

Long-Term Impacts of Short Exposure to Conditional Cash Transfers in Adolescence:

Evidence from the Philippines

Ervin Dervisevic*, Elizaveta Perova*, Abhilasha Sahay*

This paper evaluates the long-term impact of the national conditional cash transfer program in the Philippines - *Pantawid Pamilyang Pilipino Program (4P)* - on beneficiaries who were exposed to it during a short yet critical period of their life, i.e. while transitioning from adolescence to adulthood. We estimate impacts on men and women who were enrolled in the program for up to 1.5 years when they were aged between 12.5 and 14 and are currently in their early twenties. We find evidence of impacts on marriage and fertility for women: participation in the program is associated with delay in marriage and in the first birth by approximately one year and six months, respectively. We do not find any effect on education, employment, and welfare-related outcomes. We also do not find strong and consistent evidence of changes in empowerment or gender norms.

Key words: cash transfers; long-term impact; gender; randomized intervention; Southeast Asia

JEL Codes: C93, I38, J12; J13; J16

We gratefully acknowledge funding from the Umbrella Facility for Gender Equality (UFGE) and Australia's DFAT to carry out this work. The data collection protocols were approved by Health Media Labs IRB. We thank Berk Özler, Amber Peterman, Gabriel Demombynes, Jorge Araujo, Yoonyoung Cho and participants of East Asia and Pacific Chief Economist Office seminar for insightful comments. IPA Philippines carried out data collection. We are grateful to Jorge Franco, Reinier Guzman, Ma Angelica Cruz and Alana Teman for excellent support during data collection. All errors are ours.

* World Bank

1. Introduction

The primary objective of most conditional cash transfer (CCT) programs is breaking inter-generational transmission of poverty through investments in human capital. An extensive body of research suggests that CCTs succeed in increasing investment in human capital (see Saavedra and Garcia, 2012, and Fiszbein and Schady, 2009, for reviews of evidence on impacts of CCTs on education and health outcomes). There is relatively less evidence on whether children whose families benefitted from CCT programs enjoy better labor market and welfare outcomes as adults. Last years have seen impressive increase in the literature on longer-term impacts of CCTs. However, a consensus on the CCT's ability to reduce inter-generational transmission of poverty has not emerged. Some studies find impacts of large magnitude on education, labor market and welfare outcomes (for example, Parker and Vogl, 2018 in case of *Progresa* in Mexico); other suggest zero or very low impacts (for example, Araujo et al., 2016 in case *Bono de Desarrollo Humano* in Ecuador or Baird, McIntosh, and Özler (2019) in Malawi).

Moreover, while a significant share of evaluations of short-term impacts of CCTs relied on randomized controlled trials (RCTs), longer-term evaluations are mostly either limited to capturing relative, rather than absolute impacts or face greater threats to rigorous identification of impacts, than in a case of an RCT. In the former group, studies based on randomized roll out of the programs identify long-term impacts from differential length of exposure (Behrman, Parker and Todd, 2009; Barham, Macours and Maluccio, 2016). While they generally suggest that additional length of exposure results in improvements in education and labor market outcomes, they do not capture absolute impacts - differences in outcomes compared to what would have happened in the absence of the program. To our knowledge, the only exception is Baird, McIntosh, and Özler (2019), who evaluate an impact of approximately two years of exposure to CCTs and UCTs two years after the program completion in a cluster randomized study, with pure control group preserved until the end of the data collection. The fact that CCT/UCT evaluated is not a large-scale governmental program also sets this study apart.

In the latter group, several non-experimental studies capture absolute impacts relying on a range of non-experimental econometric methods, including matching and regression discontinuity (Baez and Camacho, 2011), cohort difference-in-difference (Parker and Vogl, 2018) or OLS weighted by inverse probability of receiving treatment, based on propensity scores (Kugler and Rojas, 2018). While the counterfactual in these studies are the outcomes in the absence of the program, identification frequently requires several assumptions, validity of which may be violated. Several studies combine both elements: evaluation of

relative impacts, based on randomized rollout, and evaluation of absolute impacts based on non-experimental methods (Behrman, Parker and Todd, 2011; Araujo et al., 2016).

We attempt to contribute to this literature by analyzing long-term impacts of *Pantawid Pamilyang Pilipino Program* (4P), the flagship program of the Government of the Philippines. We take advantage of randomized rollout of the program in 2008 and the fact that children close in age to eligibility cut-off did not benefit from the program, even when it was rolled out to their *barangays*. Thus, we can provide experimental evidence on absolute long-term impacts of the 4P program.

Specifically, we exploit 2 important aspects of the 4P program. First, the program was initially rolled out in 2008 in few randomly selected barangays. However, after 2.5 years the control barangays were also included in the program. Second, the program initially targeted children in the age group 0-14 years old. However, the age limit was later increased to 18 years in 2015. Therefore, among control barangays, children who were 12.5 years or older during initial roll out never received the program; they were neither eligible when the program first arrived at their barangays (due to age restriction of 14) nor when the age restriction was raised to 18 in 2015 (since this cohort was above 18 by 2015). Thus, they constitute a pure control group. To eliminate spillovers from younger siblings, we focus on families where the youngest child was aged 12.5 years at the moment when eligibility was determined.

This identification strategy differentiates our paper from other studies focused on the long-term impacts of CCTs in two additional dimensions. First, we capture impacts of relatively short-term exposure: treated individuals in our sample benefitted from the program during a year and a half at most. Three studies which capture relative impacts of CCTs find positive impacts of comparably small number of *additional* years of exposure: 1.5 years in case of *Progresa* (Behrman, Parker and Todd, 2009; Behrman, Parker and Todd, 2011) or 2 years in case of *Red de Proteccion Social* in Nicaragua (Barham, Macours and Maluccio, 2016). However, the least amount of exposure in the sample (for control group) was 4.5 years in case of *Progresa*, and 3 years in case of *Red de Proteccion Social*. While current evidence establishes positive impacts of *additional* 1.5 years of receiving benefits, we will attempt to answer the question whether just 1.5 years of exposure to a CCT program may significantly improve adult outcomes. In this respect, this study is most similar to Baird, McIntosh, and Özler (2019), who focus on impacts of approximately two years of receiving CCT benefits and find no evidence of sustained impacts on average¹. With the exception of Baird, McIntosh and Özler (2019), current long-term evaluations of CCTs capture impacts of at least 5

¹ Baird et al. (2016) find large effects on school attainment, marriage and childbearing among school dropouts at baseline.

and up to 10 years of exposure. Second, the 4P beneficiaries in our study participated in the CCT program during a critical juncture, when adolescents and their parents may make important decisions about continuing schooling, which in turn affect marriage and fertility.

We also expand the set of outcomes explored in the current literature of long-term impacts of CCTs. In addition to impacts on schooling, labor market outcomes and welfare, this study relies on data on a rich set of empowerment indicators. Given complexity of the concept of empowerment, and lack of consensus about optimal measurement of empowerment, we collect data on several of its proxies: (i) socio-economic proxies of empowerment, such as age at marriage, at first birth and prevalence of transactional sex; (ii) direct measures of empowerment, such as self-efficacy, relative autonomy or participation in intra-household decision making. The only study that includes measures of empowerment is Baird, McIntosh and Özler (2019) in Malawi. Finally, we explore whether benefitting from a CCT program in adolescence could impact exposure to gender-based violence (GBV) in adulthood and gender norms.

Thus, the contributions of this paper to the literature on the long-term impacts of the CCTs is threefold. First, we provide experimental estimates of absolute impacts of short-term exposure to a CCT program during adolescence on adult outcomes, 10 years after receipt of benefits stopped. Second, we contribute to the literature by exploring long-term impact of CCTs on a broad range of empowerment impacts, including GBV and norms. In that way, we build on evidence provided in Baird, McIntosh and Özler (2019) by focusing on longer term impacts (approximately 10 years after exposure) and expanding the set of empowerment indicators. Finally, to our knowledge, current literature on long-term impacts of CCTs is limited to Americas and one African country. We expand this literature to a different region. The paper is structured as follows: we provide a brief description of the program and its rollout in the next section. Section 3 reviews literature on long-term impacts of CCTs. Section 4 describes the data. Section 5 presents theory of change and lists outcomes of interest. Section 6 describes identification strategy. Section 7 presents the results. Section 8 describes deviations from pre-analysis plan. Section 8 concludes.

2. Program overview

The 4P program was initiated in 2008. In its first year, it registered approximately 300,000 beneficiary households. In succeeding years, the program continuously expanded its coverage and currently more than four million poor households nationwide are benefitting from it. With the expansion in coverage, the budget allocated to this program has also increased; from 50 million pesos in 2008 to 78 billion pesos in 2017. This constituted 0.5 percent of the nation's GDP in that year. With approximately 60 percent of the

poorest quintile of households in the country covered (World Bank, 2018), the program is a core pillar of the government's social protection strategy.

The 4P program provides three types of transfers: education grant, health grant and rice-subsidy². The education grant is provided to every child who complies with the education conditions of the program: is enrolled in school and attends 85 percent of the school days every month. Children enrolled in daycare/kindergarten or elementary schools receive 300 pesos per month while children enrolled in high school receive 500 pesos per month. These grants are provided for a period of 10 months per year. The program maintains a limit of three child beneficiaries for the education grants.

The health grant amounts to 500 pesos per month and is given to households subject to compliance of all health conditionalities. These conditionalities include, (i) all children under the age of five must regularly visit a health center or rural health unit for growth monitoring, vaccines and preventive health check-ups; (ii) pregnant women must visit their health center monthly for antenatal and postnatal care, and must deliver in a health facility attended by a trained health professional; (iii) all school-aged children (6-14 years old) are to comply with de-worming protocol at schools; and (iv) among households with children 0-14 years old, the household grantee (mother) and/or spouse must attend Family Development Sessions (FDS) at least once a month.

