0.003	0.026	0.036	0.023	0.000

Notes: Standard errors in parenthesis clustered at the cluster level. ICC Intra-cluster correlation. p-values in brackets, * p < 0.1 ** p < 0.05; *** p < 0.01. All estimations control for baseline outcome variable and province.

6. Theory of change analysis

The cost-effectiveness analysis of the original paper is somewhat limited, as the interpretation does not take into account the uncertainty inherent to the underlying (cost-effectiveness) data. Hence, we undertake a particular type of health economic simulation analysis – a probabilistic sensitivity analysis – for all three modes of food assistance to investigate the influence of uncertainty in the data and quantify the probability of choosing the (most) cost-effective strategy (cash, food or vouchers).

A crucial element in judging the cost-effectiveness of an intervention is clearly a decision maker's willingness to pay (WTP) for a given effect size, say, a 1 percent increase. An intervention is cost-effective when the costs are below the applicable WTP and not cost-effective when above this WTP. In many cases, however, a decision maker's explicit WTP is not available. Moreover, because WTP can vary – e.g. across context or time – it is advisable to conduct the cost-effectiveness analysis in such a way that the results are available for a range of WTP. For the WTP explored in our analysis, we finally set an upper limit USD10,000 per unit of effect estimates, as larger values did not change the interpretation of our findings.

A probabilistic sensitivity analysis is particularly suited to accounting for uncertainty in the parameters in health economic evaluations. Although this type of analysis can be used for various sources of uncertainty (e.g. structural uncertainty) we solely focus on the uncertainty in the parameters, i.e. the sampling error in the regression coefficients (Briggs et al. 2006; Drummond 2015). A probabilistic sensitivity analysis is run for several hundred or thousand times while, for each run, the input parameters are drawn from a stochastic distribution.

For our replication study, the effect parameters can be taken from the statistical estimation model, i.e. the effect estimates of the statistical models documented in the original paper. However, a limitation of this probabilistic sensitivity analysis lies in the lack of an estimator for the uncertainty of the cost data, as these data were not derived from sampling data but through activity-based accounting. Hence, we assume a standard error of 30 percent of the mean of the input parameter for the cost estimate. This is an often-used conservative assumption when standard errors are not available (Westwood et al. 2013; Briggs et al. 2012).[1] Another challenge is that effect and cost data (the input parameters of a probabilistic sensitivity analysis) are usually correlated with each other. For the effect estimates, we can use the variance-covariance matrix of the respective regression model to calculate the correlation structure. For the cost data, we are unable to obtain such a dependency structure, as we do not have sampling data.

A full probabilistic sensitivity analysis will yield several thousand cost-effectiveness estimates, i.e. one estimate for each draw from the stochastic distribution of the input parameters. These are then depicted in a cost-effectiveness plane. After a full

[1] Using 30 percent of the parameter values as a conservative estimate of the standard error is a rule-of-thumb approach that ensures a parameter is still considered statistically significant, i.e. the confidence interval based on this assumed standard error does not overlap with zero. A slightly larger assumption about the standard error – e.g. 40 percent of the parameter in question – would translate into considering the parameter statistically not significant at the 1 percent level.

probabilistic sensitivity analysis, we then calculate the percentage of cases for which an intervention is more cost-effective for a given WTP and estimate this number for varying WTP thresholds. Calculating this metric provides information about which intervention has the highest probability of being the most cost-effective when implemented.

6.1 Input data

In this section, we present the input data for our additional theory of change analysis. In Table 31, we present the effect estimates (including standard errors) for which we conduct the analysis. These values are taken from the original paper (OP-Table 3). We focus on these effect estimates, as those are the estimates used by the original authors for their cost-effectiveness analysis (OP-Table 6).

In Table 32, we present the cost estimates used for the probabilistic sensitivity analysis, assuming a standard error of 30 percent of the mean. The cost data are the modality-specific implementation costs, as used by the original authors for the cost-effectiveness analysis (OP-Table 6). However, whereas the original authors used the cost per transfer for their analysis, we use the amount per beneficiary, as presented in OP-Table B.8. The latter allows a more intuitive interpretation of the results of our additional analysis, but does not change the implications.[1]

In Table 33, we present the variance-covariance structure of the effect estimates. We calculate the values using the regression models that are used to calculate the effect estimates. For the cost data, such data are not available, as those are not based on sampling data.

[1] The cost per transfer are exactly one sixth of the cost per beneficiary (six transfers per beneficiary in each mode) and, hence, perfect linear combinations.

Table 31: Effect estimates incl. standard error as input to probabilistic sensitivity analysis

		Log caloric intake (per capita)	HDDS	DDI	FCS	PCS
Food	point estimate	0.21	0.61	2.36	6.96	–0.05
	standard error	0.04	0.12	0.44	1.22	0.02
Cash	point estimate	0.12	0.47	2.64	6.57	–0.02
	standard error	0.04	0.11	0.42	1.29	–0.02
Voucher	point estimate	0.18	0.6	3.13	9.56	–0.04
	standard error	0.04	0.12	0.45	1.39	0.02

Note: Source is OP-Table 3.

Table 32: Cost estimates incl. standard error as input to probabilistic sensitivity analysis

		Cost per Transfer
Food	Point estimate	68.75
	Standard error (30% of point estimate)	20.63
Cash	Point estimate	17.97
	Standard error (30% of point estimate)	5.39
Voucher	Point estimate	19.61
	Standard error (30% of point estimate)	5.88

Note: Source is OP-Table B.8 (point estimates).

Table 33: Variance-covariance matrix of effect estimates as input for probabilistic sensitivity analysis

(a) Log caloric intake (per capita)

	Food	Cash	Voucher
Food	0.001762	0.000880	0.000844
Cash	0.000880	0.001413	0.000840
Voucher	0.000844	0.000840	0.001525

(b) HDDS

	Food	Cash	Voucher
Food	0.014538	0.009724	0.009889
Cash	0.009724	0.013039	0.009874
Voucher	0.009889	0.009874	0.014477

(c) DDI

	Food	Cash	Voucher
Food	0.196567	0.109073	0.110377
Cash	0.109073	0.174209	0.109006
Voucher	0.110377	0.109006	0.201812

(d) FCS

	Food	Cash	Voucher
Food	1.480792	0.671373	0.709252
Cash	0.671373	1.675531	0.679698
Voucher	0.709252	0.679698	1.931010

(e) PFC

	Food	Cash	Voucher
Food	0.000256	0.000152	0.000156
Cash	0.000152	0.000264	0.000154
Voucher	0.000156	0.000154	0.000226

6.2 Incremental cost effectiveness ratios

For the pairwise comparison of health economic interventions, an often-used measure is the incremental cost-effectiveness ratio (ICER), defined as (C_1 - C_0) / (E_1 - E_0), where the marginal costs of two interventions (C) are divided by the marginal effect (E). The ICER shows the relative cost-effectiveness of an intervention to the next best (i.e. most cost-effective) intervention and allows for the identification of dominated interventions (Table 29). An intervention is dominated when another intervention has larger effects and is less costly. An intervention is also dominated when the intervention is not on the efficiency frontier of all cost-effective interventions. That is, assuming one can arbitrarily split the target population between two interventions, a third intervention that is above the line between these two interventions is (extendedly) dominated by these two interventions.

For example, this can be observed in the top-left panel of Figure 3 (cost-effectiveness plane for log caloric intake). The ICER for vouchers (V) is smaller than the ICER for cash (C). The ICER calculated from the origin ("no intervention") with zero costs and zero effects, is USD149.5 for cash and USD108.94 for vouchers. Although a single voucher is costlier per beneficiary, one can (on average) reach the same total population-level increase in log caloric intake by spending only two thirds of the amount necessary when using cash treatment.

