Can Public Works Increase Women's Autonomy? Experimental Evidence from Four Countries*

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

March 23, 2021

Abstract

Can women's participation in public works contribute to closing the gender gap in autonomy? We randomize participation in and participant gender for World Food Programme cash-for-work programs that target poor households in four 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.

JEL Classification Codes: O15, J16, J22, I38, D13

Keywords: Women's empowerment, Public works, Intrahousehold decision making

*This design benefited from comments from Emma Frankham, Macartan Humphreys, Sarah Khan, Jeremy Magruder, and two anonymous reviewers. This research benefits from generous funding from the Office of Evaluation of the World Food Programme. We thank Saahil Karpe, Felipe Dunsch, and Kristen McCollum for superb research coordination. Rasha Absi, Sylvia Cesar, Marc Andrea Fiorina, Mariachiara Iannuzzi, Eric Jospe, Edwin Kasila, and Miguel Melendez provided excellent research assistance. Finally, we thank the management and technical staff at WFP Country Offices for being outstanding research partners. This study was approved by Solutions IRB (2020/07/27), and was registered in the AEA RCT Registry (AEARCTR-0005933) prior to collection of outcome data. The views expressed do not reflect the views of the World Bank nor the World Food Programme. All errors are our own. At time of publication of this pre-analysis plan, four of an anticipated six countries have been selected (El Salvador, Kenya, Rwanda, and Syria). Selection of the two remaining countries is ongoing. In the event that we are only able to implement the experimental design described in this pre-analysis plan in these four countries, we will proceed with the analysis as described. 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

Can equalizing opportunities for men and women to work outside the home 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 labor market opportunities, and gender equality in autonomy are all strongly linked, but causality is unclear. Furthermore, institutions ("*Women cannot drive*") and norms and attitudes ("*Women should not work outside the household*") are root causes of gender inequality in opportunities to work outside the home, 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 access to labor markets, estimating their impacts on gender equality in autonomy, and observing how these changes influence institutions, attitudes, and norms.

We estimate the impacts of shifting men's and women's participation in public works on household decision making and the autonomy gender gap. While public works programs have been demonstrated to be an effective tool for economic development through increased earnings (Imbert & Papp, 2015; Gazeaud et al., 2019; Adjognon et al., 2020), there is less evidence on the impacts of participant gender in public works. To fill this gap, we work with World Food Programme (WFP) across six countries to implement two experimental arms. First, WFP's Food Assistance-for-Assets (FFA) program ("Cash-for-Work"), which has public works as a core component. Second, a modified version of FFA which 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 opportunities to work (Cash-for-Women's Work relative to Cash-for-Work) on household decision making.

The FFA programs we study provide households with a cash transfer conditional on a specific household member participating in public work.[1] These programs target particularly

[1]In many contexts, including some of the countries our experiments take place in, these public works

vulnerable communities, and focus on poor households within these communities with the objective of promoting resilience. 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.

We anticipate that manipulating men's and women's participation in public works may impact gender gaps in autonomy and labor force participation through two primary channels. First, we anticipate a direct "wage effect" for the duration of public works — household members make labor supply decisions by trading off household consumption gains with the opportunity costs of work outside the home, including foregone leisure and home production. A striking observation is women working for a wage often substitute away from leisure which creates a "second shift" (Hochschild & Machung, 2012), while men do not shift into home production (Bertrand et al., 2015). In a unitary household model, 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 being that men and women have agency over "separate spheres" of household decisions (Lundberg & Pollak, 1993).

Second, we anticipate a persistent "empowerment effect" — women's temporary participation in public works may result in lasting shifts in labor market attachment through changes in intrahousehold agency, attitudes, and perceptions of norms. Recent experimental work has demonstrated 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 participation in public works may also affect household decision making through these channels. These impacts may cascade, leading to persistent shifts in women's labor supply in response to temporary women's labor demand

include work supporting the creation of quasi-public or private assets, such as pens for livestock and kitchen gardens. We elaborate on country-specific details of implementation in Section 2.2.

shocks (Alesina et al., 2013; Goldin & Olivetti, 2013).

We implement our analysis of the impacts of men's and women's participation in public works in six steps. First, we estimate the short run impacts of men's and women's participation in public works on their labor supply. We begin by estimating the first stage impacts of Cash-for-Work and Cash-for-Women's Work on household and women's participation in public works. We anticipate Cash-for-Work primarily increases household participation in public works, and "Targeting Women" (Cash-for-Women's Work relative to Cash-for-Work) increases women's participation in public works without shifting household participation in public works. As such, these increases in women's participation in public works accompany decreases in men's participation in public works, and we interpret the effects of Targeting Women as the combined impacts of these shifts.

Second, we estimate the short run effects of men's and women's participation in public works on men's and women's time use. We anticipate the effects of men's participation in public works will depend on contextual features of the program. We expect women's participation in public works will increase time women spend working outside the home, while causing women to substitute away from leisure, sleep, and home production. We also anticipate the relative magnitudes of these effects will depend on baseline women's agency, norms, attitudes, and household structure.

Third, we estimate the short run impacts of women's participation in public works on household decision making and household welfare. We collect and leverage rich data on women's agency, attitudes, perceptions of norms, household consumption, intimate partner violence (IPV), and women's subjective well being. We do not have particular expectations for how these outcomes might change.

