

Gender gaps in earnings and economic empowerment: Experimental evidence from 6 countries*

Paul Christian[†] Lelys Dinarte[‡] Jonas Heirman[§] Saahil Karpe[†]
Erin Kelley[†] Florence Kondylis[†] Gregory Lane[¶] John Loeser[†]

May 26, 2020

This pre-analysis plan is not yet registered.
Please do not cite or circulate.

Abstract

Can equalizing market opportunities for men and women contribute to closing the gender gap in autonomy? We randomize participation in and participant gender for WFP cash-for-work programs that target poor households in 6 countries. We leverage this variation to estimate the impacts of women’s participation in the program, adjusting for household impacts of program participation, on women’s autonomy.

*This design benefited from comments from Macartan Humphreys and Sarah Khan. This research benefits from generous funding from the Office of Evaluation of the World Food Programme. Rasha Absi, Sylvia Cesar, Mariachiara Iannuzzi, Eric Jospe, Edwin Kasila, Kristen McCollum, and Miguel Melendez provided excellent research assistance. Finally, we thank the management and technical staff at WFP Country Offices for being outstanding research partners. The views expressed do not reflect the views of the World Bank nor the World Food Programme. All errors are our own. Corresponding authors: Florence Kondylis (fkondylis@worldbank.org) and John Loeser (jloeser@worldbank.org).

[†]Development Impact Evaluation, World Bank

[‡]Development Research Group, World Bank

[§]Office of Evaluation, World Food Programme

[¶]American University

Contents

1	Introduction	3
2	Experimental designs	8
2.1	Program details and sampling	8
2.2	Countries	9
2.2.1	Country 1	10
2.2.2	Country 2	10
2.2.3	Country 3	11
2.3	Overall	12
3	Data	12
3.1	Timeline	12
3.2	Survey	13
3.2.1	Consumption	13
3.2.2	Earnings	14
3.2.3	Time use	14
3.2.4	Agency	14
3.2.5	Attitudes	15
3.2.6	Perceptions of norms	15
3.2.7	Well being	16
3.2.8	Outcomes for WFP M&E	17
4	Analysis	17
4.1	Empirical Strategy	17
4.1.1	Discussion of potential exclusion restriction violations	22
4.2	Power calculations	23
4.3	Results	26

1 Introduction

Can equalizing market opportunities for men and women contribute to closing the gender gap in autonomy? Gender inequality in autonomy is pervasive, and its welfare implications are particularly concerning in developing countries (Jayachandran, 2015). Economic development, gender equality in economic opportunities, and gender equality in autonomy are all strongly linked, but causality is unclear. Further, institutions (“*Women cannot drive*”) and gender norms and attitudes (“*Women should not work outside the household*”) are both root causes of gender inequality in economic opportunities, and also mediate impacts of economic development on gender equality in autonomy. Unpacking these relationships requires experimentally manipulating both economic development and men’s and women’s equality of opportunities, estimating their impacts on gender equality in autonomy, and observing how these changes reflect back on institutions, attitudes, and norms.

We estimate the impacts of experimentally manipulating men’s and women’s wages on household decision-making and the autonomy gender gap. We design and run these experiments in the context of livelihoods programs implemented by the World Food Programme (WFP) across 6 countries.¹ These livelihoods programs provide households with a cash transfer conditional on a particular household member engaging in productive activities, which we interpret as a wage. In each country, we implement two experimental arms: the standard livelihoods program (“Cash-for-Work”), and a modified version that targets women (“Cash-for-Women’s Work”). These two arms allow us to separately estimate the impacts of economic development (Cash-for-Work) and gender equality in economic opportunity (Cash-for-Women’s Work relative to Cash-for-Work) on household decision-making.

We experimentally manipulate men’s and women’s wages in the context of “livelihoods” programs, a class of widely implemented development programs targeting poor households in developing countries (Banerjee et al., 2015). The specific programs implemented by WFP

¹At the time of publication of this pre-analysis plan, El Salvador, Kenya, and Syria have been selected as 3 of the 6 countries. Selection of 3 additional countries is ongoing.

that we study are relatively uniform in their design across contexts, enabling comparisons across diverse contexts, and have successfully been scaled.² Typically, these programs selectively target communities identified as particularly vulnerable, and focus on poor households within these communities with the objective of promoting resilience. In contrast to many programs which target the “ultra-poor”, these programs do not make large unconditional cash or asset transfers. Instead, households are paid a wage to engage in activities linked to the creation of productive assets (e.g., chicken and cow pens, kitchen gardens, public works, etc.), with monitoring to ensure compliance. The typical wage in these programs is enough to purchase a standard food basket for a family of four; this is typically below market wages, but the work is easier than market work and upon program completion participating households may benefit from the produced assets.

To interpret impacts of equalizing wages, we consider how shifts in wages should impact household decision-making in a unitary household model. In this standard framework, household members make labor supply decisions trading off household consumption gains with the opportunity costs of working for a wage, including foregone leisure and home production. A striking empirical observation is that women taking on work for a wage often substitute away from leisure (“second shift”, [Hochschild & Machung \(2012\)](#)), while men do not shift into home production ([Bertrand et al., 2015](#)). In the standard framework, this is explained by differences in men’s and women’s utility functions or their home production functions. However, a large body of empirical work rejects the unitary household model ([Browning & Chiappori, 1998](#); [Ashraf, 2009](#)), with a key mechanism that men and women have agency over different household decisions (“separate spheres”, [Lundberg & Pollak \(1993\)](#)). Recent experimental work has demonstrated that attitudes ([Dhar et al., 2018](#); [McKelway, 2019](#)) and norms ([Beaman et al., 2009](#); [Bursztyn et al., 2018](#)) shape women’s agency and, in turn, women’s labor supply. In practice, norms, attitudes, and women’s agency are also likely endogenous to women’s labor supply decisions; if so, shifts in women’s wages may also af-

²The government of Ethiopia adopted a similar WFP program and scaled it nationally (CITE).

fect household decision-making through these channels. These impacts may further cascade, leading to persistent shifts in women’s labor supply in response to temporary shocks to women’s wages (Alesina et al., 2013; Goldin & Olivetti, 2013).