To be eligible for the program, households must be identified as poor by the *Listahanan*, formerly known as National Household Targeting System for Poverty Reduction Program (NHTS-PR). They also must have a pregnant member or at least one child aged 0-18 years old at the time of selection and must be willing to commit to meeting the program conditionalities. The *Listahanan* targets beneficiaries through a household assessment and application of a Proxy Means Test (PMT) methodology to predict income of households based on characteristics such as household composition, education, housing conditions, access to basic services, ownership of assets, etc. Predicted incomes are then compared with poverty thresholds at the provincial level to identify households below (poor) or above (non-poor) these thresholds.

The program has undergone two key changes since its inception in 2008. First, during initial rollout of the 4P program, few randomly selected barangays were given treatment. However, 1.5 years after the program was rolled-out, the control group was also exposed to the program. Second, initially the program was designed to provide benefits for children 0 – 14 years old. However, this age limit was extended to

² We do not provide detailed description of the rice subsidy as it was introduced in 2017.

18 years old in 2015, with the motivation of supporting beneficiary children to at least finish high school and thereby increase their chances of getting better employment and higher income.

3. Previous literature

Recent years have seen emergence of several studies focused on long-term impacts of the CCT programs. These studies aim to establish whether indeed CCT programs achieve the objective of breaking inter-generational cycle of poverty through investments in human capital. They largely focus on impacts on education, labor market outcomes and wealth proxies. We organize the review below by these outcomes, placing stronger emphasis on government-run CCT programs, which are more akin to the 4P program.

Education

All currently available studies find some impacts on education, captured either in the likelihood of high school completion or number of years of education, at least for some subgroups. Kugler and Rojas (2018) find that exposure to *Progresa* increased the likelihood of completing high school, the number of years of education attained and the likelihood to attend tertiary educational institution. They find that impacts are weaker for girls and for children of illiterate women. In their sample, average exposure to the treatment is around 7 years. Results from Parker and Vogl (2018) as well as Behrman, Parker and Todd (2011) corroborate this conclusion: they find increase in number of grades completed for both sexes.

Baez and Camacho (2011) also find that beneficiary students of *Familias en Accion* in Colombia are more likely to graduate from high school. In their study, impacts are stronger for girls and beneficiaries from rural areas. Barham, Macours and Maluccio (2018) find positive impacts of Nicaraguan CCT *Red de Proteccion Social* on the number of grades attained for men (the study is limited to men). Araujo et al. (2016) also find that *Bono de Desarrollo Humano* increased schooling attainment for women, but by a small margin – between 1-2 percent above 75 percent counterfactual completion rate. Notably, Ham and Michelson (2018) find impacts of PRAF in Honduras on years of schooling only when transfers were complemented with supply-side incentives to medical facility. Baird, McIntosh and Özler (2019) find, perhaps, the most discouraging results, suggesting that only girls who were dropouts at baseline experienced improvements in passing primary school and increasing the number of completed grades.

The evidence on learning achievement is less conclusive: Barham, Macours and Maluccio (2018) find large and significant improvements in math and language scores in Nicaragua. However, Baez and Camacho (2011), Behrman et al. (2009a) and Araujo et al. (2016) find zero impacts on academic achievement in Colombia, Mexico and Ecuador, respectively.

Labor market outcomes

Similar to learning achievement, in the case of labor market outcomes up to date evidence suggests a mix of zero and positive impacts. Rodriguez-Oreggia and Freije (2012) do not find impacts of *Progresa* on employment, wages or inter-generational occupational mobility. Similarly, Araujo et al. (2016) find zero impact of *Bono de Desarrollo Humano* in Ecuador on the likelihood of working. Behrman, Parker and Todd (2009a) and Behrman, Parker and Todd (2011) find evidence of negative impact of *Progresa* on employment for younger boys (15-16) – possibly due to the fact that they are more likely to remain in school longer. Baird, McIntosh and Özler (2019) do not find sustainable impacts of CCTs on labor market outcomes, even for baseline dropouts who experienced gains in education.

However, several studies find evidence of positive impacts of CCTs on extensive (labor force participation, employment) and intensive (quality of employment) margins, at least for one gender. Thus, Behrman, Parker and Todd (2011) find increased likelihood of working for older girls (19-21) due to participation in *Progresa*. Similarly, Ham and Michelson (2018) find that combination of transfers and supply side incentives under *PRAF* in Honduras is associated with increased labor force participation and probability of working from home for women only.

Parker and Vogl (2018) find positive impacts of *Progresa* on labor market outcomes: their results suggest that exposure to *Progresa* is associated with increased labor force participation and labor earnings for women, and increased labor supply, lower likelihood to work in agriculture and higher likelihood to work in better paid occupations and industries. This result is corroborated in Kugler and Rojas (2018), who use different identification: they find significant increases in employment, hours, income, and employment arrangements secured through contracts and non-wage benefits. They see impacts largely being driven by men and children of literate women.

Other outcomes: income, wealth proxies, migration and delayed marriage

Most long-term evaluations of CCT programs focus on education and labor market outcomes. However, some expand the set of explored outcomes. Parker and Vogl (2018) analyze impact on household income as well as housing and asset indices, which can serve as proxies for wealth. They find positive impacts on housing and asset indices for men and women, and on household income for women only. They also suggest that increase in household income may partially stem from the fact that former *Progresa* beneficiaries are more likely to marry more educated people who earn better incomes. Similarly, Baird, McIntosh and Özler (2019) find lasting impacts on marriage, fertility and desired lifetime fertility, limited

however to baseline dropouts only. Behrman, Parker and Todd (2009) find that *Progresa* decreases likelihood to migrate for boys and delays the age of marriage.

4. Data

To identify causal effects, we rely on an RCT methodology, exploiting the initial randomized roll out of the program at the barangay level. We collect the data on individuals who are currently in the age group of 23-24.5 years old, and compare young adults who were residing in treatment barangays during adolescence (and thus were eligible to receive the 4P transfer) to young adults who were residing in control barangays during adolescence (and weren't eligible to receive the transfer at any point in their life).

To identify and track these young adults, we use data from *Listahanan* (2008-09). Using the *Listahanan*, we were able to identify households who had children in the age-group 12.5-14 years during the time of initial targeting and whether a household was identified to be eligible to receive the 4P transfers. To ensure that we did not pick up any spillovers across siblings within the same household, we limited the sample to families where the child aged 12.5 years or older was the youngest child in the family when the program was expanded to control barangays. We provided the data collection firm with the list of eligible households for each of the 71 barangays included in analysis.

The firm randomly drew a required subsample of households, stratified by gender in each barangay, and validated that the respondent households were indeed residents of the barangay with local officials. If the household was not recognized by the local official, the household was replaced. 81% of replacements are from the same barangay; 19% - from a different barangay.

After the validation with barangay officials, the field team visited households and administered a short tracking questionnaire with the objective of (i) confirming that at the time of initial 4P rollout there was a child aged 12.5 to 14 in the household; (ii) collecting information on where this child currently resides. If the household could not confirm that one of its members back in 2008 was in the age group and gender as described by *Listahanan*, the field team would replace this household. In case the child (who is now a young adult and our respondent of interest) continued to reside in his/her original HH, the field team directly approached him/her and sought permission to conduct the primary IE survey (ideally on the same day). In case the child has moved out, the field team collected information on his/her current location from the parents or older members in the original household. Thereafter, the field team tracked the child who had moved out and sought his/her permission to participate in the study. We instructed the survey

firm to track respondents who moved since the time of the initial rollout of the program as long as the household/respondent has moved within the same municipality or if she/he has moved to one of the main economic centers in Philippines, such as, metro Manila, Cebu, Davao and Cagayan de Oro. About 29% of our sample have moved to major cities. In instances, where the young adult is married or is living with a partner, we also interviewed the partner. Note that the household questionnaire was administered to the current, not original household.

Based on our power calculations, our target was to interview 1,576 young adults: 438 beneficiary women, 438 beneficiary men, and 700 partners in 876 households. Early stages of fieldwork proved that we miscalculated co-habitation rate in the Philippines. To ensure that we have sufficiently large sample of partners, we increased the number of young adults' households to 932. However, largely because the co-habitation rate was lower than originally programmed, our final sample is 87% of the target sample: 1,372 respondents, 870 of them are former beneficiaries of the 4P program, 502 partners of the former beneficiaries. This work focuses on the sample of former beneficiaries.

We collected the data between October 2019 and January 2020 from 932 young adult households, spread across approximately 71 barangays. The survey collected extensive data using two questionnaires: household and individual. The household questionnaire included modules on (i) basic demographic characteristics, (ii) economic participation, (iii) housing conditions, (iv) asset ownership, and (v) participation in the 4P program. The individual questionnaire includes modules on (i) decision making, (ii) time use, (iii) self-efficacy, (iv) wellbeing, (v) social capital, (vi) attitudes towards gender norms, gender-based violence and violence against children, (vii) pregnancy history, (viii) exposure to gender-based violence, and (ix) exposure to transactional sex. The last three modules are only administered to the women in the sample. The sub-sample of female respondents was randomly split in two groups for sections 8 and 9: half of the sample was interviewed using Face-to-Face Direct Questioning (FTFDQ) technique, another half responded to the same question in Audio Computer Assisted Self-Interview (ACASI). As we do not find differences in prevalence rates reported via FTFDQ vs. ACASI³, we pool the responses in the current analysis, but control for the type of the interview.

When collecting data on GBV, we followed WHO recommendations in elaborating data collection protocols, including working with same-sex enumerators, specialized training for enumerators, protocol

³ Detailed results available upon request.

for potential interruptions, and provision of information on support services for GBV survivors. We received research ethics review approval from Health Media Labs IRB.

