PERUVIAN ECONOMIC ASSOCIATION

Empowering women through multifaceted interventions: Long-term evidence from a double matching design

Stanislao Maldonado

Working Paper No. 170, October 2020

Empowering women through multifaceted interventions: Long-term evidence from a double matching design*

Stanislao Maldonado†

Universidad del Rosario

This version: September 30, 2020

For the latest version of this paper, click here.

Abstract

Empowering women is a policy goal that has received a lot of interest by policy-makers in the developing world in recent years, yet little is known about effective ways to promote it sustainably. Most existing interventions fail to address the multidimensional nature of empowerment. Using a double matching design to construct the sampling frame and to estimate causal effects, I evaluate the long-term impact of a multifaceted policy intervention designed to improve women's empowerment in the Atlantic region in Colombia. This intervention provided information about women's rights, soft-skills and vocational training, seed capital, and mentoring simultaneously. I find that this intervention has mixed results: improvements in incomes and other economic dimensions along with large political and social capital effects, but limited or null impacts on women's rights knowledge and control over one's body. Using a list experiment, I even find an increase in the likelihood of intra-household violence. The results highlight the importance of addressing women's empowerment multidimensional nature in policy innovations designed to foster it, incorporating men in these efforts.

Keywords: Women, empowerment, multifaceted interventions, matching.

JEL Classification Numbers: I38, J16, J24, O17.

*I thank Darwin Cortes, Lorena Trujillo and seminar audiences at the Universidad del Rosario and Universidad del Pacifico for helpful comments and advice. Data collection for this project was funded by the National Planning Department (DNP) and carried out by the Centro Nacional de Consultoria (CNC). I specially thank Cristina Querubin and Carlos Castro as well as the CNC's team for amazing field work. I acknowledge financial support from Fulbright-Colciencias and Colombia Cientifica – Alianza EFI 60185 contract FP44842- 220-2018, funded by The World Bank through the Scientific Ecosystems, managed by the Colombian Ministry of Science, Technology and Innovation (MINCIENCIAS). Juliana Aragon, Maria Paula Medina, and Julian Peña provided excellent research assistance.

†Address: Department of Economics, Calle 12C N0 4-59, Universidad del Rosario, Bogota, Colombia, E-mail: stanislao.maldonado@urosario.edu.co. Web: http://stanislaomaldonado.org/

1 Introduction

Women's empowerment is a critical component of economic development, yet very uneven progress has been observed in this regard (Doepke and Tertilt, 2011 and Jayachandran, 2015).[1] Although women's empowerment is a relevant policy goal now everywhere, it is in the developing world where progress towards this goal has been very slow (Duflo, 2012 and World Bank, 2012). Female participation in the labor market is still low, with labor earnings typically below those of men for similar occupations (Klasen, 2019 and Jayachandran, 2020). Many women have children at a younger age due to limited access to contraceptive methods, and lack of information (Upadhyay et al., 2014). Educational opportunities are still unequal (Aslam, 2013 and Heath and Jayachandran, 2018), and violence towards women is more prevalent and acceptable due to social norms (Vyas and Watts, 2009). Gender gaps in political participation and voice are still prevalent, which has led to a large underrepresentation in political office and leadership positions (Milazzo and Goldstein, 2019). Although some progress in many economic, political and social dimensions have been documented (World Bank, 2012 and Buvinic and Furst-Nichols, 2016), there are many areas in which women still face considerable disadvantages with respect to men, despite the efforts of governments and the civil society to close the gender gaps in many of these dimensions in recent years.

The standard policy response to this dramatic situation is based on a set of interventions that aim to attack one particular barrier to women's empowerment, yet these efforts have been shown to have limited success.[2] Since many women in developing countries seem to be in a low-level empowerment equilibrium, a promising venue of improvement can be based on the design of

[1] I focus on empowerment as defined by Bandiera et al. (2020) with economic empowerment, political empowerment, and control over one's body as critical dimensions. Empowerment was introduced into economics by Sen (1999) and is related to fundamental concepts such as power and agency. Common elements in existing frameworks (Fox and Romero, 2017) include aspirations and self-efficacy as well as the behavioral dimension of action, the fact that individuals operate in environments with formal and informal behavioral constraints, and the idea that empowerment is a process and an outcome. Fox and Romero (2017) consider empowerment as a set of multidimensional outcomes to be approached by attitudes and behaviors in economic, political, social, and psychological domains. Although there are some important differences among them, these approaches are interpreted as complementary in the rest of the paper.

[2] A review paper of (mostly) unidimensional empowerment interventions by Baird and Ozler (2016) concludes that "...studies that exist generally do not point to meaningful and sustained effects on key indicators of economic empowerment." The authors covered a broad set of interventions for women's empowerment including life and vocational training, cash transfers and financial incentives, sexual and reproductive health services, information and awareness campaigns, role models, and aspirations. Notice that the authors focus on young girls, the group for which more evidence is available in the current literature.

multifaceted interventions that simultaneously target different dimensions of women's disadvantage but little is known about the effects of this kind of program.[3] Very few interventions address empowerment in a multidimensional way, and there is limited evidence on their impacts, either in the short or long-term. Moreover, when such interventions exist, they are typically implemented without rigorous evaluation designs. Therefore, whether a multifaceted policy intervention can sustainably break this low-level empowerment equilibrium remains largely unanswered.

In this paper, I study the effects of "Transformate Tu Mujer" (TTM), a multifaceted empowerment intervention implemented by the Government of the Atlantico in the Colombian Caribbean, on a large set of economic, political, and social outcomes. This program was designed to provide training on soft-skills (including training modules on women's rights and self-esteem), vocational training with an emphasis on technical skills (including the development of a business plan with experts' support), mentoring, and in-kind seed capital to entrepreneur women older than 18 years old with a business or business idea. I evaluate the long-term impacts of this program 4 and 5 years after the end of its implementation.