We first estimate the impacts of men’s and women’s wages on their labor supply. We begin by estimating the first stage impacts of Cash-for-Work and Cash-for-Women’s Work on wage earnings. We anticipate Cash-for-Work increases household wage earnings, and that “Targeting Women” (Cash-for-Women’s Work relative to Cash-for-Work) increases women’s wage earnings without increasing total household wage earnings or income. As such, these increases in women’s wage earnings come with decreases in men’s wage earnings, and we interpret the effects of Targeting Women as the combined impacts of these shifts. Next, we estimate the effect of Cash-for-Work and Targeting Women on time use. We anticipate the effects of Cash-for-Work will depend on contextual features of the program. We anticipate that Targeting Women will increase the time women spend working outside the home, while causing women to substitute away from leisure, sleep, and home production. We also anticipate that the relative magnitudes of these effects will depend on baseline women’s agency, norms, attitudes, and household structure.

Building on our conceptual framework, we estimate the impacts of shifting women’s time spent working on household decision making and explore mechanisms. First, we estimate impacts on household consumption decisions. To do this, we use the control group to estimate the cross-sectional relationship between women’s income and consumption expenditures by good, controlling for total household consumption. We anticipate that Targeting Women increases “predicted women’s income”, providing evidence of changes in household decision-making over consumption. Second, we estimate the impacts on mediating forces underlying these changes in household consumption patterns. To do so, we leverage rich data we collect on women’s agency, attitudes, perceptions of norms, and intimate partner violence (IPV). We do not have strong priors over how these might shift; shifts in each of these outcomes reflects a potential mechanism through which changes in women’s time spent working can

impact household decision making.

We then estimate the impacts of these anticipated changes in women’s autonomy on women’s labor supply after the end of the livelihoods programs. As the program will have ended, there are no longer differences in wages between women in the Cash-for-Work and Cash-for-Women’s Work arms. Therefore, any differences in time allocations can be attributed to shifts in women’s labor supply, which we assume are caused by changes in women’s agency, attitudes, and perceptions of norms. We will therefore estimate impacts of Targeting Women, after the end of the livelihoods programs, on both women’s time allocations and women’s agency, attitudes, and perceptions of norms.

We explore additional mechanisms by estimating heterogeneous responses. First, we test for intensive and extensive margin impacts. We compare impacts across women who previously engaged in salaried work to those who did not. Second, we test for heterogeneity with respect to baseline women’s agency. While women with greater agency may be more likely to substitute away from home production than leisure in response to Cash-for-Women’s Work, these women may also see smaller changes in their agency (Field et al., 2019). Lastly, we test for heterogeneity with respect to household structure.³

Do these heterogeneous impacts explain differences in the impacts of Targeting Women across the 6 countries? In practice, for each outcome we test whether heterogeneity in impacts of Targeting Women across the 6 countries are jointly significant, once heterogeneity in impacts with respect to the observable characteristics described above is controlled for. This provides a test of external validity, by testing whether these characteristics mediate heterogeneity in the impacts of the intervention across countries (Wilke & Humphreys, 2020).

We build on a deep literature at the nexus of economic development, gender equality, and women’s autonomy. Our experimental design is closest to studies that have experimentally varied the gender of the recipients in cash transfer program (Akresh et al., 2013; Benhassine

³Recent evidence on divorce in Sweden suggests that wage shocks for women may differentially impact household decision making for households with larger age gaps (i.e., relatively older male spouses) between spouses (Folke & Rickne, 2020).

et al., 2015; Armand et al., 2016; Haushofer & Shapiro, 2016).⁴ In contrast to cross-sectional (Thomas, 1990) and quasi-experimental evidence (Lundberg et al., 1997; Duflo, 2000), this experimental work has typically found limited impacts of shifting beneficiaries from men to women on household consumption decisions. This contrasts further with evidence that women are willing to forgo income to be named the recipient Almås et al. (2018), and that consumption decisions are impacted by their observability Ashraf (2009). We provide additional evidence by instead manipulating the design of cash-for-work programs to generate experimental variation in the wage men and women are offered.

As such, our research questions are closest to recent work that has studied the impact of women’s labor force participation on agency, norms and attitudes. One strand of this literature has leveraged historical shocks to women’s productivity and found persistent impacts on women’s labor force participation (Alesina et al., 2013; Goldin & Olivetti, 2013). Our work is closest to Field et al. (2019), who finds that experimentally increased control over earnings causes women to increase their labor supply, and that these impacts persist with changes in attitudes towards women working as a mechanism. Relatedly, McKelway (2019) finds that success in applying for jobs increases generalized self efficacy, and that increased generalized self efficacy increases women’s labor supply. Relative to this work, we contribute by controlling for income effects through our research design, which is particularly important given strong evidence that shocks to income reduce IPV, regardless of the gender of the recipient (Haushofer et al., 2019). We further contribute on external validity, by providing evidence from 6 countries and shedding light on determinants of cross country heterogeneity.

Lastly, we complement work that directly shifts women’s agency, attitudes, and norms, and estimates impacts on women’s labor supply. This work is reviewed in Jayachandran (2019), and has found that interventions that target attitudes (Dhar et al., 2018; McKelway, 2019) and norms (Beaman et al., 2009; Bursztyn et al., 2018) can shift women’s labor supply, especially where attitudes and norms are biased against women. We contribute by showing

⁴An exception among these papers is (Armand et al., 2016), who find that food shares increases when cash transfer recipients are shifted from men to women.

that temporary shifts to women’s labor supply can also shift women’s agency, attitudes, and norms, generating sustained decreases in gender gaps in both labor force participation and autonomy.

The rest of the concept note is structured as follows. Section 2 describes the programs and experimental designs. Section 3 describes the data. Section 4 describes the planned analysis, and links to a conceptual framework.