Cumulatively, the survey took about 1-1.5 hours (per person) to administer. We complemented face-to-face data collection with a phone survey in April-May 2020 to correct the data on receipt of the 4P benefits: in our initial round of data collection the question about receipt of the 4P benefits was addressed to the respondent of the tracking questionnaire, which resulted in high number of missing values and inconsistencies. The phone survey was administered to target respondents: young women and men who resided in treatment and control barangays at the time of program rollout in 2008.

5. Theory of change and outcomes of interest

A key objective of the CCT programs is breaking the cycle of inter-generational transmission of poverty. Thus, we expect that adolescents, who have benefitted from the 4P program at the age of 12.5-14, will enjoy better socio-economic outcomes as adults. we explore this hypothesis by analyzing impacts on education, labor market outcomes and proxies of welfare. For education, we estimate impact on the likelihood to complete schooling, the number of years of education and probability of attending vocational training. Our labor market outcomes include indicator variables for labor force participation, employment, paid employment and underemployment. We collect a rich set of variables on asset ownership and housing conditions and construct three types of indices: an index based on principal component analysis, average z-scores index based on Katz, Kling and Liebman (2007), and an Inverse Covariance Weights (ICW) index obtained by weighting z-scores of variables with inverse of their covariance matrix (Anderson, 2008). The variables used in the construction of housing conditions index include roof, walls, floor materials, access to piped water and electricity, number of bedrooms, presence of kitchen and type of toilet, type of cooking fuel used. We create asset index based on ownership of residential plot, non-agricultural buildings, farm machinery and implements, livestock and a long list of durables, such as radio, sewing machine, television set, microwave oven, etc.

We expect CCTs to affect empowerment through several channels. Empowerment is a complex construct, with a wide range of measures used to capture empowerment empirically, from socio-economic proxies to more direct measures, such as relative autonomy index (RAI) or self-efficacy index. In this study, we rely on a wide range of measures of empowerment used in previous literature. In the discussion of the theory of change below we link channels to specific measures of empowerment. However, we use these connections as an organizational device, rather than assert dominance of specific channels behind changes in empowerment. The measures we use represent the same latent construct and are

interconnected. For example, greater autonomy may lead to higher decision making within the household and higher likelihood to leave/not enter an abusive relationship, resulting in lower exposure to GBV.

Changes in empowerment through education and better employment prospects

Better employment prospects may increase opportunity costs of early marriage and childbearing. Consequently, we expect to find positive impacts on two socio-economic proxies of empowerment: marriage and fertility, which we capture in two variables: age at marriage and age at first birth. We also check for impacts on prevalence of transactional sex.

Similarly, better education, health and employment opportunities may affect aspirations and self-perception, leading to increase in direct measures of empowerment, such as generalized self-efficacy index (Jerusalem and Schwarzer, 1995) and RAI (Vaz et al., 2016). We construct generalized self-efficacy index based on respondent's agreement or disagreement with eight statements about one's ability to achieve their goals, such as "I believe I can succeed at most any endeavor to which we set my mind"⁴. We use both: the sum of raw scores, which totals a range of 8 to 32 and increases in self-efficacy, and also construct a z-score of answers, following Katz, Kling and Liebman (2007).

RAI is based on Self-Determination Theory in psychology, which postulates that a person is autonomous when his or her behavior is experienced as willingly enacted and when he or she fully endorses the actions in which he or she is engaged or the values expressed by them (Vaz et al., 2016). The index captures respondent's ability to act on what she or he values through vignettes which describe people making decisions in different domains. In our survey, we use 6 domains: making small purchases, large purchases, using own earnings, household budget, schooling of children and treatment of sick children.

Each vignette suggests a motivation for action: external (when one's action is effectively coerced – by another person, or by force of circumstances), introjected (in which the individual acts to please others or to avoid blame – regardless of whether he or she personally values this particular course of action) and internal. External and introjected are relatively controlled forms of motivation. Internal motivation is relatively autonomous and is comprised of intrinsic (when the individual enjoys activity in itself), identified (when the person's behavior reflects conscious valuing of self-selected goals and activities) and integrated (when the person's actions are shaped based on his or her own system of values, goals, and identities).

Table 1 below shows the vignettes for decision making on small purchases, where the first vignette reflects external, the second – introjected and the third – internal motivation. Please note that vignettes were

⁴ Please refer to Appendix for the full list of questions.

selected depending on the gender of participants: men listened to vignettes with male characters, and women – with female. The respondents are asked if they are completely similar, somewhat similar, somewhat different or completely different from individuals described. The highest score (4) is assigned for being completely similar, the lowest (1) for being completely different. We calculate the average of scores in each motivational category (external, introjected and internal) and construct RAI as the weighted sum of the scores of the three subscales, with the external motivation subscale weighted as (-2), the introjected motivation subscale as (-1) and autonomous motivation subscale weighted as (+3). The RAI ranges from -9 to +9. Higher scores represent a more autonomous sense of motivation whereas negative scores indicate that the respondent's motivation is more controlled.

Table 1: Construction of RAI – vignette for decision making about small purchases

	Now I'm going to read you some descriptions of various people, who make decisions about small purchases. Can you tell me if you're like this person or not like this person, when you are making decisions about small purchases (food, clothing)?
RA01A	Maria buys the amount of rice she buys, because she is afraid her husband will scold her for being a spendthrift if she buys more. Albert buys the amount of feed for the farm because he is worried that he wants to avoid a fight with his wife. She would be furious if he got more feed.
RA01B	Angela bought a dress because her friends think it is in fashion now. She wants to be seen as fashionable and elegant by her friends. Dodong bought a phone because his friends praised that brand. He cares about the opinion of his friends.
RA01C	Jenny buys this amount of rice because she thinks that this is enough to make a nutritious meal for her family. Manny buys this amount of feed because he believes that this is good amount for the animals on the farm to stay strong and healthy.

Changes in empowerment through changes in perception of gender norms

Second, CCTs may have indirect impact on adolescents' empowerment through changing their perception of gender norms. These impacts are theoretically ambiguous. On the one hand, CCTs may induce gender equitable norms by increasing women's bargaining power (through channeling cash directly into women's hands). On the other hand, CCTs may perpetuate unequal gender norms, by imposing conditionalities that generally becomes a purview of women and are aligned with traditional gender roles (such as women taking care of the children).

We first check whether exposure to the 4P program in adolescence is associated with differences in adhered to and perceived gender norms. We capture own views on gender norms and acceptability of domestic violence by asking whether respondent agrees with a gender equitable statement, justification of wife-being and justification of child-beating. Then, to capture views on norms by respondent's community we ask respondent to imagine that we invited 10 men from her or his community and ask how many of them would agree with the same statement. We repeat the same exercise, asking the respondent to imagine 10 women from her/his community. We construct a z-score of agreement with gender-equitable statements (or statements condoning wife-beating or child beating) to capture personal attitude. We use the reported fraction of men/women who endorse gender-equitable norms, disagree with wife-being or child-beating to capture perceptions of attitudes in the community.

A change in own views on gender norms, or in perception of gender norms by others may affect such manifestations of empowerment as intra-household bargaining power. We attempt to capture it through questions on intra-household decision making and time-use, focusing on paid and unpaid activities, as well as household chores.

We asked two sets of questions to capture decision-making. First, we ask detailed questions about decision-making process in 7 domains (daily purchases, large household purchases, use of contraception, use of own earnings, use of the household budget, schooling of children and treatment of sick children)⁵. We ask the traditional question (who usually makes the decision on a domain) but nuance it by adding questions about decision-making process, asking whether there is a discussion, whether others in the family weigh respondent's opinion and whose opinion prevails in case of a disagreement. Decision-making in a specific domain may be an expression of empowerment, but may also be a burden (for example, if none in the household members wants to take responsibility). To capture this aspect of decision-making, we are asking respondents if it is important for them to make decisions in specific domain. Based on these questions, we calculate our outcome variables: fraction of domains for which respondent makes the final decision (either independently or jointly with a spouse), fraction of domains in which respondent's opinion is heard, and fraction of domains in which respondent cares about making the decision.

Our second measure of decision-making is based on vignettes about decision-making styles, based on Bernard et al. (2020). We present respondents with vignettes that capture 5 decision-making styles: Dictator, Contributor, Separate Spheres, Norms-Based, and Knowledge-Based. We ask the respondents if their family is like the family in the vignette. For example, for Dictator decision-making style, respondents

⁵ Only people with children are asked about the last two domains.

are read the following: "Antonio and Lilibeth are married. Antonio makes all the financial decisions for the family and does not seek Lilibeth's opinion because he makes all the decisions for the household⁶." We create a dummy variable which is equal to 1 if a respondent identifies with a specific decision-making style and evaluate the impact of CCTs on increasing the likelihood of identification with a specific decision-making style, relative to all other styles.

Finally, we explore the impact on time-use. We check whether exposure to CCTs has affected the fraction of time spent on chores, paid work and unpaid work. The impacts for men and women may be in different directions.

Changes in exposure to gender-based violence

Third, benefitting from CCTs in adolescence may have long-lasting impacts on extreme form of disempowerment, or exposure to GBV through change in behavior/relationship of parents. Buller et al. (2019) demonstrate in a review of literature on the relationship between cash transfers and GBV that CCTs are generally associated with reduction of GBV. Exposure to violence in childhood (either direct exposure or by witnessing violence perpetrated or suffered by parents) is an important predictor of exposure to and/or perpetration of GBV in adulthood (Garcia-Moreno et al. 2006). Thus, the 4P program could reduce incidence of GBV among our sample through lowering GBV their mothers were exposed to decade ago.

We collect data on intimate partner violence (IPV), domestic violence perpetrated by other household members⁷, and violence outside home⁸⁹. For each type of violence, we construct indicators of physical, emotional, sexual and economic violence, with questions largely borrowed from Demographic and Health Surveys (DHS). For IPV, we collect data on prevalence and incidence during the 12 months preceding the survey. For domestic violence and violence outside home, we only administer questions about life-time prevalence.