Fourth, we estimate the medium run impacts of women's participation in public works on household decision making. We anticipate that the short run impacts on household decision making we document in the previous paragraph may persist even after household participation in the public works programs has ended.

Fifth, we estimate the medium run impacts of these anticipated changes in women's autonomy on women's labor supply after the public works programs end on men's and women's time use and household welfare. As the program will have ended, there will no longer be differences in access to public works between women in the Cash-for-Work and women in the Cash-for-Women's Work arms. We therefore interpret differences in time use as attributed to shifts in women's labor supply, and differences in household welfare as attributed to shifts in women's autonomy and women's labor supply. We assume these medium run shifts in women's labor supply are caused by estimated impacts on women's labor market history and occupational choice, agency, attitudes, and perceptions of norms.

Lastly, we test the external validity of our results across the six countries where we implement these interventions. We implement two tests of external validity. First, we follow the cross-country analysis of Banerjee et al. (2015) and test the hypothesis of homogeneous treatment effects across countries. Second, we allow treatment effects to be heterogeneous with respect to observable characteristics of individuals that are likely to mediate impacts of participation in public works, and again test the hypothesis of homogeneous treatment effects across countries. While exploring these dimensions of heterogeneity might itself be interesting in unpacking underlying mechanisms, it also permits a test for whether allowing for heterogeneity in these observable characteristics is sufficient to achieve external validity. As observables, we include women's baseline labor force participation, age gaps between female and male decision makers in the household,[2] and the presence of other women in the household.

Our experimental design is closest to studies that have experimentally varied recipients by gender within households in cash transfer programs (Akresh et al., 2013; Benhassine et al., 2015; Armand et al., 2016; Haushofer & Shapiro, 2016).[3] In contrast to cross-sectional

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

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

(Thomas, 1990) and quasi-experimental evidence (Lundberg et al., 1997; Duflo, 2000) leveraging variation in "unearned income," experimental work has typically found limited impacts of shifting recipients of cash transfers 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 instead experimentally vary participants by gender within households in public works programs in order to provide complementary causal evidence on the impacts of women's share of earned income.

We build on a deep literature at the nexus of economic development, gender equality, and women's autonomy. In particular, our research questions are closest to 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 builds most closely on Field et al. (forthcoming), who estimate the impacts of payment modality in public works for women participants; in contrast, we estimate the impacts of women's participation in public works conditional on household participation. Our theory of change also closely follows Field et al. (forthcoming), who find increased control over earnings causes women to increase their labor supply, and these impacts persist with changes in attitudes toward women working as a mechanism. Relatedly, McKelway (2019) finds success in applying for jobs increases generalized self–efficacy, and increased generalized self–efficacy in turn increases women's labor supply. Relative to McKelway (2019), 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 also expand the external validity of our findings by providing evidence from six countries and revealing 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, reviewed in Jayachandran (2019), 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 that temporary shifts to women's labor demand also shift women's agency, attitudes, and norms, generating sustained decreases in gender gaps in both labor force participation and autonomy.

The rest of this pre-analysis plan 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 and contexts

2.1 Across countries

The program of interest in all countries is a version of WFP's Food Assistance-for-Assets (FFA) program. FFA functions similarly to cash-for-work programs, in that households are paid a cash transfer conditional on participating in household level or community work.[4] Exact details vary by context, and are described in Section 2.2. The programs are described as "Food Assistance-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 FFA, with additional arms to adjust for household or community level impacts of the FFA program. A generic version

[4]The programs are called "Food Assistance-for-Assets" instead of "Cash-for-Assets" to highlight the objective of the programs to increase food security.

of this experimental design is presented in Table 1. 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 FFA program, and the eligible man is a primary decision maker in the household.[5]

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").

Optional. Each country, where possible, includes an unconditional cash transfer (UCT) arm, with transfer recipient randomized where possible.

Table 1: Template Experimental Design

Control	# of Obs. (# of Clusters)
Cash-for-Work	# of Obs. (# of Clusters)
Cash-for-Women's work	# of Obs. (# of Clusters)

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 participants. 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,

[5]When feasible, both the man and the woman should be interested in participation, as this will improve power as demonstrated in Section 4.2.

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 household income fixed 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.

The generic timeline of implementation and surveys is presented in Table 2; additional details on the survey are discussed in Section 3. Baseline surveys will take place just prior to the start of the intervention. The WFP programs in Rwanda and El Salvador both feature 3 months of transfers, with the work component in El Salvador continuing for up to 3 months after the completion of transfers. A midline survey will take place during the implementation of the cash transfers (that is, after the first transfer has been made, but before the last transfer has been made, and while the work component is ongoing), and its reference period (one month) will lie entirely during the period during which cash transfers are being made and the work component is ongoing.[7] An endline survey will occur sufficiently after the end of the intervention such that the reference period for the endline survey (one month) will exclude the period of the intervention.[8]

[6]In theses 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.

[7]This is necessary such that all questions during the midline, particularly time use and income, can be used to estimate the direct impacts of Cash-for-Work and Cash-for-Women's work. We discuss this further in Section 4.

[8]This is necessary such that all questions during the endline can be used to estimate the persistent indirect impacts of Cash-for-Work and Cash-for-Women's work. We discuss this further in Section 4.