Studying the effects of this multifaceted intervention faces a non-trivial set of methodological challenges. As it is common in many retrospective evaluation designs, baseline data were not collected before starting the intervention in 2012. Therefore, this study's fundamental methodological challenge consists of constructing a valid control group to identify the counterfactual state for program's participants. I deal with this challenge by exploiting pre-treatment data recovered from the National System of Social Programs' Beneficiaries (SISBEN). I secured access to socio-economic information of more than 2.5 million individuals in the areas of the program's coverage. Using the program's administrative records, I identify a sub-group of the program's beneficiaries that are also registered in SISBEN records. For this sub-group of former program participants, I recover a broad set of pre-treatment characteristics for 2011, a year before the start of the intervention. To construct the control group, I identify women of similar socio-economic characteristics as this previous group from SISBEN's records using a genetic matching algorithm (Diamond and Sekhon, 2013). This matching algorithm is used as a pre-processor tool in the sense of Ho et al. (2007). For each identified program's participant in SISBEN's records, I identify at least three individuals of

[3]Few papers have explored women's empowerment using multifaceted interventions. Recent examples include Austrian et al. (2020), Bandiera et al. (2020), Buehren et al. (2017), Elsayed and Roushdy (2017), and Prennushi and Gupta (2014).

similar pre-treatment characteristics based on the matching algorithm. This process allowed me to identify a sampling frame for the study that is based on a set of individuals that are similar in a large set of socio-economic characteristics.[4]

Using this sampling frame, a survey company carried out the fieldwork for collecting a household survey with information on a large number of empowerment outcomes and related information in March 2017. This company followed all standard statistical procedures to collect this survey but using a sampling frame designed to ensure that treated and comparison units are similar ex-ante in a relevant set of socio-demographic dimensions. To the best of my knowledge, this strategy has not been used before in the way proposed in this paper. Using this sample, genetic matching is used again, but this time to identify a causal relationship in combination with the Abadie and Imbens (2006) and Abadie and Imbens (2011) nearest neighbor matching estimator. This combination is based on the advantages of both algorithms. On the one hand, genetic matching has been shown to have the best performance in achieving covariates balance among a large set of alternative matching and machine learning algorithms (Colson et al., 2016). On the other hand, the matching estimator proposed by Abadie and Imbens (2006) and Abadie and Imbens (2011) has the best performance among non-experimental estimators to recover experimental results (McKenzie et al., 2010).

As it is well known, matching depends critically on the "selection on observables" assumption (Imbens, 2004), also known as the conditional independence assumption (CIA). Therefore, this paper's results are susceptible to violations of this assumption. This assumption was relaxed using a sensitivity analysis proposed by Rosenbaum (2002). This is an essential element of my research design as there is literature that shows the difficulty of matching methods to recover causal effects (Arceneaux et al., 2006 and Smith and Todd, 2005). I also address the problem of multiplicity of outcomes by correcting for multiple testing using the method proposed by Benjamini and Hochberg (1995) to control for the false discovery rate (FDR).

Colombia offers an ideal environment to study these issues. Although Colombia has made important progress in terms of gender equality in education and health (World Economic Forum,

[4] A common approach in retrospective evaluations is to collect data based on descriptive statistical designs, similar to the ones used in household surveys designed to estimate a population parameter like the poverty rate. Because this study seeks to establish a causal relationship between a treatment and a set of outcomes, the corresponding statistical design should be analytical, which implies that a set of hypotheses regarding differences between two groups is being tested. Rather than using a descriptive sample to apply techniques for analytical designs as it is common, this paper builds the sampling frame to produce a sample that is ex-ante balanced. For a discussion of descriptive versus analytical designs, see Roy et al. (2016).

2019), major gender disparities remain relevant. Important gaps in labor market participation rates, unemployment, and wages prevail (ONU MUJERES, 2018), and differences in political participation and voice between men and women are still persistent (IDEA et al., 2019). Sadly, Colombia is among the Latin American countries with a higher prevalence of physical and sexual violence (Bott et al., 2019). Unsurprisingly, the incidence of poverty is higher among women (ONU MUJERES, 2020).

The TTM program has mixed effects on women's empowerment. In the case of economic empowerment, I find a positive set of impacts on economic well-being indicators. TTM is causally related to a 31% increase in monthly income, a 26% reduction in a multidimensional poverty indicator, and a 50% increase in the likelihood of being employed. I also document large effects on labor supply (either measured in terms of occupation, working hours, or related measures) and an increase in savings (56% for women and 31% for other household members). On the other hand, I find no impact on household assets.

In terms of political and social empowerment, the evidence is mixed. I find no changes in terms of household decision-making power, but an increase in several political participation and social capital indicators. Unfortunately, I also find an increase of 15% in the likelihood of experiencing intimate partner violence (IPV) using a list experiment. Regarding dimensions of knowledge and attitudes on women's rights, I find the program play no role in changing outcomes related to these dimensions. In sum, the program is related to valuable empowerment dimensions on the economic, political, and social sides but with more mixed evidence regarding impacts on decision-making within the household as well as attitudes and beliefs about women's rights. Even worse, the program is associated with an increase in IPV. Most of these results are robust to violations in the "selection on observables" assumption using the sensitivity analysis proposed by Rosenbaum (2002).

This paper is related to a growing literature on women's empowerment in developing countries. Existing evidence is mostly non-experimental and have been surveyed elsewhere (Duflo, 2012, Buvinic and Furst-Nichols, 2016, Buvinic and O'Donnell, 2019, and World Bank, 2012). A small yet growing experimental literature has been produced in recent years, mostly focused on the effects of multifaceted interventions on young girls. Perhaps the best example is Bandiera et al. (2020), who study the Empowerment and Livelihood for Adolescent (ELA) program in Uganda, which provides vocational and soft-skills using social clubs.[5] This program successfully increased girls' participation in economic activities, reduced early pregnancy and marriage, and raised aspirations regarding marriage, childbearing, and fertility.[6] Another example is Austrian et al. (2020), who study the AGEP program's effects in Zambia on empowerment outcomes for young girls. This program provided vocational and socio-emotional skill training using clubs, vouchers for health services, and a saving account. The intervention did not lead to relevant changes in economic assets, educational, or fertility outcomes.