⁶ The full list of vignettes is included in the Appendix.

⁷ Our questionnaire included the following options for household members: (i) Mother/Step-mother/Mother-in-law; (ii) Father/step-father/father-in-law; (iii) sister/brother/sister-in-law/brother-in-law; (iv) Other relative (please specify).

⁸ Our questionnaire included the following options for non-household members: (i) Employer/someone at workplace; (ii) Community leader; (iii) Health worker; (iv) Teacher; (v) Police/soldier.

⁹ This study's human subjects' protection protocols received research ethics review approval in accordance with the requirements of the US Code of Federal Regulations for the Protection of Human Subjects (45 CFR 46 & 45 CFR 46.110) from HML IRB.

6. Estimation

The study design and all analysis methods were pre-registered before data analysis began¹⁰. We rely on randomized rollout at *barangay* level and calculate the intent to treat effect using the following estimating equation:

$$y_{i,b} = \beta_0 + \beta_1 Treat_b + \mathbf{X}_i \boldsymbol{\alpha} + \sum_j^n \theta_j X_{ij} Treat_b + \varepsilon_{i,b} \quad (\text{Eq.1})$$

where y_{ib} is the outcome variable measured for individual i who was residing in barangay b during adolescence. $Treat_b$ is a dummy variable taking value 1 if the barangay where the respondent resided as an adolescent was randomly assigned to treatment during initial roll-out of the 4P program. X_i is the row vector of n demeaned control variables: age, gender, dummies for respondent's religion and ethnicity, dummy variable for treatment LR assignment among women, years of schooling in 2008, and a dummy variable for being enrolled in school in 2008. Following Lin (2013) we also include a full set of control variables interacted with treatment indicator. Table 2 provides summary statistics on the control variables. ε_{ib} is the error term. β_1 is the main coefficient of interest. It will measure the intent-to-treat (ITT) effect of residing in a barangay that was randomly assigned treatment, compared to the control group. We cluster standard errors at barangay of residence level. We carry out estimation of gender-specific impacts by splitting the sample into women and men and running regression (1) separately on these two subsamples. As we are estimating a large-scale governmental program, ITT is our primary estimand of interest. The discussion of results is focused on ITT effects; however, we also estimate the treatment on the treated (ToT) effect by instrumenting participation in the 4P program with the random assignment to the treatment group at barangay level. We present the results in the Appendix.

7. Results

7.1. Balance

We take advantage of the data collected in 2008 for an impact evaluation of the 4P program to test whether randomization was carried out successfully. We can identify households in our sample in these data, which includes a rich set of characteristics of households where our respondents resided in 2008, including characteristics of the household head, spouse of the household head, access to social services, housing conditions and asset ownership. Of 48 available indicators, we find statistically significant differences in only three. 4P beneficiary households are less likely to be headed by a person with no

¹⁰ <https://ridie.3ieimpact.org/index.php?r=search/detailView&id=923>

education, less likely to use a community coop and to have shared water source (Table 3). Our results do not qualitatively change if we include these variables in baseline controls.

7.2. Attrition

Of originally selected 886 households, 450 were replaced either because the household barangay officials were not able to identify the household among current barangay residents, or because the field team could not confirm that a child of eligible age and gender was a member of approached household. We do not find difference in the likelihood to be lost to follow-up by treatment status (estimated coefficient is -0.016, with standard error of 0.044).

We also check whether attrition is affected by baseline characteristics in the control group, as well as for evidence of differential attrition by observable characteristics between treatment status. Specifically, we first estimate the probability of receiving the 4P transfer based on the wide range of 2008 characteristics and construct a propensity score. We then check for differential attrition in the following regression:

$$A_{i,b} = \xi_0 + \xi_1 Treat_b + \xi_2 P_i + \xi_3 Treat_b P_i + \xi_{i,b} \quad (\text{Eq.2})$$

where $A_{i,b}$ is a binary variable equal to 1 if individual i was not tracked, $Treat_b$ is a barangay level indicator of treatment and P_i is the value of propensity score from a regression of treatment status on 2008 household characteristics – we include all variables reported in Table 3. Coefficient ξ_2 captures whether attrition differs depending on the baseline characteristics which capture the likelihood of participation in the 4P program among the control group. Coefficient ξ_3 captures whether attrition differs by observable characteristics between our treatment and control group. Table 4 shows that none of the coefficients is significantly different from 0.

Even though we do not find evidence of differential attrition, given that the fraction of households that we were unable to track is high, we report Lee bounds (Lee, 2009) and Kling-Liebman bounds (Kling and Liebman, 2004) for all the outcomes.

7.3. Main effects

7.3.1. Impacts on primary outcomes

As the core objective of CCT programs is to reduce inter-generational transmission of poverty through investment in human capital, we first focus on impacts on education and labor market outcomes. We find no evidence on long-term impacts of the 4P program on completion of compulsory education, years of

education, and enrollment in non-compulsory education for women and men¹¹ (Table 5). Similarly, regressions of labor force participation, employment, paid employment, and underemployment rates yield null results for both genders. Perhaps, not surprisingly, we also do not find impacts on proxies of welfare (Table 6). We use three alternative indices of housing conditions and assets: average of z-scores, following Katz, King and Liebman (2007), the first component of PCA and ICW. None of them yields a significant impact.

7.3.2. Impacts on empowerment

Although we do not find impacts on primary outcomes of the program, our results suggest that it increased age at marriage for women by almost a year and the age at first birth by almost half a year (Table 7). We do not find any impact on the prevalence of transactional sex.

While we can interpret increase in age at marriage and first birth as a manifestation of women's empowerment, we do not find strong evidence of increase in empowerment when using other measures. We estimate zero impact on generalized self-efficacy index, either raw value or z-score, or RAI for both women and men (Table 7).

We also do not find impact on professed as well as perceived gender norms for either women or men. The estimates in regressions of index of agreement with gender equitable statements, of agreement with child-beating, or estimated percentages of men and women in the community who would agree with gender equitable statements, wife-beating or child beating are very low and insignificant (Table 8).

However, we do find impacts on some measures of empowerment within the household for women (Table 9). The fraction of decision-making domains in which respondent's opinion is heard and the fraction of domains in which respondent cares about making decisions increases by 3.6 and 4.1 percentage points, respectively. This is a rather small increase from respective averages of 0.728 and 0.762. We do not find evidence of change when using alternative measures of decision-making within households, based on vignettes: exposure to the 4P program in adolescence is not associated with a change in likelihood to identify with a specific decision-making style. Lastly, we find zero impacts on time use (Table 10).

7.3.3. Impacts on GBV

We collected rich data on physical, emotional, social and economic gender-based violence, inflicted by three types of perpetrators: intimate partner, other members of the household or non-household

¹¹ We carry out estimation on the pooled sample of men and women; however, for succinctness of the narrative present them only when they are significant while impacts on either gender is not.

member. Our estimates suggest no impacts on physical, sexual or emotional violence perpetrated by intimate partner. However, we find that exposure to the 4P program is associated with 6 percentage points increase in prevalence, and 8 percentage points increase in incidence of economic IPV. These increases are significant at 10 and 5 percent level, respectively, and are large in magnitude: average prevalence of economic IPV is 14 percent in the control group (Table 11).

The indicator of economic intimate partner violence is constructed as a dummy variable equal to 1 if a respondent answered yes to one of the three questions about her partner (in the last 12 months for incidence, and ever for prevalence):

- Did your partner not allow you to engage in legitimate work nor practice your profession?
- Did your partner deprive you or threaten to deprive you of financial resources and the right to use and enjoy, conjugal or commonly owned property?
- Did your partner control your own money or properties or force you to work?
- Did your partner control your personal property, pets or belongings, household property, or threaten to do so?

We checked which of these four behaviors drives increase in economic IPV. Regressions on corresponding binary variables suggest that we find significant impact on economic IPV due to increase in prohibiting to allow respondents to engage in legitimate work or practice profession. It is significant at 5% level at 0.337. Coefficients in other regressions are close to zero and not significant at conventional levels. Combined with weak evidence of some increase in empowerment, the observed change in economic IPV indicator may suggest a story of more empowered women, interested in outside of home employment, facing backlash.

We do not find impacts on prevalence of GBV perpetrated by other household members or someone outside home.

7.3.4. Robustness

Given high attrition rates, we checked robustness of our results to different assumptions about attritors. Specifically, for each estimate we constructed Lee bounds, with and without control variables (Lee 2009), as well as Kling and Liebman bounds at 0.1 and 0.25 standard deviations (Kling and Liebman, 2004). Table 12 presents these estimates for outcomes where we found significant impacts.

Only the impact on age at marriage is highly robust to different assumptions about attrition: the estimates for both Kling-Liebman bounds and Lee bounds remain positive and significant. For the age at

first birth, Kling-Liebman bounds remain positive and significant when estimated at 0.1 difference, and for the upper Lee bounds. However, coefficients lose significance in other specifications. Other outcomes are not robust to changes in assumptions about attrition: for each of them, ITT impacts are bounded within an interval of different signs, both of which are sometimes significant.

7.3.5. *ToT Estimates*

Our paper contributes to the existing literature by providing experimental estimates of absolute long-term impacts of a large-scale governmental program, thus complementing the work of Baird, McIntosh and Ozler (2019), who also provide experimental estimates of absolute impacts, but of a smaller program run by an NGO. Thus, ITT are of primary interest. However, we also carry out ToT estimation by instrumenting household level participation in the 4P program with its barangay level assignment. We present tables with ToT estimates in the Appendix. Generally, our ToT estimates are similar to ITT estimates, and somewhat higher in magnitude. The impacts on age at marriage and age at first birth increase to 2.4 and 1 years, respectively. Similarly, the impacts on incidence and prevalence of economic IPV increase to 17 and 20 percentage points, respectively. Empowerment coefficients (increase in the fractions of decision domains where respondents feel heard and which respondents care about) lose their significance. However, we find small increase in the likelihood to be enrolled in non-formal school for women.