A small numerical example with a target population of 100 beneficiaries: Giving each beneficiary a cash intervention would cost USD1,797 (100 * USD17.97) and increase the log caloric intake by 12 units (100 * 0.12). Giving 67 of the 100 beneficiaries a voucher (and nothing to the remaining target population) would only cost USD1,313.87 (67 * USD19.61) but yield the same approximate overall increase in units of log caloric intake (67 * 0.18 = 12.06). This example also shows clearly where the limitations of using an ICER as a decision criterion lies. First, one might not be able to split the target population into beneficiaries and non-beneficiaries for reasons of social justice or for practical purposes. Second, there might be a decreasing marginal utility of the effect in question. An increase in 0.12 units of log caloric intake for 100 beneficiaries might provide more overall utility than a corresponding increase of 0.18 units in 67 beneficiaries.

For all effect estimates, cash is dominated by vouchers, i.e. when one is able to split the target population into beneficiaries and non-beneficiaries arbitrarily, vouchers will always yield a higher overall impact for the same costs. Moreover, for two outcomes (DDI and FCS), vouchers also dominate food as an intervention, making vouchers the most cost-effective intervention, independent of the WTP of the decision maker. Despite being costlier per beneficiary than cash, it might be advisable to select vouchers to increase the cost-effectiveness of a program – e.g. in the case of log caloric intake, one has to spend less than 10 percent more to receive a 50 percent larger effect.

Table 34: Calculation of incremental cost-effectiveness ratios (all included and excluding dominated)

(a) Log caloric intake (per capita)

	Food	Cash	Voucher
Effect	0.21	0.12	0.18
Costs	68.75	17.97	19.61
Rank	1	3	2
ICER (all included)	1638.00	149.75	27.33
ICER (excl., dominated)	1638.00	Dominated	108.94

(b) HDDS

	Food	Cash	Voucher
Effect	0.61	0.47	0.6
Costs	68.75	17.97	19.61
Rank	1	3	2
ICER (all included)	4914.00	38.23	12.62
ICER (excl., dominated)	4914.00	Dominated	32.68

(c) DDI

	Food	Cash	Voucher
Effect	2.36	2.64	3.13
Costs	68.75	17.97	19.61
Rank	3	2	1
ICER (all included)	29.13	–181.36	3.35
ICER (excl., dominated)	Dominated	6.81	3.35
	Dominated	Extendedly dominated	6.27

(d) FCS

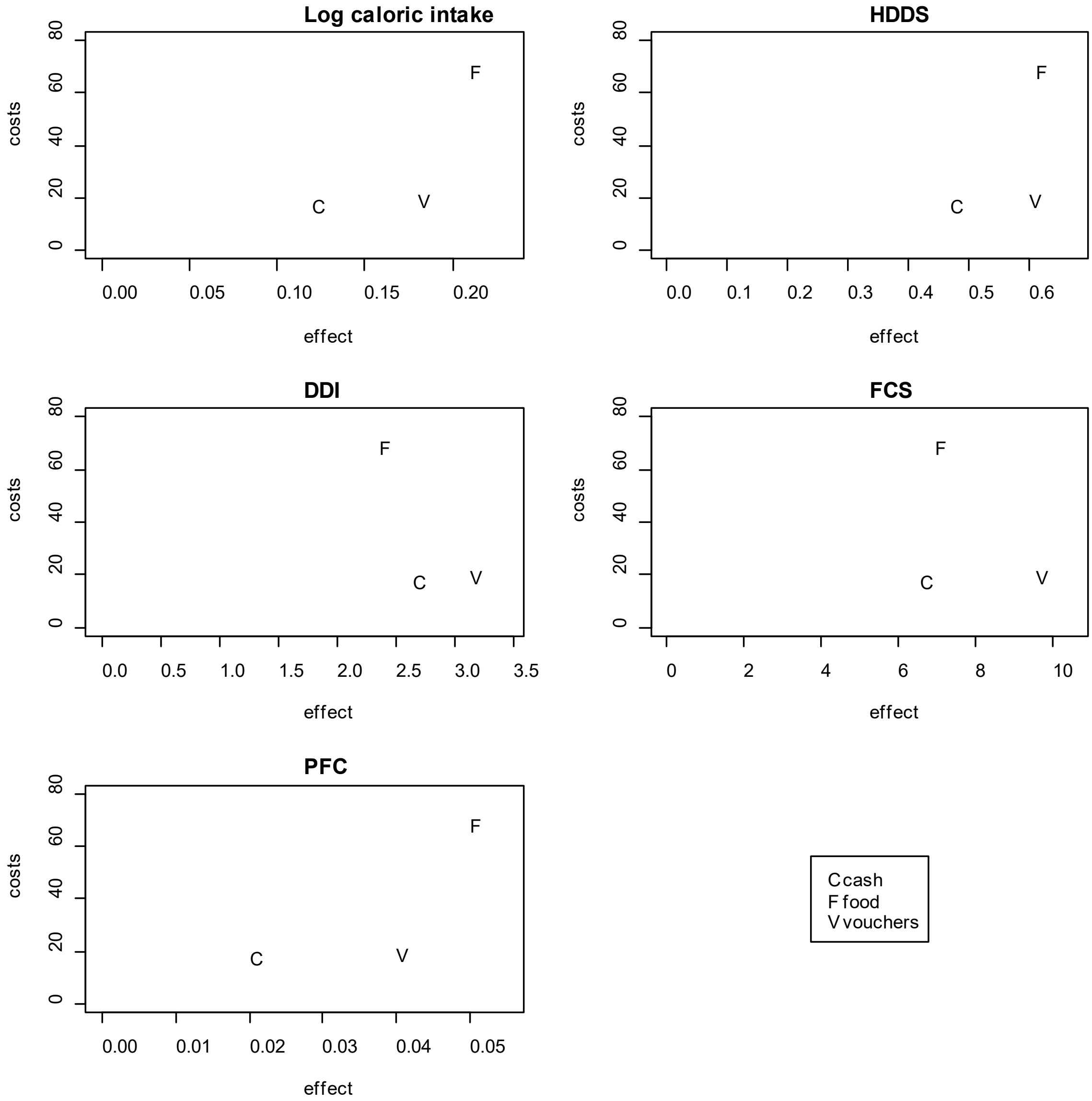
	Food	Cash	Voucher
Effect	6.96	6.57	9.56
Costs	68.75	17.97	19.61
Rank	2	3	1
ICER (all included)	130.21	2.74	–18.90
ICER (excl., dominated)	9.88	Dominated	–18.90
	Extendedly dominated	Dominated	2.05

(e) PFC (set to absolute values)

	Food	Cash	Voucher
Effect	0.05	0.02	0.04
Costs	68.75	17.97	19.61
Rank	1	3	2
ICER (all included)	4914.00	898.50	82.00
ICER (excl., dominated)	4914.00	Dominated	490.25

Note: An ICER is calculated by first ordering all interventions by effect size (see “rank”). Then, the ICERs are calculated by comparing an intervention with previous lowest intervention in terms of effects. In our case, an intervention with the rank 3 (lowest effect estimate) is compared with “no intervention” (zero costs, zero effects). The intervention ranked 2 is compared with the intervention rank 3, and so on. This yields the first sets of ICER. In our example in the top left panel (log caloric intake), the ICER from “cash” to “no intervention” is clearly higher than the ICER from “voucher” (rank 2) to “cash” (rank 3). Hence, “cash” is dominated and therefore excluded, yielding a new ICER for “voucher” (now compared with “no intervention”).