Non-experimental evidence about multifaceted empowerment interventions is also scarce. Elsayed and Roushdy (2017) analyze the effect of an empowerment intervention in Egypt based on vocational, business, and life skills training (including some basic support for legal registration and opening bank accounts) on labor market and entrepreneurship outcomes. Using a matching-differences in difference design, the authors find medium-size effects of this intervention on labor outcomes. Prennushi and Gupta (2014) study the effects of an empowerment program in India that provided seed funds, access to financial services, and training in business and life skills on a broad set of assets and expenditures, human development, and women's empowerment outcomes. The authors find positive impacts on asset accumulation, increases in household expenditures, and larger investments in education, along with positive changes in women's autonomy, mobility, and participation in social activities.

This paper differs from this previous literature regarding the dimensions of empowerment addressed by TTM and the scope in terms of evaluation outcomes. Except for Prennushi and Gupta (2014), most of the existing multifaceted interventions are based on vocational and life skills training, sometimes bundled with financial services. TTM incorporates these dimensions and includes mentoring and seed capital (including inputs and equipment), which is expected to tackle multiple and interlinked constraints to women's well-being by exploiting potential complementarities between the program's components. In terms of outcomes, this paper studies a larger set of economic, social, political, and psychological outcomes, including effects on IPV. In this sense, this paper adds to the existing literature by analyzing the effects of a more complex multifaceted in-

[5]This program was implemented by the NGO BRAC and was expanded to several countries in Africa and Asia, including Liberia, Nepal, Sierra Leone, South Sudan, and Tanzania.

[6]Buehren et al. (2017) study a similar ELA program in Tanzania that included a microcredit component, and explored similar outcomes as Bandiera et al. (2020) without finding improvements in any of these outcomes but some evidence in terms of increases in savings.

tervention (compared to the ones previously studied) on a more rich set of outcomes. In this way, it also complements a related literature on the effects of multifaceted interventions for vulnerable populations.[7] Besides, by including women of different ages, this paper provides evidence on the effects of multifaceted empowerment interventions beyond young girls.

This paper is also related to literature that explores the effects of interventions that focus on partial aspects of women's empowerment regarding treatments and outcomes. These include papers that explore the effects of one dimension of women's empowerment on multidimensional or uni-dimensional outcomes, which is limited since improvements in some dimensions can be accompanied by reductions in others (Fox and Romero, 2017). For instance, a large literature has explored the effects of vocational and business training with mixed results[8], and a recent scholarship has shown that life or socio-emotional skills can have large effects on economic empowerment.[9] Some recent evidence suggests that mentoring can be a powerful yet cost-effective strategy to promote women's economic empowerment.[10] At the same time, scholars have shown mixed effects of providing capital to entrepreneurs.[11] This paper builds on the existing literature about the effects of these programs by analyzing the impacts of a bundled empowerment intervention that include all these dimensions simultaneously.

The rest of the paper is organized as follows. Section 2 provides details about the institutional

[7] A recent scholarship has explored the role of multifaceted interventions for the poor. Banerjee et al. (2015) study the effects of multifaceted programs that include the provision of capital, skills training, cash transfers, savings and educational, and health services in six countries. The authors find large effects on consumption, incomes and assets as well as sizable effects on food security, time use, and mental health. Unfortunately, these programs did not affect women's decision-making. Bandiera et al. (2017) study a similar multifaceted program for poor women in Bangladesh, finding that it caused sustained poverty reduction and asset accumulation.

[8] The evidence for vocational training suggests that these interventions have some effects on labor market outcomes but they are typically modest (McKenzie, 2017). These results do not seem to be different for women. Regarding business training, the evidence suggest that they are effective in improving business practices but their effects on profits and business survival are limited (Woodruff and McKenzie, 2012).

[9] A recent scholarship has shown that soft-skills can boost profits, employment, and crop adoption for women (and also men). See, for instance, the evidence provided by Campos et al. (2017) for Togo, Montalvao et al. (2017) for Malawi, and Croke et al. (2017) for Nigeria.

[10] Two recent papers have shown the positive impacts of mentoring on microenterprise-level outcomes. Brooks et al. (2018) study the effects of a mentorship program for inexperienced female microentrepreneurs in Kenya finding a 20% increase in profits. On the other hand, LaFortune et al. (2018) evaluate a similar intervention in Chile were personalized consulting sessions are compared to role models. These role models are former participants of a training program. They find that both interventions raised household income by 15%, being the role model more cost-effective.

[11] The evidence in this regard is mixed. de Mel et al. (2008) documents large effects of small infusions of capital on profits for male-headed small enterprises. The short-term effects are close to 6% and five years later are close to 12% (de Mel et al., 2012). The effect for female-headed small firms is indistinguishable from zero. Fafchamps et al. (2014) study a similar intervention in Ghana, finding differential effects by gender. In the case of female-owned microenterprises, only in-kind grants seem to have a positive effect, especially for businesses of larger size. However, if one focused on smaller female-headed firms, the returns to capital disbursement are close to zero.

settings, and the TTM program. Section 3 introduces the research design, and section 4 presents the empirical results. Section 5 explores additional results, and section 6 presents the results of the sensitivity analysis. Section 7 concludes by discussing the implications of this paper's findings.

2 Institutional setting

2.1 Women's empowerment in Colombia

According to the Global Gender Gap Report 2020 (World Economic Forum, 2019), Colombia ranks 22 out of 153 countries in terms of the global gender gap index, an indicator designed by the World Economic Forum to measure progress towards the gender equality goal. This good performance is driven by significant progress in terms of educational attainment and health conditions where gender gaps have disappeared. However, important gaps remain in terms of economic opportunities and political empowerment.

In terms of education, women's participation in tertiary education almost doubled in the past decade. Now, 6 out of 10 women between 17 and 21 years old are enrolled in a university or technical-level institution (ONU MUJERES, 2018). There are no gaps in terms of enrollment in primary and secondary education, and the literacy rate is close to 100% for both sexes, but some gaps persist in terms of learning as measured by standardized tests (World Economic Forum, 2019 and Sanchez, 2018).

Regarding health, almost universal and gender-equal coverage was achieved by 2017, although quality issues remain a challenge for the country (ONU MUJERES, 2018). Healthy life expectancy is high (70 years), close to the standards of developed nations like Sweeden (73.4) or Canada (74.3), and above most of its neighbors (World Economic Forum, 2019). This contrasts with women's exposure to IPV, where Colombia ranks second in a sample of Latin American countries, with a prevalence of physical violence of 17.5% and sexual violence of 3.8% (Bott et al., 2019). Even worse, 2.7 women per day were murdered in 2017 (IDEA et al., 2019).