Table 2: Template Timeline

	Time	Survey Reference Period	Notes
Registration	Month 0		
Baseline	Month 1	Month -5 - 0	Pre-intervention, post-sampling
Intervention start (Work)	Month 2		
Intervention start (Transfers)	Month 3		
Midline	Month 4	Month 3	After transfers start, while work ongoing
Intervention end (Transfers)	Month 5		
Intervention end (Work)	Month 5		
Endline	Month 8	Month 7	Reference period after work and transfers

Notes: The template timeline of key events associated with implementation and surveying are presented in this table. Events occur at the start of each month in this stylized timeline.

2.2 Countries

At this point, four of the six countries (from three continents) that will be participating in the evaluations have been selected — El Salvador, Kenya, Rwanda, and Syria. In Section 2.2.1, we provide for each country a table with the experimental design, including number of observations in each cell, with number of clusters in parentheses. In Section 2.2.2, we then present additional details on sample selection and program implementation in each country, and in Section 2.2.3 we present timelines for each country. Kenya and Syria are currently under preparation, so many details for these countries are currently under discussion.

2.2.1 Experimental designs

El Salvador In El Salvador, the Food Assistance-for-Assets (FFA (Female)) program includes a cash transfer conditional on participation in a community works program. All participants are female. There is also an unconditional cash transfer (UCT (Male)) program in which unconditional cash transfers are made to households; households select the recipient, and in practice select almost exclusively male recipients. Random assignment will occur at the site level (with each community corresponding to a single site).

Table 3: Experimental Design: El Salvador

Control	500 (25)
UCT (Male)	500 (25)
FFA (Female)	500 (25)

For our analysis, UCT (Male) is the "Cash-for-Work" arm, and FFA (Female) is the "Cash-for-Women's work" arm. UCT (Male) does not have a work component, and implications of this for the analysis are discussed in Section 4.

Kenya In Kenya, the Food Assistance-for-Assets (FFA) program includes an unconditional cash transfer and an asset transfer, with promotion of the livelihood tied to that asset. The FFA (Female) program provides chickens (which only women traditionally raise in that region), and the FFA (Male) program provides pasture seeds (which only men traditionally raise in that region). The UCT arm will provide only the cash. Random assignment will occur at the site level (with each community corresponding to a single site).

Table 4: Experimental Design: Kenya

Control	To Be Determined
FFA (Male)	To Be Determined
FFA (Female)	To Be Determined

FFA (Male) is the "Cash-for-Work" arm, and FFA (Female) is the "Cash-for-Women's work" arm. UCT is not included in the primary analysis, but will be used for robustness to validate that estimates are robust to either using FFA (Male) or UCT as the "Cash-for-Work" arm.

Rwanda In Rwanda, the Food Assistance-for-Assets (FFA) program includes a cash transfer conditional on participating in a livelihoods program. FFA (Female) restricts to female participants, while FFA (Mixed) does not make restrictions based on gender. Random assignment will occur at the community level, with multiple communities per site, and randomization is stratified at the site level.

Table 5: Experimental Design: Rwanda

Control	387 (26)
FFA (Mixed)	387 (26)
FFA (Female)	399 (26)

FFA (Mixed) is the "Cash-for-Work" arm, and FFA (Female) is the "Cash-for-Women's work" arm.

Syria In Syria, the Food Assistance-for-Assets (FFA) program includes a cash transfer conditional on participating in a livelihoods program, with the exact activities to be determined through qualitative work to be implemented by a development partner. FFA (Female) and FFA (Male) restrict to male and female participants, respectively. Transfers are always made to the participant. Random assignment will occur at the household level.

Table 6: Experimental Design: Syria

	E-Voucher	Unrestricted cash
Control	To Be Determined	
FFA (Female)	To Be Determined	To Be Determined
FFA (Male)	To Be Determined	To Be Determined

FFA (Male) is the "Cash-for-Work" arm, and FFA (Female) is the "Cash-for-Women's work" arm. We do not plan to use variation in modality in the pre-specified analysis in this paper.

2.2.2 Program details and sampling

We present key details of the programs across these four countries in Table 7, including community and household selection (for both the program and survey samples) and details of implementation of the Cash-for-Work and Cash-for-Women's work arms. Households and communities were selected for the program based on level vulnerability. Transfers tied to FFA participation are typically between 100 and 200 USD PPP per month, and last for between 2 and 3 months.