Figure 3: Cost-effectiveness planes for effect estimates



6.3 Probabilistic sensitivity analysis

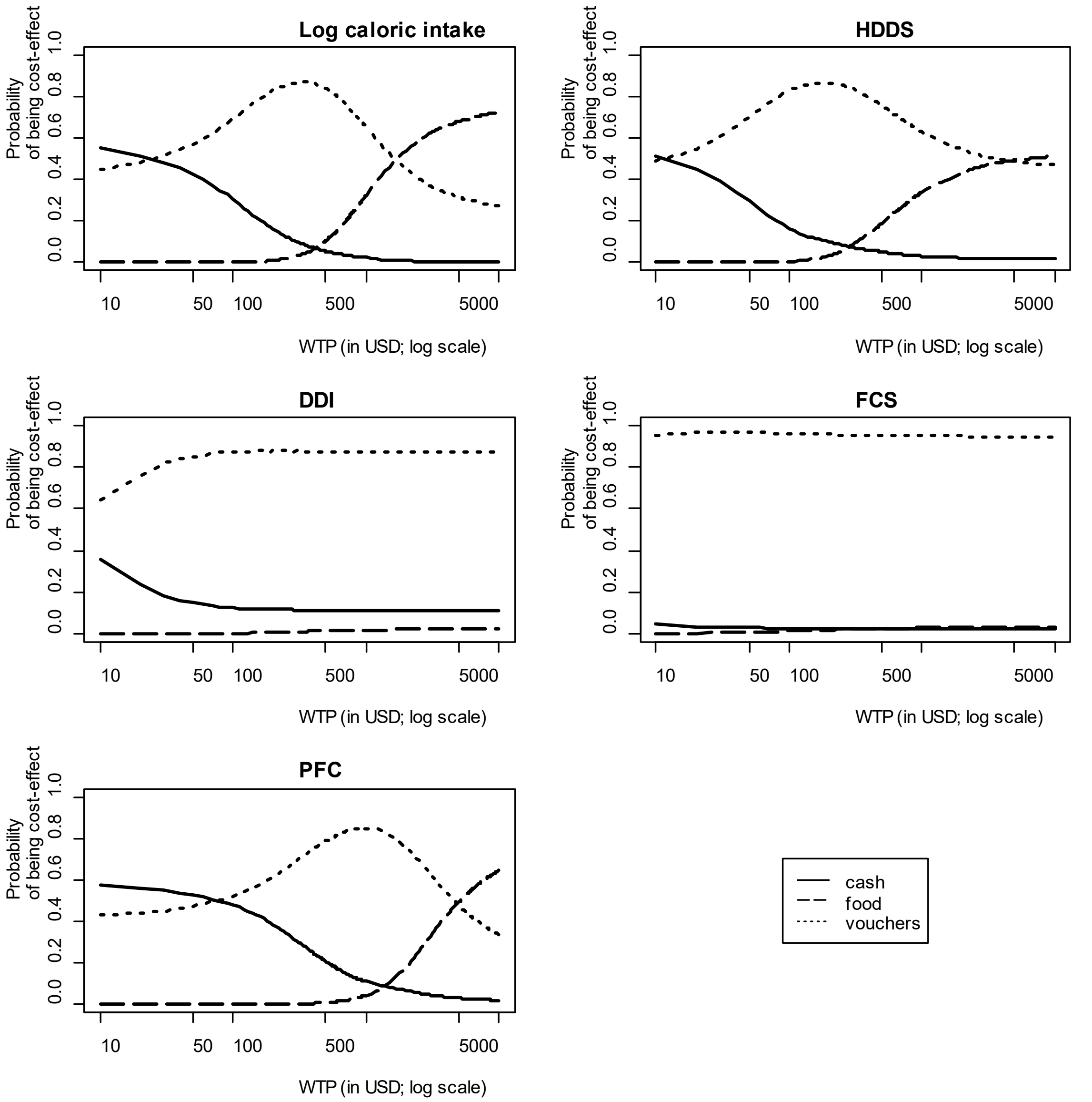
In the previous section, we investigated the implications of the point estimates for costs and effects. However, those data have been measured with uncertainty (i.e. sampling error). In a probabilistic sensitivity analysis, we can account for this uncertainty. In our probabilistic sensitivity analysis, we run 10,000 simulations in which we draw, in each run, new effect and cost estimates from a probability distribution. For our effect estimates, we draw from a multivariate normal (Gaussian) distribution that is defined by the mean effect estimates (i.e. coefficients from the regression, as reported in Table 32) and the variance-covariance structure (Table 33) in order to account for the dependency structure across treatments. For the cost estimates, we use the Gamma distribution, using the mean effect estimates and our assumptions concerning the standard errors. We do not have information on the dependency structure of the cost data; hence, we assume zero correlation across the cost estimates. A negative correlation can reduce uncertainty. A positive correlation, on the other hand, increases overall uncertainty and is mathematically comparable to increasing the magnitude of standard errors. Cost data are often positively correlated and considering the items and goods used to derive the costs in the original paper, we do not expect a strong negative correlation of cost across treatment arms.

These simulations lead in total to 10,000 distinct cost/effectiveness estimates for each treatment mode.[1] Unfortunately, in the case of more than two interventions, displaying the ICER visually is not very informative (Stollenwerk et al. 2015). Hence, the suggested graphical display to interpret the results are cost-effectiveness acceptability curves (Barton et al. 2008). For a given value of WTP for a one-unit increase, we identify the most cost-effective treatment mode in each simulation run of the probabilistic sensitivity analysis. Then, by calculating the number of times a treatment is the most-cost effective across the 10,000 simulations, we obtain the probability of a treatment being cost-effective for a given WTP. This calculation is then repeated across WTP values, from USD0 to USD10,000.

The results of the probabilistic sensitivity analysis are shown in Figure 4. For three outcomes (log caloric intake, HDDS and PFC), cash has the highest probability of being cost-effective for lower values of the WTP range (log caloric intake: approximately USD 20; HDDS: approximately USD10; PFC: approximately USD60). For increasing values of WTP, vouchers have the highest probability of being cost-effective. But with further increasing values of WTP, food has, in all three cases, the highest probability of being cost-effective (starting at approximately USD1,700 for log caloric intake, approximately USD6,600 for HDDS and approximately USD5,300 for PFC). The latter can be explained by food's having the highest effect estimates of the three outcome measures; thus, the more one is willing to pay, the more one goes for the strategy with largest effect (albeit having the highest costs attached to it). In the case of DDI and FCS, the shape of the cost-effectiveness acceptability curves reflects that vouchers have the largest effects size at only slightly larger costs than the next cheapest intervention.

[1] The number of simulations used in a PSA is somewhat arbitrary and depends on the underlying data. We carefully checked whether with this number of simulations runs the estimates were sufficiently stable.

Figure 4: Cost-effectiveness acceptability curve for all effect estimates



6.4 Expected value of perfect information

Decisions should always be based on the existing cost-effect data. But, as the cost-effectiveness acceptability curves in Figure 4 show, one may take an erroneous decision, i.e. the probability of a strategy's being cost-effective can vary quite substantially across different levels of WTP. That is, although on average we make the "right" decision by choosing the dominant strategy, another decision could yield a higher overall benefit. In principle, additional research can help reduce the uncertainty concerning the decision, e.g. decreasing the magnitude of the measurement errors by conduction an additional trial.

An important task for policymakers is to maximize the overall benefit, given budget constraints. Considering the possibility that a decision measured with uncertainty could be erroneous, and hence does not maximize the potential benefit, an important question is how much more to spend to reduce this uncertainty. This can be captured using the expected value of perfect information (EVPI; Ekwunife and Lhachimi 2017; Oppen et al.

2008). For a decision situation in which the uncertainty has been quantified using cost-effectiveness acceptability curves, one can calculate the potential value of “perfect information” (no uncertainty) for a given level of WTP. This would also yield an upper bound on what to spend for additional research to make “better” decisions (with a higher probability of being cost-effective). The EVPI is calculated by choosing the strategy with the highest possible (net monetary) benefit[1] from each simulation run, i.e. having perfect information about the state of the world. We then compare the average value with the strategy that yields the highest average net monetary benefit (the optimal decision under uncertainty). The difference is the opportunity cost of taking a decision under uncertainty compared with no uncertainty.