In terms of economic empowerment, progress to gender equality has been slow and very unequal. According to official statistics, the gender gap in labor market participation is still high (21%) and has remained almost the same for the last decade (ONU MUJERES, 2020). Unemployment is higher for women (14% over 8% for men), and a woman earns 88% of a man's wage (ONU MUJERES, 2018 and Fonseca, 2019). Moreover, the fraction of women without income is 3 times the same proportion for men (ONU MUJERES, 2020). It is not surprising that the incidence of poverty among women is higher: 120.3 women are poor for every 100 men ((ONU MUJERES, 2018)). This indicator has been worsening in the last decade.

Progress in terms of political empowerment has also been limited (IDEA et al., 2019). Although 50% of ministerial cabinets are female, the fraction of women in the parliament is only 27% (World Economic Forum, 2019). In 2018, only 12.1% of elected mayors and 17.6% of elected city council members were female (ECLAC, 2020). Part of the gap is explained by the fact that few women decide to run for political office. For instance, the fraction of women running for Congress was only 34.5% (IDEA et al., 2019). Even though some progress has been observed in the past decade (World Economic Forum, 2019), Colombia's political empowerment performance is still below the regional average.

These numbers hide important heterogeneities and inequalities. Many of those indicators are worse for poor women in the Atlantico department, the typical TTM participant in the evaluation sample. For instance, in terms of economic empowerment, the participation rate, the unemployment, and the wage gap are larger in Atlantico than the national average (Fonseca, 2019).[12] This is similar in other dimensions of empowerment. Therefore, although Colombia shows areas of substantial progress in terms of women's empowerment along dimensions that required additional effort, it is crucial to have in mind that the Atlantico department is behind in terms of economic and social opportunities for women concerning the country's averages.

2.2 "Transformate Tu Mujer" program

TTM was implemented by the Women and Gender Equality Secretariat of the Government of Atlantico to improve women's empowerment along multiple dimensions. The program was designed to incorporate several components that are expected to simultaneously affect the psychological, economic, political, and social dimensions of women's empowerment by delivering of training on business and socio-emotional skills, mentoring, and in-kind seed capital.

The program was open to women of legal age with a trade. With the support of local govern-

[12] In the Atlantico department, women's participation rate is 52% versus 78% for men. The unemployment rate for women is 11% versus 5% for men, and the wage gap is 19%. See Fonseca (2019) for details.

ments, social workers were hired to set up temporary offices across the department to reach the target population and help potential participants fill out application forms. This form collected applicants' personal information and details about the business or business idea. Then, applicants were interviewed by experts in entrepreneurship to classify them according to how innovative their business ideas were and by their level of commitment to their businesses. Applicants were ranked according to these two criteria, and a final list with the selected applicants was published. The same procedure was used for the first two cohorts of the program that will be analyzed in this paper (2012 and 2013).

Program's participants were exposed to the following programs' phases:

(i) **Phase 1: Training on soft-skills, women's rights, and self-esteem**

The first phase of the intervention was designed to train participants on soft-skills, emphasizing women's rights, agency, and self-esteem. Participants received a 90-hour long training program delivered by specialized trainers of the Universidad del Norte, one of the most prestigious universities in the department. This program was composed of 5 modules that emphasized aspects such as self-control and autonomy (module 1), health and lifestyles (module 2), women's rights (module 3), empowerment and solidarity (module 4), and women as an agent of change and social transformation (module 5).

(ii) **Phase 2: Vocational training with an emphasis on technical skills and the development of a business plan with experts' support**

The second phase was designed to improve technical skills related to participants' line of business. In this phase, participants were accompanied by instructors with large experience and knowledge in the business line selected by participants during the enrollment phase. It also included the development of a business plan with the instructor's support.

This phase was composed of 4 modules, of which the last two were focused on developing a business plan and support during its implementation. The first two modules emphasized entrepreneurship and business associativity. In total, 170 hours were assigned to these phases. Women who completed the first two phases received a certification by the Universidad del Norte.

(iii) **Phase 3: Mentoring**

In this phase, participants worked with mentors in strengthening their business plan and were exposed to specialized training in accounting and communication tools.

(iv) **Phase 4: In-kind seed capital**

In this final step, the program distributes inputs and equipment to participants according to the business plan and budget developed in the second phase.

All these phases were implemented over six months. The first and second cohorts of TTM covered 6,202 participants. In this period, 623 businesses were constituted among TTM participants. The total program cost was USD 2,916,060.38 for the first two years. It was funded by the Colombian National Government through the national oil royalties system, the Government of Atlantico, the Mario Santo Domingo Foundation, and the Universidad del Norte (SMEG, 2020). The cost per participant was USD 470.[13]

As it is common in many of these interventions, it is tough to disentangle the role of specific program's components. Due to some implementation issues, some of these components faced implementation challenges and were partially delivered. In the evaluation sample, all participants reported to have completed the TTM's training component, and 96.06% reported to have graduated from the program. Regarding the training component, a large proportion of participants considered that training materials (88%), topics covered (98%), activities implemented (98%), facilitators (98%), class duration (97%), and class participation (92%) were good.[14] Hence, the training component in phases 1 and 2 were delivered to almost all participants, and their evaluation of them is positive.

Regarding phase 3 and 4, the results are mixed. A large fraction of participants worked with mentors (76%) during the development and implementation of their business plans. Most of them also received inputs and equipment for their businesses (90%), but participants' evaluation of this component is less optimistic. For instance, qualitative interviews carried out as part of TTM evaluation (CNC, 2017) suggest that participants perceived that the distribution of inputs and

[13]Own calculation based on program information. Unfortunately, detailed information about the program's cost structure is not available to perform a cost-benefit analysis. See SMEG (2020) for details.

[14]These and the following calculations in this section are based on the evaluation survey carried out by CNC in March, 2017.

equipment was insufficient.

One critical element of the program's implementation was the goal of fostering associations between participants of the same line of business. Still, this goal was not effective because participants' businesses were highly heterogeneous, and they found of little value the kind of associations fostered by the program. Therefore, it seems that these elements of TTM were less effective as perceived by participants.