2 Experimental designs

2.1 Program details and sampling

The program of interest in all countries is a version of WFP’s cash-for-assets (CFA) program. CFA functions similarly to cash-for-work programs, in that households are paid a transfer conditional on participating in household level or community work. Exact details vary by context, and are described in Section 2.2. The programs are described as “cash-for-assets” because the work component is intended to produce an asset that generates sustained increases in income or welfare for the household or the community, with examples of work ranging from rehabilitating feeder roads to constructing improved latrines to caring for livestock.

All experimental designs build on random assignment of households or communities to treatment arms which vary whether women participate in CFA, with additional arms to adjust for household or community level impacts of the CFA program. This allows us to estimate the impacts of participation (modeled as a shift in women’s wages) on intrahousehold decision-making. This requires:

1. Sampled households have both a man and a woman who would be eligible to participate in the cash-for-asset program, and the eligible man is a primary decision maker in the household⁵

⁵When feasible, both the man and the woman should be interested in participation, as this will improve

2. Each country has a control arm
3. Each country has one arm that induces the household to participate in the program (“Cash-for-Work”)
4. Each country has another arm that causes women to participate in the program (“Cash-for-Women’s Work”)

Ideally, all details of the program will be held fixed across Cash-for-Work and Cash-for-Women’s work. This includes the work requirement and the nature of the work itself. In some cases, this will not be feasible. In many countries, the assets are selected jointly by the household and the community, and this selection may be affected by the gender of the participant. In other cases, assets are gender segregated, so the selection of the asset implicitly determines the gender of the program participant. Differences in these selected assets may cause differences in the household’s time requirement for program participation, to give one example. Finally, in some cases men are excluded from participating in the program.⁶ In these cases, the Cash-for-Work arm will be an unconditional cash transfer. This enables holding fixed household income across Cash-for-Work and Cash-for-Women’s work, but does not hold fixed the time the household spends participating in the program. Section 2.2 lays out these differences, and Section 4 describes how we attempt to account for them in our analysis.

2.2 Countries

At this point, three of the six countries (from three continents) that will be participating in the evaluations have been selected. Discussions are still ongoing on details on the interven-

power as demonstrated in Section 4.2.

⁶In these cases, for sampling we still require that sampled households have both a man and a woman who satisfy the eligibility requirements, and that the man satisfying the eligibility requirements is a primary decision maker. We do so for comparability, and because we do not anticipate impacts on women’s empowerment within the household in households that do not have men involved in decision making for the household.

tions, so the countries themselves are not yet public. In this document we will refer to them by Country 1, Country 2, and Country 3.

2.2.1 Country 1

Table 1: Experimental Design: Country 1

	Female recipient	50% female, 50% male recipient
Control		TBD (TBD)
UCT	TBD (TBD)	TBD (TBD)
CFA (Female)	0 (0)	TBD (TBD)

In Country 1, the cash-for-asset (CFA) program includes a cash transfer conditional on participation in a community works program. All participants are female.

UCT is the “Cash-for-Work” arm, and CFA (Female) is the “Cash-for-Women’s work” arm. UCT does not have a work component, and implications of this for the analysis are discussed in Section 4. We do not plan to use the random assignment of recipients in the pre-specified analysis in this paper.

2.2.2 Country 2

Table 2: Experimental Design: Country 2

Control	TBD (TBD)
UCT	TBD (TBD)
CFA (Female)	TBD (TBD)
CFA (Male)	TBD (TBD)

In Country 2, the cash-for-asset (CFA) program includes an unconditional cash transfer and an asset transfer, with promotion of the livelihood tied to that asset. The CFA (Female) program provides chickens (which only women traditionally raise in that region), and the CFA (Male) program provides cattle (which only men traditionally raise in that region). The UCT arm will provide only the cash. Random assignment will occur at the community level.

UCT and CFA (Male) are the “Cash-for-Work” arms, and CFA (Female) is the “Cash-for-Women’s work” arm. UCT does not have a work component, and implications of this for the analysis are discussed in Section 4.

2.2.3 Country 3

Table 3: Experimental Design: Country 3

	E-Voucher	Unrestricted cash
Control		600
CFA (Female)	300	300
CFA (Male)	300	300

In Country 3, the cash-for-asset (CFA) program includes a cash transfer conditional on participating in a TBD (through qualitative work to be implemented by a development partner) livelihoods program. CFA (Female) and CFA (Male) restrict to male and female participants, respectively. Transfers are always made to the participant. Random assignment will occur at the household level. Reported sample sizes are tentative.

CFA (Male) is the “Cash-for-Work” arm, and CFA (Female) is the “Cash-for-Women’s work” arm. Randomization is stratified against assignment of modality of the transfer (E-Voucher or Unrestricted cash); we do not plan to use this arm in the pre-specified analysis in this paper.

2.3 Overall

Table 4: Implementation summary

	Country 1	Country 2	Country 3
Overall			
Level of randomization	Community	Community	Household
Program eligibility requirement			Poor
Additional sample requirements			Both a man and woman in household expressed interest in participation
Female and male respondents	Yes	TBD	No
...			
Cash-for-Work			
Name of arm	UCT	CFA (Male)	CFA (Male)
Cash transfer value	TBD/TBD	TBD/TBD	TBD/TBD
Total value	TBD	TBD	TBD
Work requirement	No	Yes	Yes
Share women participants	NA	TBD	0%
Asset transfer		Cow	
Asset value			
Most common work chosen			
2nd most common work chosen			
...			
Cash-for-Women's work			
Name of arm	CFA (Female)	CFA (Female)	CFA (Female)
Work requirement	Yes	Yes	Yes
Share women participants	100%	TBD	100%
Asset transfer		Chicken	
Asset value			
Most common work chosen			
2nd most common work chosen			
...			

3 Data

3.1 Timeline

The timeline of surveys and implementation is presented in Table 5. Baseline surveys will take place just prior to the start of the intervention. The WFP programs studied last between TBD and TBD months, averaging TBD months. A midline survey will take place half way

through the implementation of the program, and an endline survey will occur just after the end of the intervention. The reference period for the endline survey will exclude the period of the intervention.