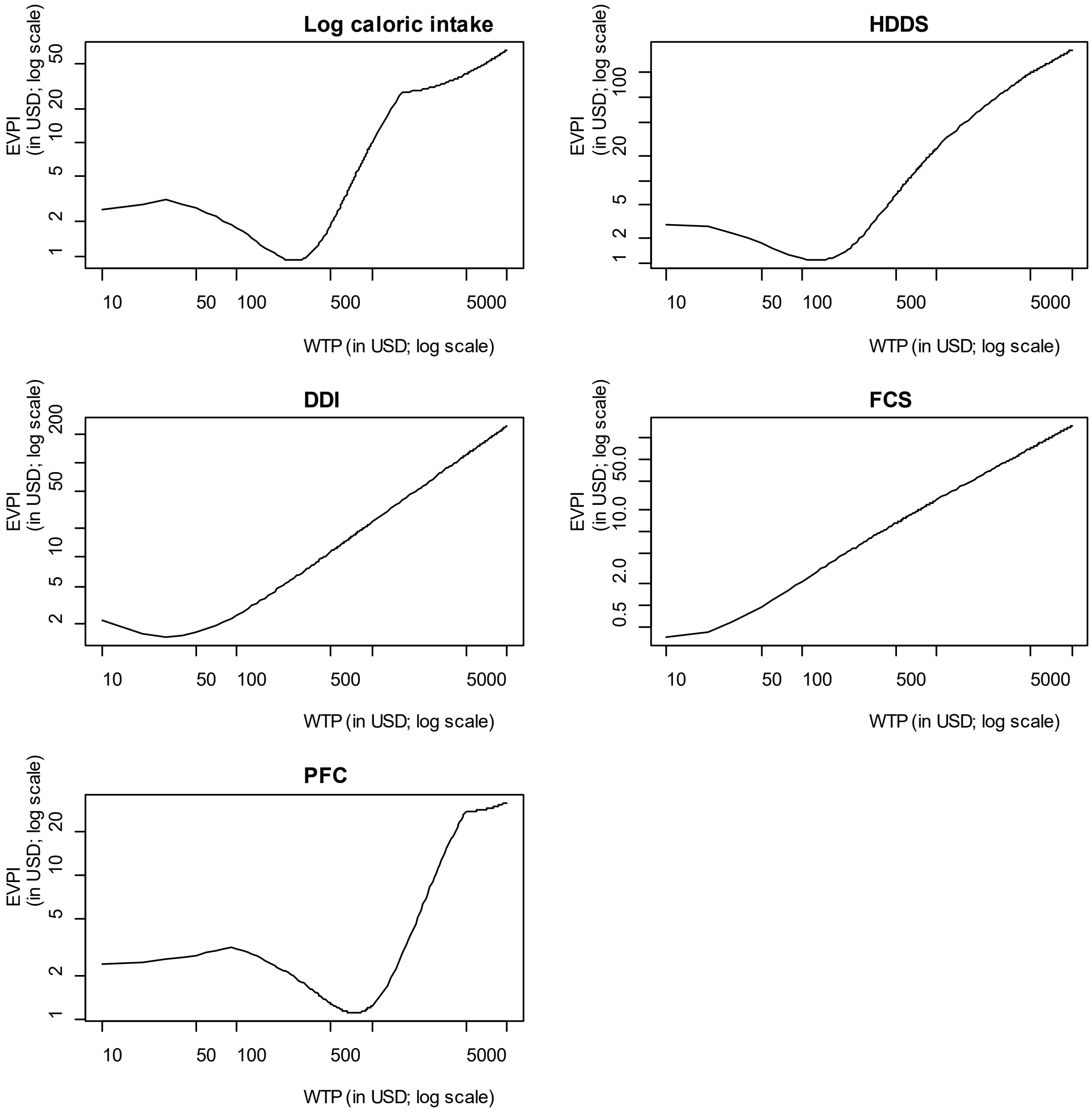
In Figure 5, the EVPI for all five outcomes with varying levels of WTP has been plotted. The shape of the EVPI curves corresponds partly to the shape of the cost-effectiveness acceptability curves. For three outcomes (log caloric intake, HDDS and PFC), the initial EVPI increases and then reaches a local minimum before steadily increasing again. The initial increase reflects that for these three outcomes, given this range of WTP, cash has the highest (but decreasing) probability of being cost-effective before being overtaken by vouchers. This reflects the relatively large uncertainty for these levels of WTP in choosing between cash or vouchers.

For the WTP ranges where vouchers clearly have a large probability of being cost-effective (and hence comparably low uncertainty), the EVPI is decreasing and reaches a local minimum. For all five outcomes, the eventually constant rise of the EVPI for large values of WTP is somewhat independent from the magnitude of the underlying uncertainty; simply put, if I have a large WTP, I am also willing to pay more for additional research to reduce uncertainty even if the remaining uncertainty is relatively small und perfect information is unlikely to change my decision.

Nonetheless, these EVPIs must be interpreted in the context of the intervention. First, perfect information will never be attained; additional research can only reduce uncertainty. Hence, one must carefully consider how much additional information additional research can truly yield. Second, it is important to understand that these figures are the value of perfect information per beneficiary, making a decision on additional research also dependent on the size of the target population. For a large target population that may profit from such an intervention, substantial additional research spending could be justified. For example, for a target population of 100,000 individuals, an EVPI of USD5 could justify additional research spending of up to USD500,000. Finally, this also shows that interpretation of these values also depends on the total underlying costs of a program. The costlier a program (e.g. an indefinite duration of assistance), the higher the calculated overall EVPI could be.

[1] The net monetary benefit is defined WTP * Effect – Cost. That is, we calculate the value of a unit of effect for a decision maker given the decision maker’s WTP and subtract the cost. If the WTP times the expected effect is larger than the costs, then an intervention is cost-effective for this level of WTP.

Figure 5: Expected value of perfect information for all effect estimates



6.5 Summary

In Section 6.2 we use only the point estimates of the cost and treatment variables (i.e. not accounting for uncertainty). For all outcome measures, cash is the least preferable option in terms of cost-effectiveness. For two outcomes (DDI and FCS), vouchers are the most cost-effective outcome. For the other three outcomes, food has a larger effect but also higher costs than vouchers. The ICERs (incremental cost for an additional unit of effect) for these outcomes are USD1,638 for log caloric intake, USD4,114 for HDDS and USD4,914 for PFC. Although we have no evidence that effects of vouchers are linear – i.e. doubling the value of a voucher leads to doubling the effect – we believe it is safe to assume that spending the amount on vouchers would be more cost-effective than spending it on direct food transfers.

The results of the probabilistic sensitivity analysis in Section 6.3, as visualized in the cost-effectiveness acceptability curves, show that for three outcomes (log caloric intake, HDDS and PFC), cash has, for (very) low values of WTP, the highest probability of being cost-effective. For increasing values of WTP, vouchers have the highest probability of being cost-effective, up to relatively high WTP levels, where food emerges as the intervention with the highest probability of being cost-effective. Hence, the probabilistic sensitivity analysis shows that for low values of transfer costs (log caloric intake: approximately USD20; HDDS: approximately USD10; PFC: approximately USD60), cash may prove to be the optimal choice, despite being dominated by vouchers when looking only at the results from Section 6.2. The expected value of perfect information analysis in Section 6.4 also suggests that, in particular for these lower levels of WTP, additional research to reduce uncertainty could be of benefit to decision makers. Hence, we believe the uncertainty we encounter in the data justifies additional research to discern more clearly whether vouchers or cash are more cost-effective when assuming reasonable levels of cost per beneficiaries, at least for some outcomes.