3 Research design

In this section, I describe the methodological steps used to evaluate the effects of TMT and provide a discussion about the design's validity. This design is retrospective and evaluates the effects of this intervention for 2012 and 2013 cohorts, 5 and 4 years after the program's end.

The main components of the research design are the following: First, I construct the sampling frame using genetic matching as a pre-processor tool (Ho et al., 2007) based on pre-treatment data from SISBEN records; second, with the sample collected using this sampling frame, I then estimate the causal effects of TTM on empowerment outcomes using genetic matching for a second time along with the matching estimator proposed by Abadie and Imbens (2006) and Abadie and Imbens (2011); and third, I perform a sensitivity analysis a la Rosenbaum (2002) to evaluate how sensitive the results are to violations of the CIA assumption.

3.1 Matching techniques

Genetic matching is an ideal approach for the research problem at hand both for pre-processing data to construct the sampling frame and the estimation of causal effects. This matching technique is based on a genetic search algorithm that automatizes the process of checking and improving overall covariate balance (Sekhon and Mebane, 1998).[15] Genetic matching extends existing matching

[15] A genetic algorithm solves complex optimization problems using heuristics inspired in the natural selection process. A population of potential solutions is evaluated iteratively. In each iteration or generation, the fitness of every individual in the population is evaluated. The more fit individuals are selected from the current population, and each individual's genome is modified to form a new generation. The new generation of potential solutions is used in the next iteration of the genetic algorithm. See Sekhon and Mebane (1998) for details.

techniques by generalizing the Mahalanobis metric in the following way:

$$GMD(X_i, X_j, W) = \sqrt{(X_i - X_j)'(S^{-1/2})'W(S^{-1/2})(X_i - X_j)}; \tag{1}$$

where X is a matrix of covariates for units i and j, S is the sample covariance matrix, $S^{-1/2}$ is the Cholesky decomposition of S, and W is positive definite weight matrix. The genetic search algorithm chooses W such that it optimizes a pre-specified loss function to maximize balance of observed covariates across matched treated and control units. Compared to other matching and machine learning approaches, genetic matching has the best performance in achieving covariates balance (Colson et al., 2016).[16]

Once covariates balance is achieved, the next step consists of comparing outcomes across treatment and control groups in the matched sample constructed with the genetic search algorithm. This can be done in several ways. The simplest one is just to perform a simple difference in means in the matched sample but this approach is unattractive. A more powerful approach consists of using the nearest-neighbor matching estimator proposed by Abadie and Imbens (2006) and Abadie and Imbens (2011). I provide some details below and refer the reader to the aforementioned papers.

The counterfactual outcome $\hat{Y}_i(0)$ for the average treatment effect on the treated (ATT) can be written in the following way:

$$\hat{Y}_i(0) = \begin{cases} Y_i, & \text{if } T_i = 0 \\ \frac{1}{\#J_M(i)} \sum_{i \in J_M(i)} Y_l & \text{if } T_i = 1; \end{cases} \tag{2}$$

where Y_i is the outcome of interest, T_i is the treatment variable, $J_M(i)$ is the set of indices for the matches for unit i that are at least as close as the Mth match, and $\#J_M(i)$ is the number of elements of $J_M(i)$. Using this counterfactual, Abadie and Imbens (2006) and Abadie and Imbens (2011) propose a **simple matching estimator**:

[16] According to the simulations implemented by Colson et al. (2016), genetic matching has the best covariate balance among 7 alternative methods even when there is poor support in terms of the propensity score. This advantage is consistent across alternative balance metrics.

$$\tau_S^{ATT} = \frac{1}{N_1} \sum_{i:T_i=1} \Big\{ Y_i - \hat{Y}_i(0) \Big\} = \frac{1}{N_1} \sum_{i=1}^{N} \Big\{ T_i - (1 - T_i) K_M(i) \Big\} Y_i; \tag{3}$$

where $K_M(i)$ is the number of times i is used as a match for all observations l of the opposite treatment group, each time weighted by the total number of matches for observation l. Also, note that $N_1 = \sum_{i:T_i=0} K_M(i)$.

One limitation of this matching estimator is that is biased in finite samples when matching is not exact. To address this issue, Abadie and Imbens (2011) propose a bias-corrected matching estimator. This estimator adjust the difference in covariate values within the matches such that:

$$\hat{\mu}_T(x) = \mathbb{E}[Y(T)/X = x] = \hat{\beta}_{T0} + \hat{\beta}_{T1} X; \tag{4}$$

where $\hat{\mu}_T(x)$ is a consistent regression estimator of $\mu_T(x)$ for $T = 0, 1$. $\hat{\beta}_{T0}$ and $\hat{\beta}_{T1}$ are the linear parameters of this regression function.

The counterfactual outcome in this scenario is:

$$\tilde{Y}_i(0) = \begin{cases} Y_i, & \text{if } T_i = 0 \\ \frac{1}{\#J_M(i)} \sum_{i \in J_M(i)} \Big\{ Y_l + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_l) \Big\} & \text{if } T_i = 1. \end{cases} \tag{5}$$

Therefore, the **bias-corrected matching estimator** is:

$$\tau_{BC}^{ATT} = \frac{1}{N_1} \sum_{i:T_i=1} \Big\{ Y_i - \tilde{Y}_i(0) \Big\}. \tag{6}$$

In the next subsections, I describe how these matching techniques are used with the goal of pre-processing data and causal effect estimation.

3.2 Constructing the sampling frame and sample

3.2.1 Sampling frame

To construct the sampling frame, I first use all the records of the program's participants for the 2012 and 2013 cohorts. In total, I have access to 3,670 participants' records. This information was provided by TTM for this evaluation under standard protocols for data protection. I then merge this information with SISBEN records for 2011. This information was provided by the National Planning Department (DNP) for this study. I am able to identify 2,445 participants in both program and SISBEN's records. In this way, I can identify pre-treatment data for a sub-set of the program's participants. However, it is important to emphasize that these are participants that are vulnerable to poverty, so the results of this study are relevant for this type of population.