Table 5: Timeline

	<u>Baseline</u>	<u>Intervention start</u>	<u>Midline</u>	<u>Intervention end</u>	<u>Endline</u>
Country 1	2020/08?	?	?	?	?
Country 2	?	?	?	?	?
Country 3	2020/08?	2020/08?	2020/11?	2021/02?	2021/05?

3.2 Survey

In all countries, a woman selected as eligible to participate in the cash-for-asset program is the primary respondent for the survey. Many questions in the survey will focus on both this woman, and also a man in the household who’s also eligible to participate in the cash-for-asset program. Some of these questions may be relatively difficult for women to answer, such as questions on the men’s time use. Therefore, in countries where it is feasible, the man who’s also eligible to participate (referred to as “primary male decision maker”) will be surveyed on a reduced set of questions. When the male respondent’s responses are used to construct an outcome, this is specified below.

All standardized indices below are constructed using inverse covariance weighting following [Anderson \(2008\)](#).

3.2.1 Consumption

Questions Expenditures over a standard reference period for up to 10 goods are asked. 5 goods are selected as the goods that most strongly predict household consumption in a household survey from the same context. 5 goods are selected as the goods that most strongly predict women’s income, controlling for total household consumption, in a household survey from the same context.⁷

⁷Additional details on the selection of goods are in Dropbox Paper.

Predicted household consumption Measured as predicted household consumption from the household’s expenditures, with coefficients used for prediction estimated in a household survey with a full consumption module from the same context.

Predicted women’s earnings Measured as predicted women’s earnings from the household’s expenditures, controlling for assigned treatments.

3.2.2 Earnings

Questions Earnings for each household member are collected during the household roster for the previous TBD months.

Outcomes Women’s Earnings, Men’s Earnings.

3.2.3 Time use

Questions The female respondent is asked for a 24 hour recall of her activities over the past two days, following the approach of American Time Use Survey. When the primary male decision maker in the household is available, he is asked about his activities over the past two days; when he is not, the female respondent is asked about his activities.

Outcomes Time spent outside the home (men and women), time spent working in self-employment (men and women), time spent working for a salary (men and women), time spent working on chores (men and women).

3.2.4 Agency

Questions The female respondent is asked, relative to the primary male decision maker in the household, how much her opinion would be considered in a series of decisions. These questions follow the DHS on consumption (“major household purchases”, “purchases from the primary male decision maker’s income”, “purchases from the female respondent’s income”,

“the female respondent’s health care”), and include additional questions on decision making over both men’s and women’s time in three productive activities (“work in self-employment”, “work for a salary”, “work on household chores”).

Outcomes Standardized indices over agency over consumption, and agency over women’s time use.

3.2.5 Attitudes

Questions (Attitudes towards time use) The female respondent is asked how much time she *should* spend, relative to the primary male decision maker in the household, on the three productive activities listed above.

Questions (Attitudes towards agency) The female respondent is asked how much her opinion *should* be considered, relative to the primary male decision maker in the household, in the same set of decisions as the Agency questions.

Outcomes Standardized indices over women’s attitudes towards women’s time use, and women’s attitudes towards women’s agency. In contexts where it is feasible to also survey men, men’s attitudes towards women’s time use and men’s attitudes towards women’s agency are included.

3.2.6 Perceptions of norms

Questions (Time use) The female respondent is asked how much time she believes women, relative to men, in her community spend on three productive activities.

Questions (Agency) The female respondent is asked how much the opinion of women in her community would be considered, relative to primary male decision makers in their households, on the same set of decisions as the Agency questions.

Questions (Attitudes) The female respondent is asked about the attitudes of people in her community. These questions mirror the above questions on attitudes towards time use and attitudes towards agency.

Outcomes Standardized indices over perceptions of norms over women’s time use, perceptions of norms over women’s agency, perceptions of norms over attitudes towards women’s time use, and perceptions of norms over attitudes towards women’s agency. In contexts where it is feasible to also survey men, men’s attitudes towards women’s time use and men’s attitudes towards women’s agency are included.

3.2.7 Well being

Questions (Subjective well being) Modules to measure locus of control, psychosocial well being, life satisfaction, and depression (PHQ9).

Questions (IPV) Questions from the DHS module on domestic violence are included, with adaptation of questions based on [Haushofer et al. \(2019\)](#).

Questions (Preferences, consumption) The female respondent is asked how much she would prefer the household spend on each of the consumption goods asked about in the survey.

Questions (Preferences, time use) The female respondent is asked how much she would prefer she and the primary male decision maker in the household should each spend on three productive activities.

Outcomes Locus of control score. Standardized index over {affect, life satisfaction score, PHQ9 score}. Standardized index over {standardized index of psychological abuse, standardized index of physical abuse, standardized index of sexual abuse}. Standardized index over preferences over consumption. Standardized index over preferences over time use.

3.2.8 Outcomes for WFP M&E

In each country, standard indicators will be collected for WFP M&E. These will be used exclusively for reports to WFP.

4 Analysis

4.1 Empirical Strategy

The main objective of the analysis is to estimate the impacts of women’s participation in the program, adjusting for any household level impacts of the program itself. To do this, we use the fact that Cash-for-Work shifts program participation at the household level, while Cash-for-Women’s work shifts both women’s participation in the program and program participation at the household level.

When comparing estimates across contexts, it is necessary to perform some standardization of effects. For example, [Banerjee et al. \(2015\)](#) report estimated effects on standardized indices of all outcomes, with weights for the index calculated using the control group in the same country and time period. For all outcomes except consumption and time use, we will similarly standardize. For consumption, we will normalize by mean household income in the control group. For time use, we will leave outcomes as hours per day.