A technical limitation in our analysis is a lack of reliable measures of uncertainty in the cost data. An extension of this analysis could investigate potential values for cost uncertainty. Moreover, instead of using only the overall cost of a mode, a more detailed analysis of the cost should aim to differentiate between different costs – e.g. initial investment cost, fixed cost and marginal cost – as those may affect the three modes of provision differently (in particular when considering provision over a longer duration). A limitation in the interpretation of the data is that we investigate the cost effectiveness in terms of a one-unit increase per outcome measure, making the analysis not comparable across outcomes. Additionally, we do not know what (minimal) levels of change in the outcomes are desirable in order to have a meaningful intervention, given a particular WTP, and whether the marginal utility of an increase in the outcomes is constant or decreasing.

7. Discussion and limitations

Hidrobo and colleagues' paper analyzes a cluster-randomized controlled trial comparing the effectiveness of different modes of food assistance (cash, food and vouchers) with a control mode (no assistance). Their statistical analysis used an ANCOVA model and included several food-related outcome measures (food consumption, several indices of food security and diet). In this paper, we report a full replication analysis of this study, based on a previously published replication plan (Lhachimi 2017).

We began with a push-button replication. For this, we obtained the data via email from Melissa Hidrobo. Only slight modifications (e.g. to path names) were necessary and we were able to replicate six of the seven tables in the original paper using the provided data and Stata code. We were able to replicate the last table by hand using the information and data from the original paper and its appendix. Hence, results of the push-button replication are comparable with the published results.

We proceeded with the pure replication. Since the data set did not contain the raw data and creation of the data set was not displayed in the provided .do file, we tried to replicate several of the main outcome variables, given that sufficient data to do so was available in the provided data set. Doing this, we found very small discrepancies for the HDDS, but larger discrepancies for the FCS. We were able to calculate the same FCS for fewer than half of the observations; for the rest, our replicated score was consistently higher than the original FCS, which also led to a smaller prevalence of PFC. Although we are unsure about the reasons for these differences, small discrepancies in the underlying operations and data to calculate the FCS could be a potential culprit. Using these reconstructed indices, we nonetheless re-estimated the tables, finding that the large majority of results changed only to a small extent, not leading to changes in the interpretation of the overall results. The main notable change in results was that cash now also seemed to be just as effective as food in reducing PFC when using the reconstructed FCS. The original paper finds food provision to be superior to cash transfers to reduce PFC. However, because we are unsure about the underlying reasons for the found discrepancies in the FCS, we are hesitant to interpret this change in results any further.

To further check the steps taken in the .do file to create the data set for analysis, and to carry out the analysis, we rebuilt the Stata code in a different statistical language (R) and were able to replicate all main results. Furthermore, we were able to replicate the cost-effectiveness analysis presented in the original paper, for which no code was included in the .do file, using the information provided in the paper and its supplementary material. In conclusion, we deem the pure replication fully satisfactory, with the caveat that we were not able to completely assess the creation of the data set. Overall, this pure replication has shown the results and conclusions of the original paper to be robust.

The additional analysis focused on potential effects of the sampling frame. We re-estimated the key results to investigate whether, in a *barrio*, only clusters with the same treatment mode were present or whether two or more treatment modes were present. The rationale is that observing individuals exposed to different treatments could affect the observer. We focused on the preferences concerning the treatment mode. By comparing *barrios* where only one treatment mode was present and *barrios* where two or more treatment modes were present, we could identify a sizable and statistically significant difference concerning the preference for vouchers. Voucher beneficiaries living in *barrios* where two or more treatment modes were present had an almost 50 percent higher total rejection rate of vouchers – i.e. preferring to receive none of the benefits of vouchers – than beneficiaries living in *barrios* where vouchers were the only treatment mode. Further we found that beneficiaries receiving cash and living in *barrios* with two or more treatment modes were more polarized in their preference concerning their treatment mode (cash). In these *barrios*, the proportion of individuals preferring to receive all or none of the transfers as cash was greater than those where cash was the only treatment mode. We must stress that this effect is not statistically significant and is small in magnitude.

Contamination between clusters is one potential explanation for the differences we found. For example, more beneficiaries may appreciate the ease of using cash when observing beneficiaries receiving vouchers or food transfers. Similarly, the increase in the share of beneficiaries rejecting cash might be related to the crowding-out effect of cash. Households receiving cash were significantly less likely to receive loans and the like, which was not the case for households receiving food or vouchers. Hence, some cash beneficiaries may perceive receiving vouchers or food transfers as more advantageous, as the available budget used for consumption does not increase exactly with the amount of the cash transfer.

However, we also found observational differences at baseline between people living in *barrios* with one cluster and *barrios* with more than one cluster that show the latter had higher food consumption levels at baseline, which could explain why their preferences shifted differently than those living in *barrios* with only one treatment mode. Therefore, although contamination may have played a role in shaping preferences, there are other potential explanations for the differences; we cannot be sure that contamination took place and, if so, to what extent.

We also investigated whether some *barrios* were more influential by conducting a leave-one-out analysis. We identified a few cases where the p-values of treatment coefficients changed and led to crossing a significance threshold. Although we could not identify a distinct pattern, such *barrios* were more likely to have two or more distinct treatment modes and to be in Sucumbíos. Similarly, the two provinces had some independent statistical effect on the outcome variables. Hence, including additional explanatory variables at the provincial and/or *barrio* level might have been worthwhile in terms of investigating the effect of spatial variation. Overall, we believe the results are quite robust to variations in the included geographical units. This robustness was also supported by the GLMM approach, which accounted explicitly for the hierarchical and geographical nature of the data. Nevertheless, some contamination across clusters with different treatment modes took place. This should be kept in mind when designing similar studies in the future.

The theory of change analysis extended the original paper's cost-effectiveness analysis. When not accounting for uncertainty in the cost-effectiveness data, we can support the authors' claim that vouchers are the most cost-effective mode of transfer. However, the probabilistic sensitivity analysis, which quantifies the underlying uncertainty, shows that for three outcomes measure at lower values of the WTP range (log caloric intake: approximately USD20; HDDS: approximately USD10; PFC: approximately USD60), cash has a higher probability of being the most cost-effective intervention. This finding, together with the analysis of the expected value of perfect information, suggests that – in particular for lower values of WTP, as it is often the case in low resource settings – additional research might be advisable to truly discern whether cash or vouchers are the most cost-effective mode of transfer.

A limitation in our replication study was that the provided data set apparently did not contain the actual raw data. The accompanying .do file (the programming commands, or "code," in Stata-readable format) did not show the steps taken to create the final data set.[1] Several variables that were based on combination of underlying variables – e.g. food or asset indices – were already present. Hence, we were unable to investigate any potential steps taken for data cleaning, variable construction, recoding or labeling during the creation of the provided data set. Moreover, our cost-effectiveness analysis was limited by the fact we had no information on the uncertainty of the cost data.

[1] This was confirmed by the authors in a personal communication.

8. Conclusion

In our replication, we do not find any meaningful differences in Hidrobo and colleagues' reported results or in their interpretation. Hence, we follow the original authors' main conclusion: all three modes of assistance – cash, food and vouchers – increase the quality and quantity of food consumption of the targeted households. Households receiving food transfers had the largest increase in calories, while vouchers had the largest effect in terms of food diversity. However, we believe the original study was underpowered – with a sample size too small to detect a statistically significant difference between the three transfer modes – to make conclusive statements regarding the relative superiority. With the available information, we are unable to exactly replicate some of the indices used as outcomes, leading to small discrepancies between the original results and our analysis. In particular, we find that with our reconstructed food consumption score, cash had a similar positive effect on PFC as food hade, while the analysis with the original score suggests a superior effect of food over cash.