Table 7: Implementation summary

	El Salvador	Kenya	Rwanda	Syria
Overall				
Level of randomization	Community/Site	Community/Site	Community	Household
Program eligibility requirements (Community)	5-6 highly vulnerable communities per municipality in highly vulnerable rural municipalities	To Be Determined	Highly vulnerable rural Sectors in 5 Districts, walking distance of potential public works site	To Be Determined
Additional sample requirements (Community)	No	To Be Determined	Not the location of site	To Be Determined
Program eligibility requirements (Household)	Doesn't expect WFP Transfer, affected by COVID or Tropical Storm Amanda, both female and male household head	To Be Determined	Ubudehe 1 or 2, not currently participating in government public works	To Be Determined
Additional sample requirements (Household)	Both a man and a woman in household registered as able to participate	To Be Determined	Both a man and a woman in household registered as able to participate	To Be Determined
Female and male respondents	Yes	To Be Determined	Yes	To Be Determined
Details of Cash-for-Work intervention				
Type of intervention	UCT (Male)	FFA (Male)	FFA (Mixed)	FFA (Male)
Cash transfer value/frequency	150 USD/1.5 months	To Be Determined	30 USD/1 month	To Be Determined
Total cash transfer value	300 USD	To Be Determined	60-90 USD	To Be Determined
Cash transfer duration	3 months	To Be Determined	2-3 months	To Be Determined
Cash transfer modality	Cash	To Be Determined	Cash or mobile money	To Be Determined
Work requirement	No	Yes	Yes	Yes
Share women participants	Not Applicable	To Be Determined	Realized post-implementation	To Be Determined
Asset transfer	No	To Be Determined	Fertilizer, seedlings, livestock	To Be Determined
Asset value	0 USD	To Be Determined	Realized post-implementation	To Be Determined
Most common activity (% of HH/% of sites)	Not Applicable	To Be Determined	Land terracing (Realized post-implementation)	To Be Determined
2nd most common activity (% of HH/% of sites)	Not Applicable	To Be Determined	Marshland reclamation (Realized post-implementation)	To Be Determined
Details of Cash-for-Women's work intervention				
Type of intervention	FFA (Female)	FFA (Female)	FFA (Female)	FFA (Female)
Cash transfer value/frequency	150 USD/1.5 months	To Be Determined	30 USD/month	To Be Determined
Total cash transfer value	300 USD	To Be Determined	60-90 USD	To Be Determined
Cash transfer duration	3 months	To Be Determined	2-3 months	To Be Determined
Cash transfer modality	Cash	To Be Determined	Cash or mobile money	To Be Determined
Work requirement	Yes	Yes	Yes	Yes
Share women participants	Realized post-implementation	To Be Determined	Realized post-implementation	To Be Determined
Asset transfer	To Be Determined	To Be Determined	Fertilizer, seedlings, livestock	To Be Determined
Asset value	To Be Determined	To Be Determined	Realized post-implementation	To Be Determined
Most common activity (% of HH/% of sites)	To Be Determined	To Be Determined	Land terracing (Realized post-implementation)	To Be Determined
2nd most common activity (% of HH/% of sites)	To Be Determined	To Be Determined	Marshland reclamation (Realized post-implementation)	To Be Determined

2.2.3 Timeline (Proposed timeline for completion of the study)

The timeline of implementation and surveys for each country is presented in Table 8, and follows the generic timeline that is described in Section 2.1.

Table 8: Timeline

	El Salvador	Kenya	Rwanda	Syria
Registration	2021/02		2020/11	
Baseline	2021/02		2020/12	
Intervention start (Work)	2021/04		2021/02	
Intervention start (Transfers)	2021/04		2021/03	
Midline	2021/05		2021/03	
Intervention end (Transfers)	2021/06		2021/06	
Intervention end (Work)	2021/11		2021/06	
Endline	2021/12		2022/02	

Notes: The timeline of key events associated with implementation and surveying are presented in this table. Dates reported are the month during which each event began.

The timelines for Kenya and Syria will be determined by when implementation begins of the FFA programs that we described for those countries in Section 2.2.1. We anticipate that implementation, starting with registration, will begin in 2021. The duration from registration to endline will be similar in both Kenya and Syria to El Salvador and Rwanda; therefore, if registration occured in December 2021, the endline survey would be completed by the end of March 2023.

3 Data

3.1 Survey

In all countries, a woman selected as eligible to participate in the FFA program is the primary respondent for the survey. Recall that sampled households have both a man and a woman who would be eligible to participate in the FFA program. Many questions in the survey will be asked about both this woman and this man. 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. As some of the questions are on sensitive topics that are strongly tied to gender, we use only female enumerators where possible. All questions at midline and endline restrict reference periods to a maximum of one month unless otherwise specified.

All standardized indices below are constructed using inverse covariance weighting following Anderson (2008). Questions that feed into outcomes indices for agency, attitudes, and perceptions of norms will be coded so higher values correspond to women spending more time in self-employment and salaried work and less time on household chores or so higher values correspond to additional decision making authority for women. For all other questions higher levels are coded naturally (consumption, earnings, time spent on specific activities or at specific locations, locus of control, subjective well being, intimate partner violence).

3.1.1 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.[9] Expenditures on education, expenditures on men's clothing, and expenditures on women's clothing will be included in all countries.

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.

[9] Additional details on the selection of goods are in Appendix B.

Women preferred consumption Measured as predicted women's earnings from the household's expenditures, controlling for assigned treatments.

3.1.2 Participation

Questions Households are asked about whether any household members participated in WFP programs or received transfers from WFP programs in the previous 6 months for the baseline survey, the minimum of the previous 6 months and the time since the baseline survey for the midline survey, and the minimum of the previous 6 months and the time since 2 weeks after the intervention was completed for the endline survey. Any participant is defined as having any household member participate or receive transfers, and Woman participant is defined as a female household member participating in the work component. These variables will be used as first stage outcomes in our analysis described in Section 4.1.

Outcomes Any participant, Woman participant.

3.1.3 Earnings and income

Questions Earnings for each household member are collected for the previous 6 months for the baseline survey, the minimum of the previous 6 months and the time since the baseline survey for the midline survey, and the minimum of the previous 6 months and the time since 2 weeks after the intervention was completed for the endline survey. Earnings are measured as total earnings from WFP plus total earnings from other paid permanent and temporary work.

For robustness, results instead measuring earnings by adding earnings from self-employment (constructed as profits from income generating activities and businesses managed by that household member), livestock (constructed as median daily wage times days of labor for laborers, and the maximum of 0 and cash profits net of household labor valued at the median daily wage for managers), and agriculture (constructed as for livestock) will be presented in

an appendix.