In our case, comparisons across contexts are particularly complicated because we are interested in the impact of one intervention (Cash-for-Women’s work) controlling for an endogenous variable (program participation). However, Cash-for-Work is a plausible instrument for program participation, suggesting an instrumental variable estimator. In addition, it is also likely the case that the extent of participation in the program, in both arms, is likely to vary across contexts; as we are interested in studying the impacts of program participation, rather than of the randomly assigned arms themselves, this suggests further using Cash-for-Women’s work as an instrument for women’s participation in the program. Lastly,

we expect program participation has impacts through the channel of household income, and women’s participation in the program has impacts through the channel of women’s earned income. In order to compare across contexts, we will therefore use Cash-for-Work and Cash-for-Women’s work as instruments for household income and women’s earned income. We further standardize household income and women’s earned income by dividing them by mean household income – we can therefore interpret the coefficient on women’s earned income as the impact of shifting all of a typical household’s income from a man to a woman.

We therefore estimate the following IV model in each country c and survey wave t . Letting Y_{hct} be outcome Y for household h in country c in survey wave t (0 for baseline, 1 for midline, and 2 for endline), we estimate

$$Y_{hct} = \beta_{1ct} \text{Women's earned income}_{hc1} + \beta_{2ct} \text{Household income}_{hc1} + X'_{hc} \gamma_{ct}^Y + \epsilon_{hct}^Y \quad (1)$$

$$\begin{aligned} \text{Women's earned income}_{hc1} = & \eta_{1ct}^T \text{Cash-for-Work}_{hc} + \eta_{2ct}^T \text{Cash-for-Women's work}_{hc} \\ & + X'_{hc0} \gamma_{ct}^T + \epsilon_{hct}^T \end{aligned} \quad (2)$$

$$\begin{aligned} \text{Household income}_{hc1} = & \eta_{1ct}^I \text{Cash-for-Work}_{hc} + \eta_{2ct}^I \text{Cash-for-Women's work}_{hc} \\ & + X'_{hc0} \gamma_{ct}^I + \epsilon_{hct}^I \end{aligned} \quad (3)$$

where X_{hc} is a vector of controls which includes the value of the outcome of interest at baseline and any stratifying variables used for randomization. The primary coefficient of interest is β_{1ct} – the estimated impact of shifting all of a household’s income from men to women.

Equations 2 and 3 are our first stage equations: the effect of treatment on women’s earned income and total household income. We expect η_{1ct}^I and η_{2ct}^I to be similar across countries – each treatment will have similar effects on household income. However, we expect $\eta_{1ct}^T \gg \eta_{2ct}^T$ – our first treatment (effectively, a program that increases women’s wages) will increase women’s earned income, while our second treatment (effectively, a program that provides unconditional cash to the household, or a program that increases men’s wages) will

have limited effects on women's earned income.

For inference, we will control false discovery rate across outcomes, using randomization inference following [Anderson \(2008\)](#). Our primary outcomes of interest are presented in [Table 6](#). Following [Banerjee et al. \(2015\)](#), for each outcome we will present average coefficients across countries using inverse variance weights, and report F-tests for equality of coefficients across countries, controlling false discovery rate across these tests. For some outcomes, these are placebo outcomes on which we do not expect impacts. This would reduce our power on other outcomes in a naive multiple inference correction. We have noted these outcomes as a 0 (instead of an X), but do not yet have a plan to improve upon a naive multiple inference correction. We have 12 placebo outcomes and 52 outcomes of interest.

Table 6: Pre-specified outcomes

	Women's earned income		Household income	
	Midline	Endline	Midline	Endline
Predicted household consumption	0	0	X	X
Predicted women's income	X	X	0	0
Time spent				
outside the home (men)	0	0	X	X
outside the home (women)	X	X	0	0
working (self-employment, men)	0	0	X	X
working (self-employment, women)	X	X	0	0
working (salary, men)	0	0	X	X
working (salary, women)	X	X	0	0
working (chores, men)	X	X	X	X
working (chores, women)	X	X	X	X
Women's agency over				
consumption	X	X		
women's time use	X	X		
Women's attitudes towards				
women's time use	X	X		
women's agency	X	X		
Men's attitudes towards				
women's time use	X	X		
women's agency	X	X		
Women's perceptions of norms				
Women's time use	X	X		
Women's agency	X	X		
Attitudes towards women's time use	X	X		
Attitudes towards women's agency	X	X		
Men's perceptions of norms				
Women's time use	X	X		
Women's agency	X	X		
Women's preferences				
over consumption	X	X		
over time use	X	X		
Locus of control	X	X	X	X
Subjective well being	X	X	X	X
Intimate partner violence	X	X	X	X

Lastly, we will conduct an analysis of heterogeneity. We estimate

$$\begin{aligned}
Y_{hct} = & \beta_{1ct} \text{Women's earned income}_{hc1} + \beta_{2ct} \text{Household income}_{hc1} \\
& + \text{Women's earned income}_{hc1} W'_{hc} \beta_{3t} + \text{Household income}_{hc1} W'_{hc} \beta_{4t} \\
& + W'_{hc} \gamma_{1ct}^Y + X'_{hc} \gamma_{2ct}^Y + \epsilon_{hct}^Y
\end{aligned} \tag{4}$$

where W'_{hc} is a vector of household characteristics for which heterogeneity is of interest. First stages for Women's earned income_{hc1}, Household income_{hc2}, Women's earned income_{hc1} W_{hc} , and Household income_{hc2} W_{hc} are analogous to Equations 2 and 3, with Cash-for-Work_{hc} W_{hc} and Cash-for-Women's work_{hc} W_{hc} also included as instruments. We report β_{3t} , focusing on heterogeneity of impacts of Women's earned income, as above controlling false discovery rate across these tests. We also report F-tests for the joint significance of β_{3t} and for the equality of β_{1ct} across countries, controlling false discovery rate across these tests. Building on [Wilke & Humphreys \(2020\)](#), we interpret the latter as a test of the null of external validity, as under this null the household characteristics W_{hc} are sufficient to explain heterogeneity in impacts of Women's earned income across contexts.