Recovering pre-treatment socioeconomic data for the program's beneficiaries helps identify non-beneficiaries with similar socioeconomic characteristics using SISBEN's records. I take advantage of 2,552,455 available records for the area of influence of this intervention. This includes 26 municipalities in the department of Atlantico, as well as the departments of Bolivar and Magdalena.[17] Figure 1 presents a map with all the municipalities included in this study.

I use a multivariate genetic matching strategy to identify three non-beneficiaries in SISBEN's records for each program's participant. As any matching design, a critical step consists of selecting the pre-treatment variables to be used to match individuals. I use a large set of household characteristics, assets, and other socioeconomic characteristics available in SISBEN's records. A full list of the variables and their definitions are included in Table S1 in the Online Appendix. I also use the SISBEN score, a welfare-index used by the Colombian government to select social programs' beneficiaries. Each household receives a score from 0 to 100 (from poorest to richest). The score is calculated using 24 variables across four dimensions: health, education, housing, and vulnerability. Given its importance, I include this variable along its square in a multivariate matching design.

As it is well-known, multivariate matching may suffer of the "dimensionality curse". One way to address this issue is by combining multivariate matching with propensity score matching. Following Rosenbaum and Rubin (1985), it is recommended to include a propensity score as one additional variable in a multivariate matching design. This would allow the design to include relevant information in the matching design without concerns about the "dimensionality curse". Given that this part of the design is descriptive, I use a simple strategy to estimate the propensity score. Table S2 in the Online Appendix presents the variables used in the estimation.

It is important to notice that this matching strategy was applied to keep the reference labor

[17] These departments were included in the sampling frame because of potential concerns of within-municipality spillovers. As mentioned later, the sample size is not large enough to exploit this feature of the sampling design in the analysis.

market constant. This implies applying this matching design in each municipality where treatment and control units were available in the Department of Atlantico. Each treated woman in a given municipality is then matched with three non-treated women in the same municipality. The final sampling frame for Atlantico is just the collection of all the matches from these municipalities.

Concerns about potential spillover effects led to the inclusion of municipalities in the departments of Bolivar and Magdalena. Due to a government's request, it was decided that a fraction of the final sample should include controls from these departments as a way to address potential within-municipality spillovers. Therefore, the procedure for these municipalities was based on applying the matching strategy using all the treated women in the department of Atlantico and match them with women of similar characteristics in Bolivar and Magdalena.[18] Due to budget constraints, the final sample size is not large enough regarding observations from these two departments to take advantage of this feature of the sampling strategy. Therefore, I do not discuss spillover effects in the rest of this paper.[19]

Table 1 presents descriptive and balance statistics for the sampling frame obtained after applying this matching strategy. Besides means and variances for the treatment and control groups, I also report the standardized difference, the variance ratio, and the p-values for t-student, Wilconxon and Fisher tests of difference in means between treatment and control groups. In general, means and variances between treated and control units are pretty similar. This is confirmed by standardized differences in means, which are-in all the cases-well below the 0.3 threshold suggested by Imbens (2015). or the 0.2 proposed by Cohen (1988).[20] Following Imai et al. (2008), I also report variance ratios to capture other moments of the empirical distribution of pre-treatment variables beyond the mean. Results are closer to 1 in most cases, which suggests that covariates are balanced in means and variances.

For the sake of completeness, I also report the p-values for the t-test with similar results, except for three unbalanced variables. Due to the results for the standardized difference in means, this

[18] One implication of this feature of my research design is the existence of repeated treated observations in the sampling frame. This characteristic is not present in the final sample. As it will be shown later, this does not affect the quality of the final sample.

[19] It is unclear whether the size of within-municipality spillovers is relevant in this setting, but some international evidence is available. For the case of young girls in Uganda, Bandiera et al. (2020) present suggestive evidence of spillovers for aspirations and control over the body indices, but weaker effects for economic empowerment. Some evidence for within-household spillovers will be discussed later.

[20] Standardized differences compare the difference in means in units of the pooled standard deviation, and it has the advantage of not being influenced by sample size. See Imbens (2015) for details.

seems just to be driven by sample size. Because t-tests are based on a particular distribution, I also use the non-parametric Wilconxon test to account that data are matched and to relax the parametric assumptions. Results do not change. Finally, I use a Fisher's exact test because many variables in my design are categorical. This test is valid regardless of the sample size. Results are similar. In sum, there is strong evidence regarding the balance of pre-treatment characteristics in the sampling frame.

A final piece of evidence on balance in the sampling frame is reported in Figure 2. As it was mentioned above, the SISBEN score is a summary measure of socioeconomic characteristics. I assess differences in distribution of this variable by treatment status, finding that the distributions are the same. This suggests, along with all the previous evidence, that the matching strategy to build the sampling frame was highly successful in achieving pre-treatment covariates balance.

3.2.2 Final sample

Using the matched sampling frame, the National Consulting Center (CNC) carried out the fieldwork to collect the final sample, using the standard statistical procedures for fieldwork and data collection. A sample of 1,738 women was finally collected with 715 women in the treatment group. Few observations were discarded after trimming observations with extreme values of a propensity score in a procedure described in the next section.[21] The final sample for the primary analyses includes 1,723 observations.

Table 2 replicates the analysis in Table 1 to check the quality of the sample.[22] In general, I find that the collected sample preserves good properties in terms of pre-treatment balance in covariates that was achieved with the matched sampling frame. I include additional pre-treatment covariates that were collected in the survey, and check whether these are balanced between treatment and control groups. Among the list of 18 covariates, only one variable is unbalanced in terms of standardized differences in means. Five variables are slightly unbalanced in terms of variance ratios (larger than 1.2), and the results for the t-student, Wilconxon, and Fisher tests suggest some imbalances but, as it was mentioned before, only in one case this is consistent with standardized

[21] Trimming of the sample did not affect the balance in terms of pre-treatment characteristics between treatment and control groups (results not shown).

[22] The full definition of these variables is the same as the ones in the sampling frame and is presented in Table S1 in the Online Appendix.

differences in means larger than 0.3 in absolute value. Therefore, the evidence presented in Table 2 is consistent with an outstanding level of balance in pre-treatment covariates between treatment and control groups.