8. Deviations from the Pre-Analysis Plan

There are several deviations from the pre-analysis plan in the current analysis. First, we did not include two types of outcomes mentioned in the pre-analysis plan: empowerment within community and subjective wellbeing. When writing the paper, we felt a lack of strong theoretical basis for inclusion of these outcomes. However, we present them in the Appendix.

We structured the questionnaire to capture impact of the CCT after accounting for his/her partner's exposure to CCTs during adolescence. By design, this estimation would only include married/partnered individuals. Unfortunately, we have very limited variation in the sample for women: partners of only 25 women were exposed to CCTs in adolescence: less than 10 percent of the sample of partnered women with non-missing information about partner's exposure to CCTs.

We did not discuss explicitly in the PAP whether we are using only FTFDQ data as our primary specification, or pool FTFDQ and ACASI. We chose to pool outcomes yielded by the two methods, given lack of significant differences across methods, and in order to increase sample size.

9. Conclusions

Despite recent proliferation of studies of long-term impacts of the CCT programs, the jury is still out on their ability to reduce inter-generational transmission of poverty through human capital accumulation. Parker and Vogl (2018), as well as Kugler and Rojas (2018) find strong evidence of impacts of *Progresa* on labor force participation, as well as quality of employment. Parker and Vogl (2018) also explore changes in household wealth and find impacts driven by increased labor market incomes and higher likelihood of marrying more educated, and hence higher earning, partners. Both studies rely on non-experimental method. Baird, McIntosh and Özler (2019) and Araujo et al. (2016) find more sobering results for a CCT experiment in Malawi and *Bono de Desarrollo Humano* in Ecuador, using non-experimental framework: no impacts on labor market outcomes.

Our study attempts to contribute to this literature by evaluating impacts of approximately 1.5 years exposure to the 4P program in the Philippines during adolescence on adult outcomes in education, labor market, proxies of wealth and empowerment. Notably, there is evidence that additional 1.5-2 years of exposure to a CCT program may significantly improve education and labor market outcomes, compared to beneficiaries enrolled for a shorter time (Behrman, Parker and Todd, 2009; Barham, Macours and Maluccio, 2016). Can just 1.5 years of exposure to a CCT program generate a lasting impact in adulthood? Findings of Baird, McIntosh and Özler (2019) suggest no on average, and only on a limited set of outcomes for most at risk beneficiaries: school dropouts at baseline experience positive impacts on marriage and fertility.

Our results are somewhat more optimistic: we find that the program delayed marriage and fertility for all female beneficiaries, and this finding is robust to attrition bounds. However, similarly to Berk, McIntosh and Özler (2019) and Araujo et al. (2016) we do not find impacts on education or labor market outcomes (school completion, number of years of education, likelihood to be enrolled in non-formal education, labor force participation, employment, and employment in paid work). Nor we find impacts on the proxies of wealth.

Although we collect a battery of empowerment measures, we do not find robust impacts. Evidence of increase in intra-household decision-making among women does not withstand tests of robustness to differential assumptions about respondents we lost due to attrition: both Lee and Kling and Liebman bounds on estimates are wide, and of different sign.

Concerningly, we find that participation in the program is associated with the higher likelihood of economic IPV; however, increase in this indicator is fully driven by increase in perception that respondent's partner does not allow her to engage in legitimate work or practice her profession. This effect is compatible with increase in women's willingness to work, without any change in their partner's attitude to women working. These estimates also are not robust to corrections for attrition using Lee and King and Liebman bounds.

Our work contributes to the literature on long-term impacts of the CCT programs by providing experimental evidence on absolute long-term impacts of short exposure in Asia. We interpret our results as sobering and cautiously optimistic at the same time. On the one hand, while a number of papers demonstrate strong impacts of *additional* exposure of 1-1.5 years to a CCT program (REFERENCES), we demonstrate that *just* 1.5 year of exposure is not sufficient to move the needle on the core objectives of most CCT programs: education, translated into superior labor market outcomes and poverty reduction. At the same time, our results suggest that just 1.5 years of cash transfer support during critical transition from adolescent to adulthood may trigger delays in marriage and fertility for girls: we find evidence of delay in marriage by one year, and delay in the first birth by approximately half a year. We do not find impacts on any other indicator of women's empowerment, though.

Overall, we interpret our results as suggestive evidence of the *potential* of the 4P program to affect long-term outcomes, but recognize the need to complement this work with study of longer exposure to the program, perhaps, employing non-experimental methods as in Parker and Vogl (2018) and Kugler and Rojas (2018).

References

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.

Araujo, M. C., Bosch, M., & Schady, N. (2016). *Can Cash Transfers Help Households Escape an Inter-Generational Poverty Trap?* (No. w22670). National Bureau of Economic Research.

Baez, J. E., & Camacho, A. (2011). *Assessing the long-term effects of conditional cash transfers on human capital: evidence from Colombia*. The World Bank.

Baird, Sarah Jane; McIntosh, Craig; Ozler, Berk. 2019. "When the Money Runs Out: Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?" *Journal of Development Economics*, Vol. 140, Sep. 2019, pp. 169-185

Barham, T., Macours, K., & Maluccio, J. A. (2018). *Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings After 10 Years*. CEPR Discussion Paper 11937

Behrman, J. R., Parker, S. W., & Todd, P. E. (2009). Medium-Term Impacts of the Oportunidades Conditional Cash-Transfer Program on Rural Youth in Mexico. The MIT Press; Chapter 7 in *Poverty, Inequality, and Policy in Latin America*. 2009.

Behrman, J. R., Parker, S. W., & Todd, P. E. (2011). Do conditional cash transfers for schooling generate lasting benefits? A five-year followup of PROGRESA/Oportunidades. *Journal of Human Resources*, 46(1), 93-122.

Bernard, Tanguy, Cheryl Doss, Melissa Hidrobo, Jessica B. Hoel, Caitlin Kieran. (2020). Ask me why: Using vignettes to understand patterns of intrahousehold decision making in rural Senegal. *World Development*, forthcoming.

Buller, A. M. B., Ranganathan, M., Bleile, A., Hidrobo, M., & Heise, L. (2018). A mixed-method review of cash transfers and intimate partner violence in low and middle-income countries. *World Bank Research Observer*.

Fiszbein, A., & Schady, N. R. (2009). *Conditional cash transfers: reducing present and future poverty*. The World Bank.

Ham, A., & Michelson, H. C. (2018). Does the form of delivering incentives in conditional cash transfers matter over a decade later?. *Journal of Development Economics*, 134, 96-108.

Jerusalem, M., & Schwarzer, R. (1995). Generalized self-efficacy scale. *Measures in health psychology: A user's portfolio. Causal and control beliefs*, 35-37.

Kling, J. R., & Liebman, J. B. (2004). *Experimental analysis of neighborhood effects on youth*. John F. Kennedy School of Government, Harvard University.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1), 83-119.

Kugler, A. D., & Rojas, I. (2018). *Do CCTs improve employment and earnings in the very long-term? Evidence from Mexico* (No. w24248). National Bureau of Economic Research.

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

Parker, S. W., & Vogl, T. (2018). *Do conditional cash transfers improve economic outcomes in the next generation? Evidence from Mexico* (No. w24303). National Bureau of Economic Research.

Rodríguez-Oreggia, Eduardo and Samuel Freije (2012). Long Term Impact of a Cash- Transfers Program on Labor Outcomes of the Rural Youth in Mexico. CID Working Paper No. 230.

Saavedra, J. E., & García, S. (2012). Impacts of conditional cash transfer programs on educational outcomes in developing countries: a meta-analysis. *RAND Labor and Population Working Paper Series, WR-921-1*.

Vaz, A., Pratley, P., & Alkire, S. (2016). Measuring women's autonomy in Chad using the Relative Autonomy Index. *Feminist Economics*, 22(1), 264-294

Table 2: Summary statistics for control variables

	All mean	All SD	Control mean	Control SD	Treatment mean	Treatment SD	Diff. (Control - Treatment)	P-val
Age of respondent	24.3858	0.635	24.4419	0.634	24.3287	0.630	0.1132***	0.008
Female respondent	0.5017	0.500	0.5011	0.501	0.5023	0.501	-0.0012	0.972
Treatment LR assignment among women	0.2560	0.437	0.2437	0.430	0.2685	0.444	-0.0248	0.403
Years of schooling in 2008	6.6831	2.097	6.6902	2.082	6.6759	2.114	0.0143	0.920
Attending school in 2008	0.7486	0.434	0.7494	0.434	0.7477	0.435	0.0017	0.953
Muslim	0.0023	0.048	0.0023	0.048	0.0023	0.048	-0.0000	0.991
Protestant	0.0287	0.167	0.0319	0.176	0.0255	0.158	0.0064	0.570
Catholic	0.8117	0.391	0.7859	0.411	0.8380	0.369	-0.0521**	0.049
Iglesia	0.0080	0.089	0.0068	0.082	0.0093	0.096	-0.0024	0.689
Aglipay	0.0161	0.126	0.0296	0.170	0.0023	0.048	0.0273***	0.001
Other Christian	0.0631	0.243	0.0752	0.264	0.0509	0.220	0.0242	0.141
Adventist	0.0115	0.107	0.0046	0.067	0.0185	0.135	-0.0140*	0.054
Tagalog	0.0023	0.048	0.0023	0.048	0.0023	0.048	-0.0000	0.991
Cebuano	0.1917	0.394	0.2118	0.409	0.1713	0.377	0.0405	0.129
Ilonggo	0.0092	0.095	0.0091	0.095	0.0093	0.096	-0.0001	0.982
Bisaya	0.7853	0.411	0.7585	0.428	0.8125	0.391	-0.0540*	0.052
Hiligaynon	0.0011	0.034	0.0023	0.048	0.0000	0.000	0.0023	0.318
Number of observations	871		439		432			