Outcomes Women's Earnings, Men's Earnings.

3.1.4 Time use

Questions The female respondent is asked for a 48 hour recall of her activities over the past two days, along with her activities over the most recent country-specific day of rest, following the approach of the 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, and we will use the female respondent's report to construct the outcome for those household-survey rounds. We will aggregate across activities by location (outside the home or inside the home) or by category (working in self-employment, working for a salary, working on chores, leisure and sleep), with WFP FFA work and other public works categorized as work for a salary.

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).

3.1.5 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.

3.1.6 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.

3.1.7 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 she *believes* 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 her *beliefs* 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 perceptions of norms over women's time use and men's perceptions of norms over women's agency are included.

3.1.8 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). At midline and endline, we set the reference period for IPV at two months.

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}.

3.1.9 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 WFP FFA, adjusting for any household level impacts of participation in WFP FFA. 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, earnings, and time use, we will similarly standardize. For consumption and earnings, we will normalize by monthly transfer size, and also report impacts normalized by total amount transferred in the Appendix. 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, varies 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.

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

$$Y_{hct} = \beta_{1,ct}\text{Woman participant}_{hc1} + \beta_{2,ct}\text{Any participant}_{hc1} + X'_{hc0}\gamma^Y_{ct} + \epsilon^Y_{hct} \tag{1}$$

$$\begin{aligned}\text{Woman participant}_{hc1} = \eta^T_{1,ct}\text{Cash-for-Work}_{hc} + \eta^T_{2,ct}\text{Cash-for-Women's work}_{hc} \\ + X'_{hc0}\gamma^T_{ct} + \epsilon^T_{hct}\end{aligned} \tag{2}$$

$$\begin{aligned}\text{Any participant}_{hc1} = \eta^I_{1,ct}\text{Cash-for-Work}_{hc} + \eta^I_{2,ct}\text{Cash-for-Women's work}_{hc} \\ + X'_{hc0}\gamma^I_{ct} + \epsilon^I_{hct}\end{aligned} \tag{3}$$

where X_{hc0} 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 $\beta_{1,ct}$ – the estimated impact of shifting household public works participation from men to women.

Equations 2 and 3 are our first stage equations: the effect of treatment assignment on women's participation and household participation. We expect $\eta^I_{1,ct}$ and $\eta^I_{2,ct}$ to be similar across countries – each assigned treatment will have similar effects on household participation.[10] However, we expect $\eta^T_{2,ct} \gg \eta^T_{1,ct}$ – Cash-for-Women's work will only increase women's participation in public works, while Cash-for-Work will primarily increase men's participation in public works.

For inference, we will use robust variance covariance estimators clustering at the level of randomization.[11] Our primary outcomes of interest are presented in Table 9. 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. We have 80 total outcomes of interest.

[10] We anticipate this because, as described in Section 2, we sampled households where both the man and the woman are eligible and, when possible, we further restricted the sample to households where both the man and the woman are interested in participation. We chose to do so because this set of households would constitute our compliers, and therefore would drive any estimates, but focusing our sampling on them improves precision. We discuss this further in Section 4.2.

[11] In an appendix, we will present the reduced form with standard errors from either cluster robust variance covariance estimators or randomization inference to test the robustness of our inference.

Table 9: Pre-specified outcomes

	Impacts of Woman participant		Impacts of Any participant	
	Midline (1)	Endline (2)	Midline (3)	Endline (4)
Predicted household consumption	X	X	X	X
Women preferred consumption	X	X	X	X
Men's earnings	X	X	X	X
Women's earnings	X	X	X	X
Time spent				
outside the home (men)	X	X	X	X
outside the home (women)	X	X	X	X
working (self-employment, men)	X	X	X	X
working (self-employment, women)	X	X	X	X
working (salary, men)	X	X	X	X
working (salary, women)	X	X	X	X
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		
Locus of control	X	X	X	X
Subjective well being	X	X	X	X
Intimate partner violence	X	X	X	X

Notes: Pre-specified outcomes are marked with an "X" and are summarised in this table. Columns (1) and (2) are outcomes of Woman participant, as estimated in Equation 1 and 4. Columns (3) and (4) are outcomes of Any participant, as estimated in Equation 1 and 4.

Before introducing additional specifications, we briefly discuss the interpretation of the

coefficients on women's participant ($\beta_{1,ct}$) and any participant ($\beta_{2,ct}$), and the test of equality across countries. First, we note that in the presence of heterogeneity in treatment effects of Any Participant, instrumental variables will not necessarily estimate a local average treatment effect of Woman Participant. However, if the shift from Cash-for-Work to Cash-for-Women's work does not affect the composition of participating households, $\beta_{1,ct}$ is the local average treatment effect among complier households who shift from having a man participate to having a woman participate. We anticipate this will approximately hold in practice, and therefore ignore bias from heterogeneity in treatment effects of Any Participant that is correlated with impacts of Cash-for-Women's work, relative to Cash-for-work, on Any Participant. Second, letting τ_{it} be the treatment effect for individual i in period t from country $c(i)$ we note that $\beta_{1,ct} = \mathbf{E}[\tau_{it}|c(i), i \text{ is a Cash-for-Women's work complier}]$. This highlights the reasons why $\beta_{1,ct}$ may differ across contexts, limiting external validity. First, the characteristics of Cash-for-Women's work compliers may differ across countries, and these characteristics may drive heterogeneity in program responses. These differences could be driven by differences between compliers and targeted populations that vary across contexts, or differences in the targeted population across contexts. Second, even conditional on participant characteristics, program impacts may genuinely differ across countries. This may be due to differences in the programs being implemented, or differences in the contexts themselves that affect the impacts of the program, potentially mediated by program implementation. To partially address this, we next allow program impacts to vary by characteristics of participants that are likely to drive heterogeneity in responses. We therefore change our test of external validity from testing whether the local average treatment effect is equal across countries, to whether the local average treatment effect conditional on observable characteristics is equal across countries. We may therefore reject the null of external validity in this test if either there are differences in unobservable characteristics of Cash-for-Women's work compliers across countries that drive treatment effect heterogeneity, or if there are differences in the impact of the program in different countries.