We include 3 household characteristics in W_{hc} for our analysis of heterogeneity. First, we include a dummy that the female respondent was previously engaged in salaried work. We interpret larger impacts when female respondents were previously engaged in salaried work as reflecting the relative importance of the intensive margin (increased earned income conditional on any earned income) relative to the extensive margin (any earned income). Second, we include baseline women's agency over time use.⁸ We interpret larger impacts when female respondents have higher agency over time use as suggestive that increasing women's income has larger impacts when women are able to make decisions over their own time use. Third, we include the age gap between the woman and the primary male decision maker in the household, that is the primary male decision maker's age minus the female

⁸In order to compare across contexts, when using women's agency over women's time use to test for heterogeneity, we use the same weights for all countries to construct the index, calculating weights for the index using the control group across all countries.

respondent's age. We interpret larger impacts when the age gap is smaller as suggestive that interventions directly targeting broad increases in women's empowerment would complement interventions to increase women's earned income.

4.1.1 Discussion of potential exclusion restriction violations

In the above model, we implicitly assume that differences in the impacts of Cash-for-Work and Cash-for-Women's work on all outcomes are, relative to the control group, explained by differences in their impacts on women's earned income and household income. Alternatively, differences could be explained by impacts on women's income (earned or unearned). Existing literature studying the impacts of shifting unconditional cash transfers from men to women has found mixed results: [Haushofer & Shapiro \(2016\)](#) find intimate partner violence falls, but no impacts on investment, while [Akresh et al. \(2013\)](#) find differences in how men and women invest unearned income. In countries where it was feasible, we include an unconditional cash transfer arm. As an unconditional cash transfer shifts unearned income, but not earned income, this allows us to test whether our effects are explained by changes in women's unearned income in those countries.

In addition, there are other likely deviations from the model in which the only difference between Cash-for-Work and Cash-for-Women's work is the gender of the participant. Although this ideally would be held fixed, gender by its nature implies differences in perceived roles, which leads to differences in the activities undertaken by men and women. Therefore, activities in Cash-for-Work and Cash-for-Women's work may generate different amounts of physical and human capital, causing differential impacts across the two arms on medium run income even when impacts on short run income are the same. As a result, medium run impacts may be driven purely by income effects. However, we can use the short run income effects that we estimate to isolate the component of medium run effects that are not explained by income effects.

Finally, WFP's FFA programs often include additional components, such as social and

behavior change communication programs. Participation in these programs may itself have impacts distinct from the impacts of work, and potentially directly shift women’s autonomy. We are not able to separate the impacts of these additional programs from the work component, in contrast to other work that has focused on estimating the impacts of these programs with and without cash (Roy et al., 2019). The intensity of these additional programs and their characteristics vary country-to-country; if we find similar impacts across contexts on key outcomes, we can interpret that as evidence that these additional programs do not drive our results. However, broadly any work has idiosyncratic characteristics that impact laborers, so these qualifications will apply to any research design shifting women’s earnings holding fixed household income.

4.2 Power calculations

For power calculations, we estimate statistical power for the reduced form

$$Y_{hc1} = \alpha + \delta_{1c}\text{Treated}_{hc} + \delta_{2c}\text{Cash-for-Women’s work}_{hc} + \epsilon_{hc1}^{RF}$$

where for convenience we ignore the presence of controls (yielding modestly conservative power calculations), and Treated_{hc} denoted that household h in country c received either the Cash-for-Work or Cash-for-Women’s work treatment. We focus on power for δ_{2c} , the effect of Cash-for-Women’s work conditional on Treated at midline. For analysis of the impacts of household income, we also estimate statistical power for the reduced form

$$Y_{hc1} = \alpha + \delta_{1c}\text{Treated}_{hc} + \epsilon_{hc1}^{RF}$$

where we pool across both Cash-for-Work and Cash-for-Women’s work. Calculated minimum detectable effects and expected effect sizes are presented in Table 7, along with country and estimating equation specific assumptions. All other assumptions and the details of the calculations are discussed below.

Table 7: Power calculations

	Country 1	Country 2	Country 3
Treated			
Number of observations			1800
Number of clusters			1800
Share treated			0.67
MDE			0.065
Anticipated take-up	0.9		0.8
Transfer size	≈ 0.15		0.3
Expected effect	≈ 0.091		0.161
Cash-for-Women's work			
Number of observations			1200
Number of clusters			1200
Share treated			0.5
MDE			0.225
Anticipated take-up	0.9		0.4
Transfer size	≈ 0.15		0.3
Expected effect	≈ 0.091		0.080

For the first power calculation, we use Predicted women's income as an outcome, as it can be calculated in any household survey. For the second power calculation, we use Predicted household consumption. We use the 2015/16 Kenya Integrated Household Budget Survey for these calculations, restricting to rural poor households, consistent with the typical households targeted by WFP CFA programs. We apply the standard formula for the minimum detectable effect, $MDE = \sigma_\epsilon (z_{0.8} + z_{0.975}) \sqrt{\frac{1+\rho(m-1)}{NP(1-P)}}$, where σ_ϵ is the standard deviation of the outcome, $z_{0.8} + z_{0.975} = 2.80$ is the sum of the two z-scores, ρ is the intracluster correlation, m is the number of observations per cluster, N is the number of observations, and P is the share of observations assigned to treatment. We set $\rho = 0.05$ for all calculations. N and P will both vary across the two reduced forms, as power for the effect of Cash-for-Women's work conditional on being treated depends on the number of treated households, and the share of treated households who receive Cash-for-Women's work.

To calculate σ_ϵ for Predicted household consumption, we first select via LASSO the 5

goods that best predict household consumption, controlling for village fixed effects and number of women, men, and children under the ages of 2, 5, 10, and 16 in the household. We assume Predicted household consumption is a surrogate for household consumption in the language of [Athey et al. \(2016\)](#). We derive power under their worst case bounds when surrogacy is violated: doing so is equivalent to scaling σ_ϵ by $1/R^2$, where R^2 is from a regression of residualized Predicted household consumption on residualized household consumption.⁹ To construct a single measure we can use across contexts, we normalize by average household consumption. Lastly, we replicate this exercise for Predicted women’s income by assuming it is a surrogate for women’s income, and we also include controls for total household consumption and total household income. This calculation yields $\sigma_\epsilon = 0.46$ for Predicted household consumption and $\sigma_\epsilon = 1.39$ for Predicted women’s income.