In our measurement and estimation analysis, we find some evidence for different treatment preferences when comparing clusters with only one treatment mode to clusters with two or more treatment modes. However, due to observational differences in food consumption and other characteristics at baseline, it remains unclear if contamination across clusters took place or if these baseline differences were affecting preferences. For at least one outcome, we find an indication that the effect estimates differed by province. We believe this small, yet statistically significant difference can be explained by the differing starting conditions in the two provinces, again showing the importance of context in which an intervention takes place.

We first extended the relatively limited cost-effectiveness analysis of the main paper to calculate ICERs using the available point estimates and identify the most cost-effective (dominant) intervention. When not accounting for uncertainty in the cost-effectiveness data, we can support the original paper's claim that vouchers are the advisable mode of transfer for all outcome measures. Even for those outcomes where direct food transfers have the largest effect estimate, we believe the additional funds needed for food transfers seem to be better spent on increasing the value of a voucher transfer.

In our theory of change analysis, we further extended the analysis by conducting a probabilistic sensitivity analysis to gauge the influence of the uncertainty inherent in the data and the resulting probability of decision makers to choose the most cost-effective intervention based on the available information. Unfortunately, we had to make assumptions about the uncertainty of the cost data, due to the unavailability of data-based uncertainty measures of the intervention costs. Our results revealed that for some outcomes (e.g. caloric intake per capita), decision makers with a relatively small budget to spend on a program carried a relatively larger risk of making an erroneous decision, i.e. not choosing the most cost-effective analysis. Related to this, our findings indicate a need for further research into the comparative cost-effectiveness of these interventions. This would be particularly worthwhile for programs intended to be implemented in low-resource settings, given the potential reduction in decision uncertainty such research could achieve for small(er) budgets.

Appendix A: Push-button replication

List of materials obtained

- Final report of impact evaluation of the project the paper is based on;
- Data file in Stata format containing data for replication; and
- Do-file to replicate results.

Classification

Comparable replication

Replication process

The data were obtained via email from Melissa Hidrobo. The path names for the files indicated in the .do file were modified in order to run the code. The code was modified to additionally report p-values. After obtaining the results, we identified which results belonged to which tables in the original paper, based on the commented code, the paper's description of each table and the generated output file. Once we identified the results, we replicated the tables as they appear in the paper. Finally, we compared the coefficients and significance level of each table and reported the results.

Push-button replication classification justification

Comparable replication. From seven tables in manuscript, six could be replicated using the provided data and Stata code. Results are comparable.

No replication code for Table 6, "Modality specific cost of improving outcomes by 15%," is available. However, this table is not driven by the results of the data analysis but by model calculations. We were able to replicate these calculations by hand using the information and data from the original paper and its appendix.

Push-button replication tables

A) Descriptions

Table 1: Comparable
Similar coefficients and p-values

Table 2: Comparable
Similar coefficients and p-values

Table 3: Comparable
Similar coefficients and p-values
Table 4: Comparable
Similar coefficients and p-values

Table 5: Comparable
Similar coefficients and p-values

Table 6: No code
Replication by hand using information from paper and supplementary material did yield similar results.

Table 7: Comparable
Similar coefficients and p-values.

B) Replication tables

	Comparable
	Minor differences
	Major differences
	No data to produce replication
	Information not reported in table

Table A1: Baseline mean characteristics by intervention arms

		Means				P-value of diff.					
	N	Control	Food	Cash	Voucher	Food - Control	Cash - Control	Voucher - Control	Food - Cash	Food - Voucher	Cash - Voucher
Attrition Rates	2357	0.11	0.08	0.09	0.11	0.11	0.34	0.80	0.43	0.18	0.49
Household head is female	2087	0.26	0.25	0.28	0.29	0.96	0.49	0.37	0.48	0.37	0.87
Household head is Colombian	2087	0.37	0.28	0.24	0.26	0.22	0.05	0.08	0.57	0.76	0.71
Household head is married	2087	0.28	0.30	0.28	0.27	0.59	0.89	0.65	0.50	0.33	0.74
Age of household head	2087	41.71	41.13	41.42	42.21	0.57	0.80	0.68	0.80	0.36	0.55
Household head secondary edu or higher	2087	0.32	0.35	0.35	0.38	0.49	0.54	0.21	0.92	0.63	0.54
Number of children 0-5 years	2087	0.59	0.66	0.59	0.62	0.25	0.95	0.58	0.28	0.56	0.62
Number of children 6-15 years	2087	1.02	0.90	0.89	0.83	0.21	0.12	0.01	0.91	0.44	0.45
Household size	2087	4.12	3.91	3.82	3.75	0.17	0.03	0.01	0.57	0.33	0.64
Floor type: dirt	2087	0.06	0.04	0.03	0.04	0.51	0.24	0.57	0.61	0.91	0.50
Owns television	2087	0.81	0.82	0.79	0.82	0.93	0.46	0.77	0.39	0.83	0.24
Owns computer	2087	0.27	0.32	0.29	0.29	0.28	0.56	0.54	0.54	0.53	1.00
Owns mobile phone	2087	0.85	0.80	0.82	0.84	0.11	0.25	0.63	0.61	0.25	0.50
Owns Car/truck/motorcycle	2087	0.24	0.22	0.24	0.24	0.47	0.81	0.99	0.58	0.43	0.77
Owns land	2087	0.13	0.12	0.12	0.13	0.78	0.74	0.80	0.95	0.65	0.57
Dietary diversity index	2087	17.02	17.44	17.41	17.28	0.53	0.51	0.64	0.95	0.77	0.77
Household dietary diversity score	2087	9.11	9.22	9.23	9.19	0.57	0.49	0.61	0.95	0.87	0.78

	N	Means				P-value of diff.					
		Control	Food	Cash	Voucher	Food - Control	Cash - Control	Voucher - Control	Food - Cash	Food - Voucher	Cash - Voucher
Food consumption score	2087	59.05	60.93	60.00	59.75	0.35	0.57	0.66	0.63	0.52	0.86
Total consumption per capita (monthly)	2087	111.03	110.46	111.02	114.35	0.93	1.00	0.59	0.94	0.55	0.58
Non food consumption per capita (monthly)	2087	65.65	63.50	62.60	65.11	0.68	0.55	0.91	0.86	0.74	0.59
Food consumption per capita (monthly)	2087	45.38	46.96	48.42	49.23	0.63	0.26	0.15	0.64	0.47	0.75
Caloric intake per capita (daily)	2087	2,021.38	1,803.24	1,922.36	1,813.00	0.26	0.60	0.25	0.25	0.92	0.17
P-value from joint F-test			0.87	0.63	0.68						

P-values are reported from tests on the equality of means for each variable. Standard errors are clustered at the cluster level. F-tests of joint significance: test of joint significance in regression of respective treatment dummies on all 17 baseline variables

Table A2: Impact of treatment arms on food and non-food consumption

	LEVELS			LOGS		
	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)	Food consumption (per capita)	Non-food consumption (per capita)	Total consumption (per capita)
Food treatment	9.22	9.22	18.50	0.20	0.15	0.17
	(2.79)***	(3.30)***	(5.02)***	(0.04)***	(0.07)**	(0.05)***
	[0.00]	[0.01]	[0.00]	[0.00]	[0.03]	[0.00]
Cash treatment	5.47	6.81	12.66	0.14	0.07	0.11
	(2.56)**	(3.93)*	(5.09)**	(0.04)***	(0.06)	(0.04)***
	[0.03]	[0.09]	[0.01]	[0.00]	[0.26]	[0.01]
Voucher treatment	6.38	6.78	13.45	0.15	0.13	0.13
	(2.58)**	(2.82)**	(4.38)***	(0.04)***	(0.06)**	(0.04)***
	[0.01]	[0.02]	[0.00]	[0.00]	[0.02]	[0.00]
R^2	0.21	0.17	0.22	0.26	0.25	0.25
N	2087	2087	2087	2087	2087	2087
Baseline Mean	47.54	64.29	111.83	3.67	3.84	4.54
P-value: Food=Voucher	0.31	0.46	0.33	0.23	0.75	0.40
P-value: Cash=Voucher	0.73	0.99	0.88	0.80	0.35	0.63
P-value: Food=Cash	0.17	0.57	0.30	0.14	0.27	0.21

Standard errors in parenthesis clustered at the cluster level. p-values in brackets, * p<0.1 ** p<0.05; *** p<0.01. All estimations control for baseline outcome variable and province.