As described above, we therefore will conduct an analysis of heterogeneity. We estimate

$$Y_{hct} = \text{Woman participant}_{hc1} W'_{hc} \beta_{1,ct} + \text{Any participant}_{hc1} W'_{hc} \beta_{2,ct} + W'_{hc} \gamma^{Y}_{1,ct} + X'_{hc0} \gamma^{Y}_{2,ct} + \epsilon^{Y}_{hct} \tag{4}$$

where W'_{hc} is a vector including a constant and household characteristics for which heterogeneity is of interest. First stages for Woman participant$_{hc1} W_{hc}$, and Any participant$_{hc2} W_{hc}$ are analogous to Equations 2 and 3, but with Cash-for-Work$_{hc} W_{hc}$ and Cash-for-Women's work$_{hc} W_{hc}$ also included as instruments, separately for each country. We report F-tests for the equality of $\beta_{1,ct}$ across countries. 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 Woman participant across contexts.[12]

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 would 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 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 respondent's age. We would 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 participation in public works. Third, we include a dummy for the presence of other women in the household. We would interpret larger impacts when other women are present in the household as suggestive that within household within gender substitution of tasks is a key mechanism through which women shift their time use.

In summary, this corresponds to three sets of tests for the 80 outcomes we specify —

[12] In supplementary analysis we will also report the cross country average of $\beta_{1,ct}$, focusing on heterogeneity of impacts of Woman participant.

a test of the null hypothesis that the average effect across contexts is 0, a test of the null hypothesis that effects are identical across contexts, and a test of the null hypothesis that effects are identical across contexts conditional on observable characteristics of households.

4.1.1 Attrition

For each regression and each country for our primary specification, we will test for differential attrition and, for questions where men respond when present, differential attrition of male respondents. When statistically significant attrition is present for a given country and a given outcome, we will estimate Lee (2009) Bounds for the reduced form for each country for that outcome, and report the average upper bound and average lower bound in robustness.

4.2 Power calculations

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

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

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

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

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 10, along with country and estimating equation specific assumptions. All other assumptions and the details of the calculations are discussed below.

Table 10: Power calculations

	El Salvador	Rwanda
Treated		
Number of observations	1500	1173
Number of clusters	75	78
Share treated	0.67	0.67
MDE	0.099	0.104
Anticipated take-up	0.9	0.9
Transfer size	0.3	0.5
Expected effect	0.180	0.302
Cash-for-Women's work		
Number of observations	1000	786
Number of clusters	50	52
Share treated	0.5	0.5
MDE	0.344	0.363
Anticipated take-up	0.9	0.6
Transfer size	0.3	0.5
Expected effect	0.181	0.201

For the first power calculation, we use Women preferred consumption 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, $\text{MDE} = \sigma_\epsilon(z_{0.8} + z_{0.975})\sqrt{\frac{1+\rho(m-1)}{NP(1-P)}}$, 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.[13] To construct a single measure we can use across contexts, we normalize by average household consumption. Lastly, we replicate this exercise for Women preferred consumption 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 Women preferred consumption.

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.[14] 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,[15] and the monthly transfer size relative to average monthly household consumption.

Summarizing Table 10, we are powered to detect the effect of Cash-for-Work in each country. We are underpowered to detect the effect of Cash-for-Women's work in any in-

[13] This 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.

[14] Although 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.

[15] Formally, 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.

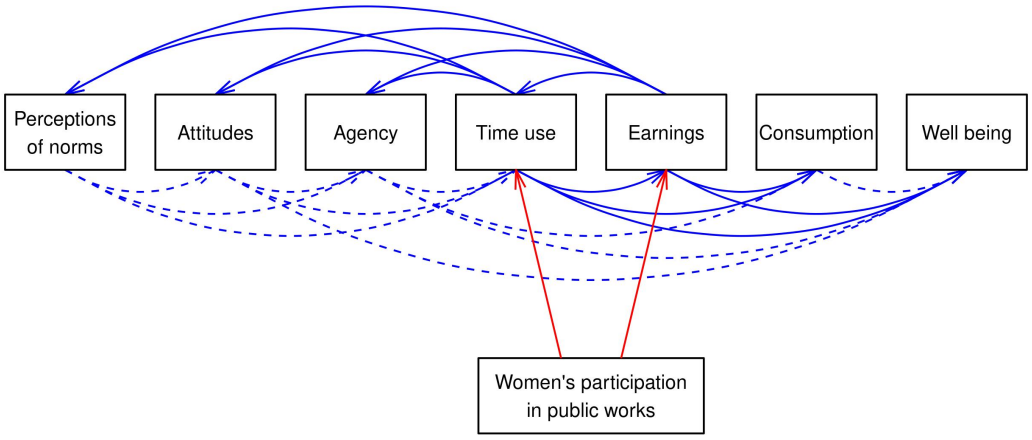
dividual country. However, we are powered to detect the effect of Cash-for-Women's work in the cross-country analysis (heuristically, the cross-country MDE is inversely proportional to the square root of the number of countries), highlighting the value of the cross-country analysis.