To calculate our expected effect size for each analysis, we focus on effects during the midline survey. For household consumption as an outcome of pooled treatment, we first apply a marginal propensity to consume from cash transfers of 0.67, estimated based on [Haushofer & Shapiro \(2016\)](#). We then multiply this by the share of households anticipated to take up the intervention, and the monthly transfer size relative to average monthly household consumption. For women’s income as an outcome of Cash-for-Women’s work conditional on being treated, we continue to apply a marginal propensity to consume of 0.67.¹⁰ We then multiply this by take-up, which is now the share of participating households who shift from male to female participants in response to Cash-for-Women’s work,¹¹ and the monthly transfer size relative to average monthly household consumption.

⁹This R^2 is biased upward as we do not do any cross validation to correct for selecting the goods in the same sample that we estimate the R^2 , which causes us to overestimate power.

¹⁰Although this is an income measure, it is predicted from consumption goods in a cross section where marginal propensity to consume is likely closer to 1.

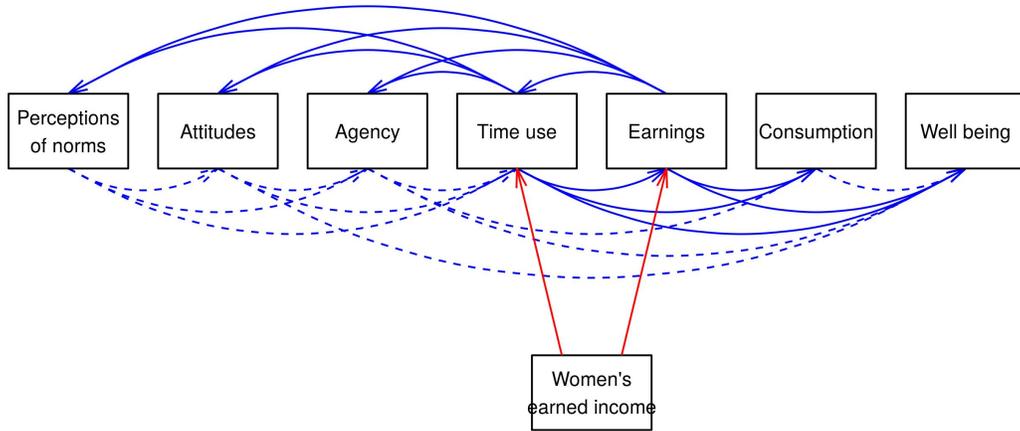
¹¹Formally, this is the share of households with female participants minus the share of households with male participants in Cash-for-Women’s work, minus the share of households with female participants minus the share of households with male participants in Cash-for-Work, all divided by 2.

4.3 Results

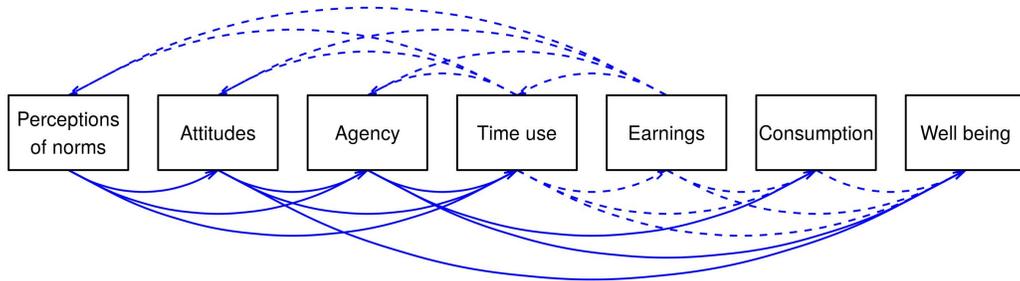
In Figure 1, we lay out a theory of change for the intervention based on the context to guide our discussion of the results, with arrows communicating directions of causality. We use this theory of change to guide our analysis described in Section 4.1, including both our primary analysis and analysis of heterogeneity. We focus on the impacts of women’s earned income (which we refer to as the impacts of Targeting Women) estimated in Equation 1, as the impacts of household income are well established. Figure 1a shows medium run effects of Targeting Women (corresponding to outcomes in the midline survey), Figure 1b shows long run effects of Targeting Women (corresponding to outcomes in the endline survey), while Figure 1c shows which outcomes in Table 6 correspond to which nodes of the theory of change.

Figure 1: Theory of change

(a) Theory of change: Medium run



(b) Theory of change: Long run



(c) Outcomes

Perceptions of norms	Attitudes	Agency
Women's PoN, women's time use	Women's attitudes, women's time use	Consumption
Women's PoN, women's agency	Women's attitudes, women's agency	Women's time use
Women's PoN, attitudes, women's time use	Men's attitudes, women's time use	
Women's PoN, attitudes, women's agency	Men's attitudes, women's agency	
Men's PoN, women's time use		
Men's PoN, women's agency		

Time use	Earnings	Consumption	Well being
Women, outside the home	Women's	HH, predicted	Locus of control
Women, self-employment	Men's	Women's earnings, predicted	Subjective well being
Women, salaried			IPV
Women, chores			Women's preferences, consumption
Men, outside the home			Women's preferences, time use
Men, self-employment			
Men, salaried			
Men, chores			

First, in Figure 1a, impacts of Targeting Women enter through the household's earnings and time use. Solid lines trace out the direct impacts of these changes in the theory of

change, while dotted lines trace out secondary impacts. To see how this theory of change guides the discussion, suppose we saw that Targeting Women shifted only agency, time use, earnings, and consumption. We would conclude that impacts on agency were caused by changes in women’s earned income and time use. However, impacts on consumption would be interpreted as driven by some combination of direct effects of changes in women’s earned income and time use, and indirect effects through changes in agency. Alternatively, suppose we saw that Targeting Women shifted only attitudes, time use, earnings, and consumption. We would then conclude that our evidence is consistent with impacts on consumption caused only by changes in women’s earned income, as we failed to find evidence of shifts in any other upstream outcomes that might cause changes in consumption.