Table 3: Balance Tests based on 2008 Data

	Non- 4P obs.	Mean for non- 4P	4P obs.	Mean for 4P	Diff. (non-4P - 4P)	P-val
Household head female	465	0.1097	464	0.0927	0.017	0.391
Number of household members	465	6.3441	464	6.3987	-0.055	0.673
Number of children in the household (0-14 years)	465	2.9527	464	2.9806	-0.028	0.784
Number of adults in the household (15-64 years)	465	3.3011	464	3.3513	-0.050	0.581
Number of children in the household (65 years and above)	465	0.0903	464	0.0668	0.024	0.237
Household dependency ratio	461	1.0925	461	1.1146	-0.022	0.678
Household head no education	465	0.0452	464	0.0237	0.021*	0.073
Household head attended elementary school	465	0.4817	464	0.4375	0.044	0.177
Household head finished elementary school	465	0.2559	464	0.2823	-0.026	0.365
Household head attended high school	465	0.0753	464	0.0948	-0.020	0.286
Household head finished high school	465	0.0903	464	0.1056	-0.015	0.434
Household head attended college	465	0.0344	464	0.0366	-0.002	0.855
Household head finished college	465	0.0194	464	0.0129	0.006	0.438
Spouse of household head no education	424	0.0189	417	0.0168	0.002	0.820
Spouse of household head attended elementary school	424	0.3726	417	0.3549	0.018	0.594
Spouse of household head finished elementary school	424	0.3443	417	0.3237	0.021	0.527
Spouse of household head attended high school	424	0.1108	417	0.1439	-0.033	0.151
Spouse of household head finished high school	424	0.0920	417	0.1271	-0.035	0.103
Spouse of household head attended college	424	0.0401	417	0.0240	0.016	0.185
Spouse of household head finished college	424	0.0212	417	0.0168	0.004	0.638
Household uses PhilHealth	465	0.3613	464	0.3405	0.021	0.508
Household uses HMO	465	0.0043	464	0.0043	-0.000	0.998
Household uses PHIC	465	0.0065	464	0.0043	0.002	0.656
Household uses community coop	465	0.0086	464	0	0.009**	0.045
Roof made of strong materials	465	0.3075	464	0.2909	0.017	0.582
Outer walls made of strong materials	465	0.1613	464	0.1789	-0.018	0.476
Roof made of light materials	465	0.5204	464	0.5388	-0.018	0.575
Outer walls made of light materials	465	0.4624	464	0.4698	-0.007	0.820

Household owns dwelling	465	0.2817	464	0.2500	0.032	0.274
Household rents dwelling	465	0.0129	464	0.0216	-0.009	0.312
Dwelling has toilet	465	0.3742	464	0.3427	0.032	0.317
Dwelling has shared water source	465	0.2753	464	0.2004	0.075***	0.007
Dwelling has own water source	465	0.0430	464	0.0366	0.006	0.620
Dwelling has electricity	465	0.4344	464	0.4289	0.006	0.865
Household has television set	465	0.2366	464	0.2565	-0.020	0.482
Household has VHS/VCD/DVD	465	0.1247	464	0.1034	0.021	0.308
Household has stereo/CD player	465	0.1763	464	0.1961	-0.020	0.439
Household has refrigerator/freezer	465	0.0839	464	0.0776	0.006	0.725
Household has washing machine	465	0.0237	464	0.0151	0.009	0.344
Household has air conditioner	465	0.0043	464	0.0086	-0.004	0.412
Household has living room	465	0.2194	464	0.2306	-0.011	0.682
Household has dining set	465	0.3118	464	0.3427	-0.031	0.317
Household has car or jeep	465	0.0086	464	0.0043	0.004	0.415
Household has telephone/cellphone	465	0.0946	464	0.0884	0.006	0.741
Household has a PC	465	0.0043	464	0.0022	0.002	0.565
Household has microwave	465	0.0086	464	0.0065	0.002	0.707
Household has motorcycle	465	0.0452	464	0.0517	-0.007	0.642

* p<0.1 ** p<0.05 *** p<0.01

Table 4: Attrition

	(1)	(2)	(3)	(4)
4P Assignment	-0.07124 (0.683)	-0.17888 (0.317)	-0.09383 (0.587)	-0.22338 (0.206)
4P Propensity Score	0.07897 (0.719)	-0.12860 (0.575)	0.06994 -0.75	-0.16800 (0.451)
4P Propensity Score X 4P Assignment	0.16151 (0.615)	0.38872 (0.246)	0.20688 (0.516)	0.47874 (0.149)
Observations	886	886	886	886
Characteristics of spouse of HH head (values for missing spouses coded as zero)	Yes	Yes	No	No
Dummy for presence of spouse of HH head	Yes	No	Yes	No
F-test of all vars (p-val)	0.677	0.609	0.590	0.455
P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01				

Table 5: long term impacts of the 4P program on education and labor market

	Completed schooling	Years of education	Enrolled in non- formal school	Employment	Labor force participation	Paid employment	Under- employment
Women							
ITT	-0.006 (0.87)	0.266 (0.40)	0.051 (0.16)	-0.050 (0.32)	-0.031 (0.51)	-0.007 (0.89)	-0.068 (0.42)
Observations	437	436	437	437	437	148	148
Adj. R2	0.33	0.39	0.02	0.00	0.01	0.07	-0.02
Control mean	0.645	10.714	0.132	0.364	0.418	0.910	0.308
Men							
ITT	0.021 (0.65)	0.238 (0.59)	-0.007 (0.89)	-0.040 (0.28)	-0.029 (0.39)	0.018 (0.56)	-0.073 (0.15)
Observations	435	435	435	435	435	317	317
Adj. R2	0.32	0.47	0.04	0.00	0.01	0.07	0.06
Control mean	0.523	9.464	0.241	0.764	0.864	0.889	0.389

P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01

Table 6: long term impacts of the 4P program on assets

	Housing conditions indices			Asset indices		
	KKL	PCA	ICW	KKL	PCA	ICW
Women						
ITT	-0.062 (0.21)	-0.339 (0.13)	-0.290 (0.46)	0.001 (0.98)	-0.275 (0.31)	0.159 (0.81)
Observations	437	348	348	437	437	437
Adj. R2	0.11	0.08	0.05	0.09	0.12	0.02
Control mean	-0.035	0.059	0.951	-0.039	-0.315	-0.653
Men						
ITT	-0.052 (0.30)	-0.182 (0.44)	-0.332 (0.43)	-0.016 (0.57)	-0.361 (0.15)	-0.434 (0.51)
Observations	435	337	337	435	435	435
Adj. R2	0.11	0.15	0.06	0.03	0.12	0.02
Control mean	-0.055	-0.138	0.529	-0.066	-0.614	-1.011

Regression results from equation 1; with demeaned covariates and interaction of full set of covariates with treatment variable. KKL denotes average of z-scores as in Katz, Kling and Liebman (2007); PCA is the first component of PCA; and ICW applies inverse covariate weighting. P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01

Table 7: long-term impacts of the 4P program on empowerment - socio-economic proxies and direct measures - ITT

	Age at Marriage	Age at First Birth	Prevalence of transactional sex	Generalized self efficacy (raw values 8-32)	Generalized self-efficacy (KKL zscore avg)	Relative Autonomy Index
Women						
ITT	0.933** (0.01)	0.457** (0.05)	-0.004 (0.74)	-0.118 (0.84)	-0.010 (0.87)	-0.044 (0.84)
Observations	176	299	437	436	436	428
Adj. R2	0.04	0.00	-0.00	-0.00	-0.00	0.05
Control mean	20.650	20.138	0.023	22.470	0.053	3.072
Men						
ITT	0.260 (0.56)			0.367 (0.54)	0.039 (0.54)	-0.179 (0.52)
Observations	89			435	435	408
Adj. R2	0.09			0.04	0.04	-0.01
Control mean	21.694			22.350	0.039	2.076
P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01						

Table 8: long-term impacts of the 4P program on gender and GBV/DV norms - ITT

Women								
Agreement with gender equitable statements (KKL z-score average)	Pc of men who agree with a gender equitable statement	Pc of women who agree with a gender equitable statement	Pc of men who disagree with wife-beating	Pc of women who disagree with wife-beating	Disagreement with child-beating (KKL z-score average)	Pc of men who disagree with child-beating	Pc of women who disagree with child-beating	
ITT	-0.034 (0.38)	-0.039 (0.13)	-0.033 (0.20)	0.005 (0.83)	-0.000 (0.99)	-0.013 (0.89)	0.002 (0.95)	-0.016 (0.65)
Observations	436	436	436	436	436	436	436	436
Adj. R2	-0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00
Control mean	0.042	0.597	0.689	0.797	0.859	-0.037	0.581	0.638
Men								
Agreement with gender equitable statements (KKL z-score average)	Pc of men who agree with a gender equitable statement	Pc of women who agree with a gender equitable statement	Pc of men who disagree with wife-beating	Pc of women who disagree with wife-beating	Disagreement with child-beating (KKL z-score average)	Pc of men who disagree with child-beating	Pc of women who disagree with child-beating	
ITT	-0.048 (0.20)	0.030 (0.30)	-0.019 (0.54)	-0.008 (0.73)	0.006 (0.79)	-0.106 (0.22)	-0.004 (0.91)	-0.024 (0.43)
Observations	435	435	435	435	435	435	435	435
Adj. R2	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00
Control mean	0.003	0.585	0.694	0.790	0.833	0.054	0.593	0.624

P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01

Table 9: long-term impacts of the 4P program on empowerment - decision-making - ITT