Table A3: Impact of treatment arms on food security outcomes

	Log caloric intake (per capita)	HDDS	DDI	FCS	Poor food consumption
Food treatment	0.21	0.61	2.36	6.96	-0.05
	(0.04)***	(0.12)***	(0.44)***	(1.22)***	(0.02)***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Cash treatment	0.12	0.47	2.64	6.57	-0.02
	(0.04)***	(0.11)***	(0.42)***	(1.29)***	(0.02)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.13]
Voucher treatment	0.18	0.60	3.13	9.56	-0.04
	(0.04)***	(0.12)***	(0.45)***	(1.39)***	(0.02)***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
R^2	0.17	0.16	0.27	0.16	0.08
N	2087	2087	2087	2087	2087
Baseline Mean	1895.43	9.18	17.27	59.86	0.11
P-value: Food=Voucher	0.40	0.86	0.07	0.07	0.73
P-value: Cash=Voucher	0.15	0.16	0.22	0.05	0.13
P-value: Food=Cash	0.03	0.12	0.48	0.77	0.09

Standard errors in parenthesis clustered at the cluster level. p-values in brackets, * p<0.1 ** p<0.05; *** p<0.01. All estimations control for baseline outcome variable and province.

Table A4: Impact of treatment arms on food frequency and caloric intake by food groups - Panel A

	Outcome variable: Number of days in the last week household consumed...											
									In-kind food items			
	Roots & Tubers	Vegetables	Fruits	Meat & poultry	Eggs	Milk & dairy	Sugar & honey	Other	Cereals	Fish & seafood	Pulses legumes & nuts	Oils & fats
Food treatment	0.30	0.15	0.31	0.27	0.04	0.38	0.06	-0.19	0.36	0.77	1.22	0.04
	(0.16)*	(0.11)	(0.16)*	(0.09)***	(0.15)	(0.17)**	(0.08)	(0.19)	(0.10)***	(0.13)***	(0.15)***	(0.11)
	[0.07]	[0.16]	[0.06]	[0.00]	[0.80]	[0.02]	[0.43]	[0.33]	[0.00]	[0.00]	[0.00]	[0.72]
Cash treatment	0.33	0.33	0.13	0.39	0.25	0.70	-0.04	0.05	0.07	0.25	0.58	-0.07
	(0.17)**	(0.10)***	(0.15)	(0.11)***	(0.15)*	(0.17)***	(0.09)	(0.19)	(0.10)	(0.09)***	(0.12)***	(0.11)
	[0.05]	[0.00]	[0.36]	[0.00]	[0.10]	[0.00]	[0.64]	[0.80]	[0.46]	[0.00]	[0.00]	[0.51]
Voucher treatment	0.48	0.40	0.28	0.35	0.47	0.98	-0.04	0.02	0.21	0.48	0.79	-0.10
	(0.17)***	(0.10)***	(0.14)**	(0.11)***	(0.14)***	(0.19)***	(0.09)	(0.19)	(0.10)**	(0.09)***	(0.12)***	(0.11)
	[0.01]	[0.00]	[0.04]	[0.00]	[0.00]	[0.00]	[0.64]	[0.93]	[0.03]	[0.00]	[0.00]	[0.33]
N	2087	2087	2087	2087	2087	2087	2087	2087	2087	2087	2087	2087
Baseline Mean	5.15	6.07	4.50	1.91	3.65	2.92	6.45	4.41	6.22	0.85	1.53	0.40
P-value: Food=Voucher	0.35	0.02	0.88	0.41	0.01	0.00	0.17	0.30	0.09	0.03	0.01	0.16
P-value: Cash=Voucher	0.43	0.44	0.34	0.75	0.19	0.16	0.98	0.87	0.13	0.01	0.09	0.73
P-value: Food=Cash	0.87	0.11	0.32	0.25	0.23	0.07	0.18	0.21	0.00	0.00	0.00	0.28

Standard errors in parenthesis clustered at the cluster level. p-values in brackets, * p<0.1 ** p<0.05; *** p<0.01. All estimations control for baseline outcome variable and province.

Table A4: Impact of treatment arms on food frequency and caloric intake by food groups - Panel B

	Outcome variable: Log per capita caloric intake (daily) ...											
										In-kind food items		
	Roots & Tubers	Vegetables	Fruits	Meat & poultry	Eggs	Milk & dairy	Sugar & honey	Other	Cereals	Fish & seafood	Pulses legumes & nuts	Oils & fats
Food treatment	0.24	0.20	0.24	0.37	0.09	0.51	0.05	0.13	0.22	1.29	0.90	0.16
	(0.12)**	(0.07)***	(0.11)**	(0.11)***	(0.09)	(0.16)***	(0.06)	(0.10)	(0.05)***	(0.17)***	(0.15)***	(0.10)
	[0.04]	[0.01]	[0.03]	[0.00]	[0.36]	[0.00]	[0.44]	[0.18]	[0.00]	[0.00]	[0.00]	[0.13]
Cash treatment	0.19	0.24	0.16	0.46	0.06	0.68	0.05	0.06	0.16	0.49	0.44	0.06
	(0.10)*	(0.07)***	(0.10)	(0.12)***	(0.08)	(0.13)***	(0.06)	(0.10)	(0.06)***	(0.14)***	(0.13)***	(0.09)
	[0.07]	[0.00]	[0.11]	[0.00]	[0.44]	[0.00]	[0.40]	[0.57]	[0.01]	[0.00]	[0.00]	[0.49]
Voucher treatment	0.16	0.24	0.21	0.42	0.19	0.89	0.09	0.04	0.17	0.57	0.63	0.07
	(0.10)	(0.07)***	(0.10)**	(0.11)***	(0.09)**	(0.15)***	(0.06)	(0.10)	(0.06)***	(0.12)***	(0.13)***	(0.09)
	[0.13]	[0.00]	[0.04]	[0.00]	[0.03]	[0.00]	[0.12]	[0.73]	[0.00]	[0.00]	[0.00]	[0.45]
N	2087	2087	2087	2087	2087	2087	2087	2087	2087	2087	2087	2087
Baseline Mean	146.31	29.53	198.52	142.76	37.33	102.50	317.73	24.21	818.46	22.20	49.16	6.72
P-value: Food=Voucher	0.46	0.53	0.72	0.66	0.30	0.01	0.47	0.35	0.32	0.00	0.04	0.32

	Outcome variable: Log per capita caloric intake (daily) ...											
										In-kind food items		
	Roots & Tubers	Vegetables	Fruits	Meat & poultry	Eggs	Milk & dairy	Sugar & honey	Other	Cereals	Fish & seafood	Pulses legumes & nuts	Oils & fats
P-value: Cash=Voucher	0.75	0.95	0.56	0.75	0.15	0.10	0.52	0.83	0.74	0.41	0.11	0.94
P-value: Food=Cash	0.63	0.50	0.39	0.45	0.81	0.22	0.94	0.45	0.15	0.00	0.00	0.31

Standard errors in parenthesis clustered at the cluster level. p-values in brackets, * p<0.1 ** p<0.05; *** p<0.01. All estimations control for baseline outcome variable and province.