Second, in Figure 1b, we assume that any long run impacts of Targeting Women would occur through long run changes in perceptions of norms, attitudes, and agency. Now, suppose we saw long run changes in attitudes and time use. We would conclude that the impacts on time use were driven by persistent changes in attitudes, as opposed to changes in perceptions of norms or agency. Alternatively, suppose we saw long run changes in attitudes, agency, and time use. We would conclude that changes in time use were driven by either persistent changes in attitudes or agency (or both), but changes in either attitudes or agency could also be caused contemporaneously by the changes in time use.

Lastly, in Figure 1c, each node of the theory of change is tied to multiple outcomes. For Perceptions of Norms, Attitudes, and Agency, we do not expect changes in every outcomes to plausibly affect every downstream outcome. To enumerate these:

1. “Perceptions of Norms” would only affect attitudes and the associated activity. For instance, “Perceptions of Norms, Attitudes, Women’s Agency” and “Perceptions of Norms, Women’s Agency” would only affect “Attitudes, Women’s Agency” and “Agency” directly.
2. “Attitudes” would only affect well being and the associated activity directly. For instance, “Attitudes, Women’s Time Use” would only affect “Well being” and “Time

use” directly.

3. “Agency over Consumption” would only affect “Well being” and “Women’s earnings, predicted” directly, while “Agency over Time Use” would only affect “Well being” and “Time Use” directly.

References

- Akresh, R., De Walque, D., & Kazianga, H. (2013). Cash transfers and child schooling: evidence from a randomized evaluation of the role of conditionality.
- Alesina, A., Giuliano, P., & Nunn, N. (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, 128(2), 469–530.
- Almås, I., Armand, A., Attanasio, O., & Carneiro, P. (2018). Measuring and changing control: Women’s empowerment and targeted transfers. *The Economic Journal*, 128(612), F609–F639.
- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484), 1481–1495.
- Armand, A., Attanasio, O., Carneiro, P. M., & Lechene, V. (2016). The effect of gender-targeted conditional cash transfers on household expenditures: Evidence from a randomized experiment.
- Ashraf, N. (2009). Spousal control and intra-household decision making: An experimental study in the philippines. *American Economic Review*, 99(4), 1245–77.
- Athey, S., Chetty, R., Imbens, G., & Kang, H. (2016). Estimating treatment effects using multiple surrogates: The role of the surrogate score and the surrogate index. *arXiv preprint arXiv:1603.09326*.
- Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., & Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236), 1260799.

- Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: does exposure reduce bias? *The Quarterly journal of economics*, 124(4), 1497–1540.
- Benhassine, N., Devoto, F., Duflo, E., Dupas, P., & Pouliquen, V. (2015). Turning a shove into a nudge? a “labeled cash transfer” for education. *American Economic Journal: Economic Policy*, 7(3), 86–125.
- Bertrand, M., Kamenica, E., & Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2), 571–614.
- Browning, M. & Chiappori, P.-A. (1998). Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, (pp. 1241–1278).
- Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2018). Misperceived social norms: Female labor force participation in saudi arabia. (24736).
- Dhar, D., Jain, T., & Jayachandran, S. (2018). *Reshaping Adolescents’ Gender Attitudes: Evidence from a School-Based Experiment in India*. Working Paper 25331, National Bureau of Economic Research.
- Duflo, E. (2000). *Grandmothers and Granddaughters: Old Age Pension and Intra-household Allocation in South Africa*. Working Paper 8061, National Bureau of Economic Research.
- Field, E. M., Pande, R., Rigol, N., Schaner, S. G., & Moore, C. T. (2019). *On Her Own Account: How Strengthening Women’s Financial Control Affects Labor Supply and Gender Norms*. Working Paper 26294, National Bureau of Economic Research.
- Folke, O. & Rickne, J. (2020). All the single ladies: Job promotions and the durability of marriage. *American Economic Journal: Applied Economics*, 12(1), 260–87.
- Goldin, C. & Olivetti, C. (2013). Shocking labor supply: A reassessment of the role of world war ii on women’s labor supply. *American Economic Review*, 103(3), 257–62.

- Haushofer, J., Ringdal, C., Shapiro, J. P., & Wang, X. Y. (2019). *Income Changes and Intimate Partner Violence: Evidence from Unconditional Cash Transfers in Kenya*. Working Paper 25627, National Bureau of Economic Research.
- Haushofer, J. & Shapiro, J. (2016). The short-term impact of unconditional cash transfers to the poor: Experimental evidence from kenya. *The Quarterly Journal of Economics*, 131(4), 1973–2042.
- Hochschild, A. & Machung, A. (2012). *The second shift: Working families and the revolution at home*. Penguin.
- Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics*, 7(1), 63–88.
- Jayachandran, S. (2019). Social norms as a barrier to women’s employment in developing countries.
- Lundberg, S. & Pollak, R. A. (1993). Separate spheres bargaining and the marriage market. *Journal of political Economy*, 101(6), 988–1010.
- Lundberg, S. J., Pollak, R. A., & Wales, T. J. (1997). Do husbands and wives pool their resources? evidence from the united kingdom child benefit. *Journal of Human resources*, (pp. 463–480).
- McKelway, M. (2019). Vicious and virtuous cycles: Self-efficacy and employment of women in india.
- Roy, S., Hidrobo, M., Hoddinott, J., & Ahmed, A. (2019). Transfers, behavior change communication, and intimate partner violence: Postprogram evidence from rural bangladesh. *Review of Economics and Statistics*, 101(5), 865–877.
- Thomas, D. (1990). Intra-household resource allocation: An inferential approach. *Journal of human resources*, (pp. 635–664).

Wilke, A. M. & Humphreys, M. (2020). Field experiments, theory, and external validity.
In L. Curini & R. Franzese (Eds.), *The SAGE Handbook of Research Methods in Political Science and International Relations*. Oxford University Press.