3.3 Estimating causal effects

The sample collected in the previous step is used in this section to estimate the causal effects of interests using the combination of Abadie and Imbens (2006) and Abadie and Imbens (2011) nearest neighbor matching estimator along with genetic matching to optimize the process of finding covariates balance. The methodological steps are described below.

3.3.1 Estimating the propensity score using the Imbens and Rubin's (2015) algorithm

In the first step, I estimate a propensity score using the algorithm proposed by Imbens and Rubin (2015). This algorithm selects covariates for inclusion in the propensity score's specification by choosing a set of linear terms as well as interactions and quadratic terms based on a data-driven approach.[23] This propensity score is used as a covariate for the matching process as in Rosenbaum and Rubin (1985) and to trim observations with extreme propensity score values following Crump et al. (2009).[24] In particular, I drop observations with propensity score larger than 0.9 and smaller than 0.1. Table S3 and S4 in the Online Appendix present the definitions of variables used to estimate the propensity score and the results of this exercise applying the Imbens and Rubin (2015) algorithm, respectively.

A critical element in evaluating my research design's validity consists of checking whether the estimated propensity score fulfills the overlap condition. The evidence in Figure 3 suggests that this is the case. Therefore, this important assumption for matching designs is valid in this setting.

[23] A detailed discussion about the steps involved in this algorithm can be found in the Appendix A of Imbens (2015).

[24] One important issue is that the propensity score is a generated regressor and, as such, standard errors should account for this fact. The recommended solution for this problem is bootstrapping standard errors, but Abadie and Imbens (2008) have highlighted the problems with this alternative for matching. Abadie and Imbens (2016) study the case where matching is done exclusively on the propensity score without additional covariates-as in this paper-and show that, for the case of ATET, ignoring the estimation error has ambiguous effects on the size of confidence intervals. Because there is no clear suggestion in the technical literature about how this issue should be addressed, I proceed like the rest of the empirical literature, by ignoring this issue.

3.3.2 Genetic matching for covariate balance

The second step consists of using a multivariate genetic matching design to find balance in covariates across treatment and control units in the sample. Table S5 in the Online Appendix presents the variables used to match treated and control women. I use 11 pre-treatment covariates, including the SISBEN score and the propensity score previously estimated. Due to sample size constraints, I set the number of neighbors in 2 per treated woman. I also use a population size of 1,000 for the genetic algorithm to search for alternative weights to achieve overall balance, following a recommendation by Diamond and Sekhon (2013).[25]

Figure 4 assess pre-treatment balance before and after matching. I check for balance for a list of 26 covariates for both the full and trimmed samples. I use a standardized difference in means of 0.2 in absolute value to assess balance.[26] In general, a large set of covariates were balanced even before matching and remain in this way after. This suggests that the sample was already well balanced before using the genetic algorithm in this phase. This speaks well about the quality of the first matching process to build the sampling frame. Only three characteristics are unbalanced before applying matching in the sample. After matching, only the variable years of education is slightly above the 0.2 standard, but well below the common 0.3 standard suggested by Imbens (2015). Therefore, it can be safely concluded that matching in this phase successfully achieved balance in pre-treatment characteristics beyond the original 11 variables included in the genetic algorithm.

It is possible that this good balance is only present for differences in means but not for differences in distribution. Figure 5 assess distributional balance for the SISBEN score before and after matching. Again, there are very minor differences between treatment and control units before matching, implying a good quality of the sampling frame in terms of balance. Although by a small margin, after matching balance is improved, being the distribution of SISBEN score practically indistinguishable across treatment and control groups.

Figures S1 to S4 in the Online Appendix display, side by side, quantile-quantile plots for both the unmatched and matched samples for a set of (continuous) pre-treatment covariates in the trimmed sample. Results suggest that matching does improve distributional balance, but the gains

[25]Recall that the population size in genetic algorithms represents the set of potential solutions for the optimization problem of interest.

[26]Using this threshold is stricter than the standard proposed by Imbens (2015). I follow this approach to be more demanding with this element of my research design.

are small because these variables were already well balanced. Consistent with the results above, only years of education seems to be slightly unbalanced. Figures S5 to S10 in the Online Appendix display kernel densities for a similar set of (continuous) pre-treatment covariates before and after matching, finding again that the covariates' distribution across treatment and control are almost the same, with marginal improvements after matching. In sum, the distributions of covariates seem to be very well balanced between treatment and control women.

Once balance has been achieved, the next step consists of estimating the causal effects with any arbitrary matching technique. I use the bias-corrected nearest neighbor matching estimator proposed by Abadie and Imbens (2006) and Abadie and Imbens (2011) to compute the ATT given its better properties compared to alternative matching estimators (McKenzie et al., 2010). Standard errors are computed following the formula for the variance derived by Abadie and Imbens (2006). Due to a large set of outcomes in this study, I first need to discuss how multiple testing will be handled in this paper.

3.3.3 Addressing multiple testing

Due to women's empowerment's multidimensional nature, a large set of outcomes will be analyzed in this study. I use the correction for multiple testing proposed by Benjamini and Hochberg (1995) to control for the false discovery rate.[27] I apply this correction for each of the set of outcomes explored in this study.

3.4 Sensitivity analysis

A final piece of this paper's research design is a sensitivity analysis to check the validity of the CIA assumption. An extensive literature shows the limitations of matching to recover causal effects because of violations of this assumption (Arceneaux et al., 2006 and Smith and Todd, 2005), so it is important to relax it using a sensitivity analysis as in Rosenbaum (2002).