	Fraction of domains for which respondent makes the decision	Fraction of domains in which respondent's opinion is heard	Fraction of domains in which respondent cares about making the decision	Dummy for dictator decision making	Dummy for contributor decision making	Dummy for separate spheres decision making	Dummy for norms-based decision making	Dummy for most-knowledge decision making
Women								
ITT	0.019 (0.45)	0.036* (0.08)	0.041** (0.03)	-0.034 (0.35)	-0.002 (0.96)	0.024 (0.65)	-0.051 (0.18)	0.011 (0.78)
Observations	437	437	437	437	437	437	437	437
Adj. R2	0.05	0.02	0.04	0.01	0.01	-0.00	0.05	0.06
Control mean	0.627	0.728	0.762	0.241	0.286	0.641	0.236	0.327
Men								
ITT	-0.004 (0.90)	-0.017 (0.52)	-0.007 (0.74)	-0.029 (0.50)	-0.046 (0.24)	-0.075 (0.19)	0.052 (0.22)	0.042 (0.39)
Observations	435	435	435	435	435	435	435	435
Adj. R2	0.03	0.03	0.03	0.02	0.01	-0.00	0.04	0.02
Control mean	0.513	0.649	0.703	0.268	0.336	0.668	0.250	0.368

P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01

Table 10: long-term impacts of the 4P program on empowerment - time-use

	Fraction of ours spent on activity; out of total 24 hours			Fraction of hours spent on activity, out of total awake hours		
	Household activity	Paid activity	Unpaid activity	Household activity	Paid activity	Unpaid activity
Women						
ITT	0.015 (0.33)	-0.013 (0.35)	-0.001 (0.80)	-0.025 (0.68)	-0.128 (0.22)	-0.004 (0.80)
Observations	436	436	436	422	422	422
Adj. R2	0.04	0.02	-0.06	-0.00	0.02	-0.06
Control mean	0.240	0.059	0.010	0.668	0.261	0.026
Men						
ITT	0.013 (0.25)	-0.002 (0.91)	0.001 (0.87)	0.082 (0.17)	-0.072 (0.27)	-0.010 (0.64)
Observations	435	435	435	418	418	418
Adj. R2	0.00	0.03	0.06	-0.02	0.02	0.10
Control mean	0.095	0.121	0.018	0.257	0.392	0.056
P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01						

Table 11: long-term impacts of the 4P program on empowerment - exposure to IPV, NPDV, and NDV - ITT

	Prevalence of physical intimate partner violence	Prevalence of sexual intimate partner violence	Prevalence of emotional intimate partner violence	Prevalence of economic intimate partner violence	Incidence of physical intimate partner violence	Incidence of sexual intimate partner violence	Incidence of emotional intimate partner violence	Incidence of economic intimate partner violence
ITT	-0.026 (0.48)	-0.027 (0.15)	-0.001 (0.97)	0.066* (0.06)	-0.003 (0.91)	-0.013 (0.39)	0.026 (0.56)	0.080** (0.02)
Observations	437	437	437	437	437	437	437	437
Adj. R2	-0.00	0.00	-0.00	0.01	-0.00	-0.00	-0.00	0.01
Control mean	0.136	0.045	0.227	0.141	0.095	0.027	0.186	0.105
	Prevalence of physical non-partner domestic violence	Prevalence of sexual non-partner domestic violence	Prevalence of emotional non-partner domestic violence	Prevalence of economic non-partner domestic violence	Prevalence of physical non-domestic violence	Prevalence of sexual non-domestic violence	Prevalence of emotional non-domestic violence	Prevalence of economic non-domestic violence
ITT	0.034 (0.28)	-0.005 (0.31)	-0.011 (0.77)	-0.008 (0.80)	-0.013 (0.46)	-0.004 (0.57)	-0.031 (0.25)	0.000 (0.99)
Observations	437	437	437	437	437	437	437	437
Adj. R2	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00
Control mean	0.091	0.005	0.205	0.109	0.041	0.009	0.091	0.009

P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01

Table 12: robustness

	ITT	Kling-Liebman bounds				Lee bounds			
		Lower with 0.1 diff.	Lower with 0.25 diff.	Upper with 0.1 diff.	Upper with 0.25 diff.	Lower w/o controls	Lower w/ controls	Upper w/o controls	Upper w/ controls
Age at marriage	0.933**	0.6863***	0.1758**	1.3670***	1.8775***	0.8807**	1.0071***	1.2106***	1.2849***
	(0.01)	(0.000)	(0.042)	(0.000)	(0.000)	(0.018)	(0.005)	(0.001)	(0.000)
	176	583	583	583	583	131	131	130	130
Age at first birth	0.457**	0.3740***	-0.1049	1.0126***	1.4916***	0.5496	0.5496	0.9122**	0.9122**
	(0.05)	(0.000)	(0.291)	(0.000)	(0.000)	(0.188)	(0.188)	(0.016)	-0.02
	299	600	600	600	600	148	148	147	147
Fraction of domains in which respondent's opinion is heard	0.036*	-0.0518***	-0.0906***	-0.0002	0.0385***	-0.0480**	-0.0466**	-0.0147	-0.0133
	(0.08)	(0.000)	(0.000)	(0.985)	(0.001)	(0.040)	(0.040)	(0.515)	(0.541)
	437	883	883	883	883	426	426	425	425
Fraction of domains in which respondent cares about making the decision	0.041**	-0.0420***	-0.0727***	-0.0012	0.0295***	-0.0398**	-0.0387**	-0.0119	-0.0113
	(0.03)	(0.000)	(0.000)	(0.905)	(0.004)	(0.029)	(0.029)	(0.542)	(0.551)
	437	883	883	883	883	426	426	425	425
Prevalence of economic IPV	0.066*	0.0071	-0.0680***	0.1073***	0.1824***	0.0534	0.0534	0.0895*	0.0895*
	(0.06)	(0.669)	(0.000)	(0.000)	(0.000)	(0.309)	(0.309)	(0.066)	(0.066)
	437	665	665	665	665	212	212	211	211
Incidence of economic IPV	0.080**	0.0099	-0.0575***	0.0996***	0.1670***	0.0520	0.0520	0.0884*	0.0884*
	(0.02)	(0.529)	(0.001)	(0.000)	(0.000)	(0.290)	(0.290)	(0.050)	(0.050)
	437	665	665	665	665	212	212	211	211

Appendix

Questions used to construct self-efficacy index

Now, I will present to you some statements, please tell me whether you think they are true for you.

<i>Statement</i>	<i>Coding category</i>
	<i>Not at all true = 1; Hardly true = 2; Moderately true = 3; Exactly true = 4</i>
I will be able to achieve most of the goals that I have set for myself	
When facing difficult tasks, I am certain that I will accomplish them	
In general, I think that I can obtain outcomes that are important to me	
I believe I can succeed at most any endeavor to which I set my mind	
I will be able to successfully overcome many challenges.	
I am confident that I can perform effectively on many different tasks	
Compared to other people, I can do most tasks very well	
Even when things are tough, I can perform quite well.	

Vignettes on decision-making typology¹²

Now, we present to you 5 stories. Each story describes one type of couple. Please tell us whether you and your partner are similar or different to this couple.

S. No.	Story or type of couple	Response
		(1= completely similar 2= somewhat similar 3= somewhat different 4= completely different)
		(DC06)
DC06A	Antonio and Lilibeth are married. Antonio makes all the financial decisions for the family and does not seek Lilibeth's opinion because he makes all the decisions for the household.	
	(Prompt: dictator decision-making style)	
DC06B	Joshua and Jasmine are married. Joshua makes all the financial decisions for the family because he is the one who brings home income	
	(Prompt: contributor decision making style)	
DC06C	Juan and Darna are married. Juan makes all the financial decisions for the family because he makes these decisions while Darna makes other decisions for the family	
	(Prompt: separate spheres decision making style)	
DC06D	Rodrigo and Dolores are married. Rodrigo makes all the financial decisions for the family because most men in the community make these decisions	
	(Prompt: decision making according to norms)	

¹² We use the vignettes from Bernard et al. (2020)

DC06E	Fernandez and Angelica are married. Fernandez makes all the financial decisions for the family because he has the most information about budgeting and finances <i>(Prompt: decision making on basis of most knowledge)</i>
-------	--

Among the 5 couples, which type of couple do you resemble the most:

A B C D E

(DC07)

Provide letter code

ToT estimation

Table A1: long term impacts of the 4P program on education and labor market - ToT

	Completed schooling	Years of education	Enroled in non- formal school	Employment	Labor force participation	Paid employment	Underemployment
Women							
ITT	0.080 (0.50)	1.180 (0.28)	0.146* (0.09)	-0.104 (0.38)	-0.055 (0.64)	0.006 (0.96)	-0.111 (0.53)
Observations	387	387	387	387	387	130	130
Control mean	0.711	11.703	0.142	0.379	0.453	0.906	0.318
Men							
ITT	-0.025 (0.89)	-0.655 (0.71)	-0.099 (0.49)	-0.014 (0.90)	-0.004 (0.97)	0.060 (0.55)	-0.101 (0.48)
Observations	344	344	344	344	344	250	250
Control mean	0.550	9.904	0.230	0.756	0.842	0.914	0.336
P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01							

Table A2: long term impacts of the 4P program on assets - ToT

	Housing conditions indices			Asset indices		
	KKL	PCA	ICW	KKL	PCA	ICW
Women						
ITT	-0.067 (0.58)	-0.176 (0.74)	0.118 (0.90)	-0.031 (0.65)	-0.653 (0.37)	-0.786 (0.55)
Observations	387	307	307	387	387	387
Control mean	0.009	0.175	1.009	-0.002	-0.120	-0.453
Men						
ITT	-0.229 (0.12)	-0.807 (0.25)	-1.324 (0.24)	-0.021 (0.80)	-1.045 (0.19)	-0.202 (0.92)
Observations	344	265	265	344	344	344
Control mean	-0.021	0.001	0.739	-0.040	-0.368	-0.892
Regression results from equation 1; with demeaned covariates and interaction of full set of covariates with treatment variable. KKL denotes average of z-scores as in Katz, Kling and Liebman (2007); PCA is the first component of PCA; and ICW applies inverse covariate weighting. P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01						