4.3 Results

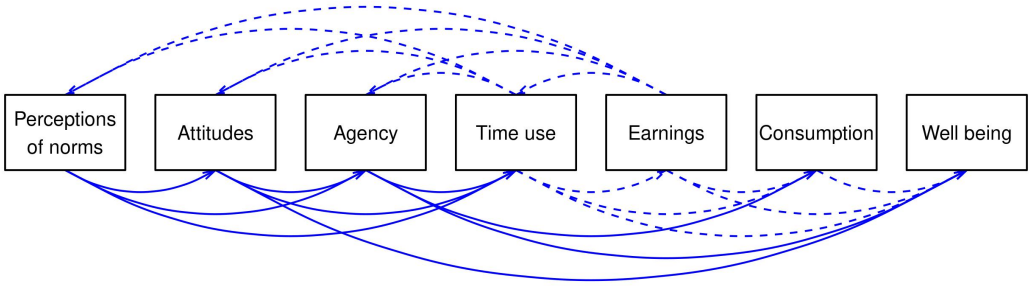
In Figure 1, we lay out a theory of change for the intervention based on the context to guide our discussion of the results, with arrows communicating directions of causality. We use this theory of change to guide our analysis described in Section 4.1, including both our primary analysis and analysis of heterogeneity. We focus on the impacts of women's participation in public works estimated in Equation 1, as the impacts of household participation in public works are well established. Figure 1a shows short run effects of women's participation in public works (corresponding to outcomes in the midline survey), Figure 1b shows medium run effects of women's participation in public works (corresponding to outcomes in the endline survey), while Figure 1c shows which outcomes in Table 9 correspond to which nodes of the theory of change.

Figure 1: Theory of change

(a) Theory of change: Short run



(b) Theory of change: Medium run



(c) Outcomes

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

Attitudes
Women's attitudes, women's time use
Women's attitudes, women's agency
Men's attitudes, women's time use
Men's attitudes, women's agency

Agency
Consumption
Women's time use

Time use
Women, outside the home
Women, self-employment
Women, salaried
Women, chores
Men, outside the home
Men, self-employment
Men, salaried
Men, chores

Earnings
Women's
Men's

Consumption
HH, predicted
Women preferred consumption

Well being
Locus of control
Subjective well being
IPV

First, in Figure 1a, impacts of women's participation in public works 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 women's participation in public works shifted only agency, time use, earnings, and consumption. We would conclude that impacts on agency were caused by changes in earnings and time use. However, impacts on consumption would be interpreted as driven by some combination of direct effects of changes in earnings and time use, and indirect effects through changes in agency. Alternatively, suppose we saw that women's participation in public works 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 earnings, 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 women's participation in public works 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

Adjognon, G. S., van Soest, D., & Guthoff, J. (2020). Reducing hunger with payments for environmental services (pes): Experimental evidence from burkina faso. *American Journal of Agricultural Economics*.

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., Pande, R., Rigol, N., Schaner, S., & Moore, C. T. (forthcoming). On her own account: How strengthening women's financial control affects labor supply and gender norms. *The American Economic Review*.

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.

Gazeaud, J., Mvukiyehe, E., & Sterck, O. (2019). Cash transfers and migration: Theory and evidence from a randomized controlled trial.

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.

Imbert, C. & Papp, J. (2015). Labor market effects of social programs: Evidence from india's employment guarantee. *American Economic Journal: Applied Economics*, 7(2), 233–63.

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.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3), 1071–1102.

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.

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.

Appendix A Administrative information

Appendix A.1 Confirmation of ethics approval

This study was approved by Solutions IRB (2020/07/27).

Appendix A.2 Status of funding

This research benefits from generous funding from the Office of Evaluation of the World Food Programme.

Appendix A.3 Conflicts of interest

Christian, Dinarte, Jeong, Kelley, Kondylis, Lane, and Loeser have no conflicts of interest to declare. Heirman is employed at WFP's independent Office of Evaluation. The WFP Office of Evaluation is independent of WFP Management and has a a track record of publishing positive and negative findings, and evaluators (including PIs) at WFP Office of Evaluation cannot be involved in the management or implementation of WFP programs, or have any vested interests the results of evaluations.

Appendix A.4 Acknowledgements

This design benefited from comments from Emma Frankham, Macartan Humphreys, Sarah Khan, Jeremy Magruder, and two anonymous reviewers. This research benefits from generous funding from the Office of Evaluation of the World Food Programme. We thank Saahil Karpe, Felipe Dunsch, and Kristen McCollum for superb research coordination. Rasha Absi, Sylvia Cesar, Marc Andrea Fiorina, Mariachiara Iannuzzi, Eric Jospe, Edwin Kasila, and Miguel Melendez provided excellent research assistance. Finally, we thank the management and technical staff at WFP Country Offices for being outstanding research partners. This study was approved by Solutions IRB (2020/07/27), and was registered in the AEA RCT

Registry (AEARCTR-0005933) prior to collection of outcome data. The views expressed do not reflect the views of the World Bank nor the World Food Programme. All errors are our own. At time of publication of this pre-analysis plan, four of an anticipated six countries have been selected (El Salvador, Kenya, Rwanda, and Syria). Selection of the two remaining countries is ongoing. In the event that we are only able to implement the experimental design described in this pre-analysis plan in these four countries, we will proceed with the analysis as described. Corresponding authors: Florence Kondylis (`fkondylis@worldbank.org`) and John Loeser (`jloeser@worldbank.org`).