Table A5: Heterogeneous impact on caloric intake

	Log adult equiv. caloric intake	Log adult equiv. caloric intake
Food treatment	0.22	0.28
	(0.04)***	(0.05)***
	[0.00]	[0.00]
Cash treatment	0.12	0.12
	(0.03)***	(0.05)**
	[0.00]	[0.01]
Voucher treatment	0.17	0.19
	(0.04)***	(0.05)***
	[0.00]	[0.00]
Food X High caloric intake (kcals>2100)		-0.13
		(0.06)**
		[0.04]
Cash X High caloric intake (kcals>2100)		0.01
		(0.06)
		[0.93]
Voucher X High caloric intake (kcals>2100)		-0.05
		(0.06)
		[0.33]
High caloric intake (kcals>2100)		0.11
		(0.05)**
		[0.04]
Constant	5.44	5.75
	(0.23)***	(0.33)***
	[0.00]	[0.00]
R^2	0.12	0.13
N	2087	2087
P-value: Food=Voucher	0.22	0.09
P-value: Cash=Voucher	0.15	0.09
P-value: Food=Cash	0.01	0.00

Standard errors in parenthesis clustered at the cluster level. P-values in brackets, * p<0.1 ** p<0.05; *** p<0.01. All estimations control for baseline outcome variable and province.

Table A6: Transfer preference, by treatment status

	Means			P-value of diff.		
	Food	Cash	Voucher	Food - Cash	Food - Voucher	Cash - Voucher
All	0.55	0.77	0.56	0.00	0.91	0.00
None	0.28	0.09	0.31	0.00	0.45	0.00
N	341.00	425.00	441.00			

P-values are reported from Wald tests on the equality of means for each variable.
Standard errors are clustered at the cluster level.

Appendix B: Code replicated in R

Has been made available online.

https://www.3ieimpact.org/sites/default/files/2021-07/RPS17-Pure-Replication-Cash-Food-Vouchers-appendix-B.pdf

References

Barrientos, A and Hulme, D, 2008. Social protection for the poor and poorest in developing countries: Reflections on a quiet revolution. *BWPI Working Paper 30.* Manchester: University of Manchester.

Barton, GR, Briggs, AH and Fenwick, EAL, 2008. Optimal Cost-Effectiveness Decisions: The Role of the Cost-Effectiveness Acceptability Curve (CEAC), the Cost-Effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfection Information (EVPI). *Value in Health,* 11(5), pp. 886–897.

Briggs, A, Claxton, K and Schulper, M, 2006. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press.

Briggs, AH, Weinstein, MC, Fenwick, EA, Karnon, J, Sculpher, MJ, Paltiel, AD and Force, I-SMGRPT, 2012. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Medical Decision Making,* 32, pp.722–732.

Brown, AN, Cameron, DB and Wood, BDK, 2014. Quality evidence for policymaking: I'll believe it when I see the replication. *Journal of Development Effectiveness,* 6, pp.215–235.

Campbell, MK, Piaggio, G, Elbourne, DR and Altman, DG, 2012. Consort 2010 statement: extension to cluster randomised trials. *BMJ*, 345, e5661.

Chow, S and Liu, J, 2003. *Design and Analysis of Clinical Trials*. Hoboken: John Wiley & Sons, Inc.

Drummond, MF, Sculpher, MJ, Claxton, K, Stoddart, GL and Torrance, GW, 2015. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press.

Ekwunife, OI and Lhachimi, SK. 2017, Cost-effectiveness of Human Papilloma Virus (HPV) vaccination in Nigeria: a decision analysis using pragmatic parameter estimates for cost and programme coverage. *BMC Health Services Research*, 17(1), p. 815.

Eldridge, S and Kerry, S, 2012. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. Chichester: John Wiley & Sons.

Fiszbein, A and Schady, N, 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington, DC: World Bank.

Fiszbein, A, Schady, N, Ferreira, FHG, Grosh, M, Keleher, N, Olinto, P and Skoufias, E, 2009. Conditional cash transfers: reducing present and future poverty. *World Bank Policy Research Report*. Washington, DC: World Bank.

Gelman, A and Hill, J, 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Gentilini, U, 2016. Revisiting the “Cash versus Food” Debate: New Evidence for an Old Puzzle? *The World Bank Research Observer,* 31, pp.135–167.

Hidrobo, M, Hoddinott, J, Peterman, A, Margolies, A and Moreira, V, 2014. Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador. *Journal of Development Economics,* 107, pp.144–156.

Kennedy, G, Ballard, T and Dop, MC, 2011. *Guidelines for measuring household and individual dietary diversity*. Food and Agriculture Organization of the United Nations.

Keogh-Brown, M, Bachmann, M, Shepstone, L, Hewitt, C, Howe, A, Ramsay, C, Song, F, Miles, J, Torgerson, D, Miles, S, Elbourne, D, Harvey, I and Campbell, M, 2007. Contamination in trials of educational interventions. *Health Technology Assessment*, 11(43).

Lagarde, M, Haines, A and Palmer, N, 2009. The impact of conditional cash transfers on health outcomes and use of health services in low and middle income countries. *Cochrane Database of Systematic Reviews*, CD008137.

Lee, DS, 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies,* 76, pp.1071–1102.

Lhachimi, SK, 2017. Technical proposal for “Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador.” *Call Replication Window 4.* Washington, DC: International Initiative for Impact Evaluation (3ie). Available from: <http://www.3ieimpact.org/media/filer_public/2017/07/26/lhachimi-replication-plan.pdf>. [Accessed 13 September 2018].

Marmot, M, Allen, J, Bell, R, Bloomer, E and Goldblatt, P, on behalf of the Consortium for the European Review of Social Determinants of Health and the Health Divide, 2012. WHO European review of social determinants of health and the health divide. *The Lancet,* 380, pp.1011–1029.

Pega, F, Walter, S, Liu, SY and Lhachimi, SK, 2015. Unconditional cash transfers for assistance in humanitarian disasters: effect on use of health services and health outcomes in low- and middle-income countries. *Cochrane Database of Systematic Reviews*, CD011247.

Pega, F, Walter, S, Liu, SY, Pabayo, R, Saith, R and Lhachimi, SK, 2017. Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low- and middle-income countries. The Cochrane Library.

Stollenwerk, B, Lhachimi, SK, Briggs, A, Fenwick, E, Caro, JJ, Siebert, U, Danner, M and Gerber-Grote, A, 2015. Communicating the Parameter Uncertainty in the IQWiG Efficiency Frontier to Decision-Makers. *Health Economics*, 24(4), pp. 481–490.

Westwood, M, Al, M, Burgers, L., Redekop, K, Lhachimi, S, Armstrong, N, Raatz, H, Misso, K, Severens, J and Kleijnen, J, 2013. A systematic review and economic evaluation of new-generation computed tomography scanners for imaging in coronary artery disease and congenital heart disease: Somatom Definition Flash, Aquilion ONE, Brilliance iCT and Discovery CT750 HD. *Health Technology Assessment,* 17, pp.1–243.

WFP, 2008. Food consumption analysis: calculation and use of the food consumption score in food security analysis. Rome: World Food Programme (WFP).

World Bank, 2011. *Safety Nets and Transfers: Glossary* [Online]. Available from: <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALPROTECTION/EXTSAFETYNETSANDTRANSFERS/0,,contentMDK:22200792~pagePK:210058~piPK:210062~theSitePK:282761,00.html>. [Accessed 15 July 2017].

Other publications in the 3ie Replication Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-replication-paper-series/

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

3ie 10 YEARS

www.3ieimpact.org