Table A3: long-term impacts of the 4P program on empowerment - socio-economic proxies and direct measures - ToT

	Age at Marriage	Age at First Birth	Prevalence of transactional sex	Generalized self efficacy (raw values 8-32)	Generalized self-efficacy (KKL zscore avg)	Relative Autonomy Index
Women						
ITT	2.406** (0.02)	1.038* (0.06)	-0.012 (0.72)	-0.120 (0.94)	-0.009 (0.96)	0.181 (0.75)
Observations	151	265	387	386	386	379
Control mean	20.976	20.167	0.026	22.242	0.030	3.119
Men						
ITT				1.165 (0.40)	0.123 (0.41)	-0.502 (0.52)
Observations				344	344	324
Control mean				22.306	0.034	1.859
P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01						

Table A4: long-term impacts of the 4P program on gender and GBV/DV norms - ToT

Women								
	Agreement with gender equitable statements (KKL zscore avg)	Pc of men who agree with a gender equitable statement	Pc of women who agree with a gender equitable statement	Pc of men who disagree with wife-beating	Pc of women who disagree with wife-beating	Disagreement with child-beating (KKL zscore avg)	Pc of men who disagree with child-beating	Pc of women who disagree with child-beating
ITT	-0.073 (0.39)	-0.084 (0.18)	-0.084 (0.17)	0.017 (0.74)	-0.014 (0.78)	0.096 (0.67)	0.030 (0.71)	-0.013 (0.87)
Observations	386	386	386	386	386	386	386	386
Control mean	0.073	0.588	0.683	0.807	0.871	-0.004	0.602	0.657
Men								
	Agreement with gender equitable statements (KKL zscore avg)	Pc of men who agree with a gender equitable statement	Pc of women who agree with a gender equitable statement	Pc of men who disagree with wife-beating	Pc of women who disagree with wife-beating	Disagreement with child-beating (KKL zscore avg)	Pc of men who disagree with child-beating	Pc of women who disagree with child-beating
ITT	-0.038 (0.71)	0.090 (0.19)	-0.041 (0.58)	-0.022 (0.73)	0.035 (0.54)	-0.275 (0.24)	-0.015 (0.85)	-0.062 (0.43)
Observations	344	344	344	344	344	344	344	344
Control mean	-0.011	0.586	0.676	0.781	0.830	0.057	0.597	0.626

P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01

Table A5: long-term impacts of the 4P program on empowerment - decision-making - ToT

Fraction of domains for which respondent makes the decision (indep/joint)	Fraction of domains in which respondent's opinion is heard	Fraction of domains in which respondent cares about making the decision	Dummy for dictator decision making	Dummy for contributor decision making	Dummy for separate spheres decision making	Dummy for norms-based decision making	Dummy for most-knowledge decision making
Women							
ITT	0.026 (0.68)	0.053 (0.30)	0.068 (0.13)	-0.117 (0.21)	-0.045 (0.69)	0.068 (0.61)	-0.142 (0.14)
Observations	387	387	387	387	387	387	387
Control mean	0.631	0.727	0.760	0.203	0.254	0.655	0.185
Men							
ITT	0.052 (0.45)	-0.044 (0.52)	0.010 (0.86)	-0.073 (0.56)	-0.112 (0.33)	-0.164 (0.27)	-0.006 (0.96)
Observations	344	344	344	344	344	344	344
Control mean	0.501	0.658	0.702	0.263	0.340	0.632	0.292
P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01							

Table A6: long-term impacts of the 4P program on empowerment - time-use -ToT

	Fraction of ours spent on activity; out of total 24 hours			Fraction of hours spent on activity, out of total awake hours		
	Household activity	Paid activity	Unpaid activity	Household activity	Paid activity	Unpaid activity
Women						
ITT	0.004 (0.92)	-0.015 (0.64)	-0.005 (0.69)	-0.141 (0.33)	-0.218 (0.28)	-0.013 (0.72)
Observations	386	386	386	374	374	374
Control mean	0.229	0.056	0.011	0.642	0.262	0.030
Men						
ITT	0.024 (0.41)	0.030 (0.58)	-0.003 (0.90)	0.191 (0.32)	-0.079 (0.67)	-0.046 (0.52)
Observations	344	344	344	329	329	329
Control mean	0.089	0.132	0.012	0.263	0.396	0.040
P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01						

Table A7: long-term impacts of the 4P program on empowerment - exposure to IPV, NPDV, and NDV - ITT

	Prevalence of physical intimate partner violence	Prevalence of sexual intimate partner violence	Prevalence of emotional intimate partner violence	Prevalence of economic intimate partner violence	Incidence of physical intimate partner violence	Incidence of sexual intimate partner violence	Incidence of emotional intimate partner violence	Incidence of economic intimate partner violence
ITT	-0.026 (0.48)	-0.027 (0.15)	-0.001 (0.97)	0.066* (0.06)	-0.003 (0.91)	-0.013 (0.39)	0.026 (0.56)	0.080** (0.02)
Observations	437	437	437	437	437	437	437	437
Adj. R2	-0.00	0.00	-0.00	0.01	-0.00	-0.00	-0.00	0.01
Control mean	0.136	0.045	0.227	0.141	0.095	0.027	0.186	0.105
	Prevalence of physical non-partner domestic violence	Prevalence of sexual non-partner domestic violence	Prevalence of emotional non-partner domestic violence	Prevalence of economic non-partner domestic violence	Prevalence of physical non-domestic violence	Prevalence of sexual non-domestic violence	Prevalence of emotional non-domestic violence	Prevalence of economic non-domestic violence
ITT	0.034 (0.28)	-0.005 (0.31)	-0.011 (0.77)	-0.008 (0.80)	-0.013 (0.46)	-0.004 (0.57)	-0.031 (0.25)	0.000 (0.99)
Observations	437	437	437	437	437	437	437	437
Adj. R2	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00
Control mean	0.091	0.005	0.205	0.109	0.041	0.009	0.091	0.009

P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01

Additional results

Table A8: long-term impacts of the 4P program on empowerment in community - ITT

	Prox. of support network: 0=family; 1=friend/ colleague/ neighbor/ NGO worker/ religious cleric	Info. on family planning: 0=friend/ relative/ media/ landlord/ employer; 1=community worker/FDS	Info. on child health: 0=friend/ relative/ media/ landlord/ employer; 1=community worker/FDS	Info. on education: 0=friend/ relative/ media/ landlord/ employer; 1=community worker/FDS	Info. on income earning: 0=friend/ relative/ media/ landlord/ employer; 1=community worker/FDS	Info. on govt services: 0=friend/ relative/ media/ landlord/ employer; 1=community worker/FDS	Collective action (KKL z-score index)	
Women								
ITT	-0.031 (0.27)	0.012 (0.44)	0.080 (0.20)	-0.023 (0.70)	-0.023 (0.67)	-0.045 (0.23)	-0.069 (0.34)	-0.085 (0.32)
Observations	431	391	315	336	306	303	244	432
Adj. R2	0.04	-0.03	0.00	0.02	0.01	-0.01	0.00	0.01
Control mean	0.899	0.021	0.545	0.570	0.303	0.130	0.425	0.127
Men								
ITT	-0.004 (0.91)	-0.043 (0.13)	-0.029 (0.67)	0.053 (0.37)	0.001 (0.98)	-0.030 (0.11)	0.202*** (0.00)	-0.053 (0.55)
Observations	428	385	248	283	281	320	229	433
Adj. R2	0.01	0.01	0.05	0.01	-0.01	0.04	0.04	0.02
Control mean	0.898	0.107	0.309	0.270	0.216	0.055	0.179	0.024

P-values in parentheses: * p<0.1 ** p<0.05 ***
p<0.01

Table A9: long-term impacts of the 4P program on empowerment - subjective wellbeing - ITT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Patient health questionnaire	Own quality of life index (KKL z-score avg)	Family's quality of life (KKL z-score avg)	Life during adolescence QoL (KKL z-score avg)	Ladder step for own quality of life in current day	Ladder step for own quality of life; 5 years ago	Ladder step for own quality of life; 5 years ahead	Ladder step for family quality of life in current day	Ladder step for family quality of life; 5 years ago	Ladder step for family quality of life; 5 years ahead	Ladder step for family quality of life; 5 years ahead	Ladder step for quality of life as an adolescent
ITT	0.428 (0.38)	-0.068 (0.53)	-0.165 (0.12)	-0.053 (0.55)	0.093 (0.55)	0.246 (0.17)	0.056 (0.68)	0.096 (0.59)	0.190 (0.33)	0.138 (0.34)	-0.039 (0.80)
Observations	436	436	436	436	433	432	419	434	427	418	435
Adj. R2	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00
Control mean	14.429	-0.078	0.081	0.082	4.312	3.493	4.860	4.384	3.739	4.906	2.486
Gender-specific impacts											
ITT	0.133 (0.75)	0.052 (0.60)	0.149 (0.11)	-0.023 (0.78)	-0.008 (0.96)	0.152 (0.39)	0.106 (0.37)	0.115 (0.54)	-0.080 (0.69)	0.050 (0.68)	0.052 (0.73)
Observations	435	435	434.00	435	433	433	428	426	422	426	434
Adj. R2	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00
Control mean	14.305	-0.030	-0.029	-0.043	4.265	3.623	4.744	4.253	3.861	4.907	2.509

P-values in parentheses: * p<0.1 ** p<0.05 *** p<0.01