Appendix B Selection of goods for consumption module

The analysis in Section 4 specifies two consumption outcomes – household consumption and women preferred consumption. To avoid doing a full consumption module, before each baseline we identified goods in each country that strongly predict household consumption and women's income. For each country, we used the full set of consumption expenditure goods available in the most recent nationally representative household survey with a full consumption module; the survey used is reported in Table A1. The analysis below restricted to a sample of households likely to be comparable to targeted beneficiaries of FFA, with details on the sample reported in Table A1.

For household consumption, in each country we first identified 5 goods that most strongly predicted household consumption. Let $y_i^{\perp}$ denote nominal annualized consumption of household i, demeaned and standardized, and let $x_i^{\perp}$ denote the vector of demeaned and standardized expenditures on each good of household i. We demeaned by regression on a vector of controls for household composition and enumeration area dummies.[16] We then estimated

[16]In El Salvador and Rwanda, controls were number of women and number of men in the household, number of household members under 2, 5, 10, and 16, and enumeration area fixed effects.

Equation A1 by LASSO using the R package `glmnet`.

$$y_i^{\perp} = x_i^{\perp\mathsf{T}}\beta + \epsilon_i^{\perp} \tag{A1}$$

We adjusted the regularization parameter until at least 5 goods were selected; these goods are reported in Table A1. Next, let y_i denote nominal annualized consumption of household i, and x_i^* denote nominal annualized expenditure of household i on the selected 5 goods. As a final step, to generate our measure of predicted consumption, we then estimate in the same household survey

$$y_i = \alpha + x_i^{*\mathsf{T}}\beta^* + \epsilon_i \tag{A2}$$

We reported the estimated coefficient β^* in Table A1, along with the R^2 from this regression. To construct predicted household consumption in each survey wave, we will simply use $x_i^{*\mathsf{T}}\hat{\beta}^*$ (omitting the constant), which we use as an outcome. Following Athey et al. (2016), in order to interpret estimated impacts on $x_i^{*\mathsf{T}}\hat{\beta}^*$, we assume that the distribution of household consumption conditional on consumption on the 5 selected goods is the same in the nationally representative household survey and conditional on treatment status in our household survey ("comparability"). To ensure comparability, the description of each consumption good and the question closely followed the consumption module from the nationally representative household survey. Under this assumption, unobserved impacts on household consumption are bounded between 1 and $1/R^2$ times as large as impacts on predicted household consumption $x_i^{*\mathsf{T}}\hat{\beta}^*$. We typically find $R^2 \approx 0.5$, so these bounds are relatively tight.

Table A1: Consumption module good selection

	El Salvador	Rwanda
Survey	ENIGH (2006)	EICV5 (2016-2017)
Sample restrictions	Rural	Rural, Outside Kigali, Ubudehe 1 or 2
Predicted household consumption		
R^2	0.54	0.50
Goods (Coefficient)	Clothing[17] (3.56)	Educational expenditures (1.83)
	Soft drinks (3.84)	Airtime (9.72)
	Airtime (7.67)	Women's footwear (24.9)
	Hygiene (5.23)	Women's tailoring (41.3)
	Beef (9.04)	Beauty and cosmetic products (20.9)
Women preferred consumption		
Goods	Utensils	Women's footwear
	Airtime	Men's footwear
	Tools	Health insurance
	Mandarin oranges	Men's haircut
	Sweet corn	
	Medicine	
	Medical services	
Other selected goods		
Goods	Men's clothing	
	Women's clothing	
	Educational expenditures	

For women preferred consumption, in each country we first identified 5 goods that most strongly predicted women's income. Let $y_i^{\perp}$ denote nominal annualized women's income of household i, demeaned and standardized, and let $x_i^{\perp}$ denote the vector of demeaned and standardized expenditures on each good of household i. Income by household member was constructed using all questions on earnings from the household roster. We demeaned by regression on a vector of controls for household composition, household income and consumption, and enumeration area dummies.[18] We then estimated Equation A3 by LASSO using the R package `glmnet`.

$$y_i^{\perp} = x_i^{\perp\mathsf{T}}\beta + \epsilon_i \tag{A3}$$

We adjusted the regularization parameter until at least 5 goods were selected; these goods are reported in Table A1. As a last step, we excluded goods that did not have statistically

[18]In Rwanda, controls were number of women and number of men in the household, number of household members under 2, 5, 10, and 16, household consumption and household income, and enumeration area fixed effects.

significant coefficients in an OLS regression clustering at the enumeration area. To construct women preferred consumption, we will estimate Equation A3 by OLS in each survey wave in each country using all 10 selected goods, and use predicted women's income as the outcome.

Lastly, in addition to the goods selected above, we included educational expenditures, one item of men's clothing, and a comparable item of women's clothing in each survey. These additional goods are listed in Table A1.