The basic intuition behind this approach is that two matched individuals with the same observable characteristics X should have the same probability of receiving treatment e. Therefore, the

[27]This procedure ranks the number m of hypotheses $H_{(i)}$ according to their uncorrected p-values $P_{(i)}$ and sets a proportion q of rejected null hypotheses that are erroneously rejected. Then, letting k be the largest i for which $P_{(i)} \leq \frac{i}{m}q$; the procedure rejects all $H_{(i)}, i = 1, 2, ..., k$ that violate this weak inequality. See Benjamini and Hochberg (1995) for details.

ratio of probabilities should be equal to 1 between one person in the treatment group with respect to other person in the control group. If CIA is violated, then this ratio should be higher than 1. Rosenbaum (2002) derived bounds of the following form using a logistic function:

$$\frac{1}{\epsilon^{\gamma}} \leq \frac{e_i(1-e_j)}{e_j(1-e_i)} \leq \epsilon^{\gamma} \tag{7}$$

where e is the propensity score for individual i or j, and γ is a sensitivity parameter. An alternative derivation based on assignment probabilities allows to establish that $\Gamma = \epsilon^{\gamma}$ since $\frac{e_i(1-e_j)}{e_j(1-e_i)} = \epsilon^{\gamma(v_i - v + j)}$.[28] Therefore, Γ can be interpreted as the size of log of the coefficient γ for the unobserved covariate v. If $\Gamma = \epsilon^{\gamma} = 1$, two matched individuals with the same X will have the same probability of participation. On the other hand, if $\Gamma = \epsilon^{\gamma} > 1$, two matched individuals with the same X will have a different probability of participation, implying that there are unobserved factors. This exercise allows me to measure how large the CIA's deviation needs to be to qualitatively change the original results under the assumption that CIA is valid.[29]

This procedure is applied via a Wilconxon sign rank test for continuous outcomes and a McNemar test for dichotomous outcomes. A set of values for Γ from 1 to 2 are chosen following Keele (2010). The goal is to evaluate how p-values change due to bias related to an unobserved confounder; in other words, how inference regarding a causal estimate is affected once confounding factors are accounted for. Upper and lower bounds on the p-value are calculated based on the test statistic for different values of Γ. A value of $\Gamma = \Gamma_0$ implies that a treated person is Γ_0 times more likely to participate in the program due to unobserved covariates with respect to someone in the control group. Then, the procedure evaluates whether original results in terms of p-values remain unaffected. Because the lower-bound is always lower than the estimated p-value, only the upper-bound is relevant. Values of Γ associated with p-values larger than the usual significance standards reflect a qualitative change regarding the original inference about the effect of TTM for a given outcome.

[28] The equivalence of the derivation based on assignment probabilities and the original derivation based on unobserved covariates is proved in Proposition 12 of Rosenbaum (2002).

[29] See Rosenbaum (2002), Chapter 4, for details.

4 Empirical results

I organize the results according to the dimensions of empowerment that are emphasized by the academic and policy literature. Due to space constraints, the sensitivity analysis results are presented in the Online Appendix and will be briefly discussed in the next section.

4.1 Economic empowerment

Table 4 presents the main ATT impacts on economic empowerment. I consider a set of economic outcomes and assets. To put results in perspective, column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4, and 5 report the standard error, the t-statistics, and the associated p-value respectively. Column 6 shows the Benjamini and Hochberg factor to adjust for multiple outcomes, and column 7 informs whether the null hypothesis is rejected after this adjustment. Column 8 reports the effect size.

Among the economic outcomes, I study monthly per-capita income and expenditure[30], a multidimensional poverty indicator[31], an indicator of employment, and social protection measured by health and pension coverage. Regarding assets, I use indicator variables for whether the household has fridge, DVD, motorcycle, car, and washing machine.

I find important increases in monthly income and expenditure per-capita. In the case of income (row 1), the point estimate is USD 45.7 (Column 2).[32] With respect to a control group mean of USD 144.8, this effect is large (31% increase) and represents 18% of the minimum wage. This coefficient is statistically significant at the usual standards (p-value of 0.002) and robust to multiple comparisons. The effect for the monthly expenditure per-capita (row 2) is also important (coefficient of USD 30.3) and statistically significant without controlling for multiple testing (p-value of 0.018), but it is not robust to multiple comparisons.[33]

[30] All monetary measures were converted from Colombian Pesos to US dollars using the exchange rate of February 2017 (USD 1= COP 2,879.57)

[31] This multidimensional poverty indicator was developed using the methodology proposed by Schreiner (2014), and it has been used in around 60 countries. It considers ten indicators of family size, education, labor supply, access to electricity, source of energy for food preparation, and household assets.

[32] The minimum wage in Colombia in 2017 was COP 737,717, about USD 256.

[33] These results are consistent the effects multifaceted programs for the poor. For instance, effects for income (0.20 standard deviations (SDs) for TTM) are smaller than the ones reported by Banerjee et al. (2015) (0.38 SDs). However, these effects do not distinguish women from men. A more similar case is Bandiera et al. (2017). They report a 21% increase in earnings with respect to the control group, smaller than the 31% increase in incomes estimated for TTM. Regarding other women's empowerment programs, Bandiera et al. (2020) find no effects on expenditures after four years, and Buehren et al. (2017) report null effects on labor incomes. Of course, these comparisons are

Row 3 shows the impact of TTM on poverty reduction. Column 1 reports an estimated coefficient of -10.7 percentage points (t-statistic of -3.76). Given a mean poverty rate for the control group of 40.2%, this represents a reduction of 26.7% (column 8). This reduction is robust to accounting for multiple comparisons. Therefore, the program was highly successful in reducing poverty.[34]

Evidence regarding impacts on labor market outcomes is mixed. On the one hand, I find large impacts in the likelihood of being employed (point estimate of 21.6 percentage points with an effect size of 52%), robust to multiple comparisons (row 4). On the other hand, there is no evidence regarding increases in health and pension coverage (rows 5 and 6).[35]

Rows 7 to 11 show the impact of the intervention on the household's asset accumulation. Although the program seems to increase the likelihood that a household has a fridge (a critical appliance in a tropical area), this change is not robust to multiple comparisons.

Table 5 presents the ATT estimates for labor market outcomes. Row 1 shows that the TTM caused an increase of 29% in the likelihood of being part of the economically active population (18.2 percentage points, significant at the 1% level). Rows 2 and 3 present evidence of a 20% increase in household labor supply, but this is not explained by the formal sector (rows 4 and 5). The program is also related to a reduction of 36.8% in the likelihood of having unpaid family workers (row 6) and increases the probability of doing a paid activity by 45.6% (row 7). There is no evidence of changes in the likelihood of doing a business activity due to the program (row 8). Finally, the program caused a reduction of 43% in the likelihood of searching for a job (row 9), and an increase of 7.5 in the number of working hours (row 10). These results are robust to multiple comparisons. Because of the importance of employment opportunities for women's autonomy and empowerment[36], these effects are of first order from a policy perspective.[37]

Table 6 presents evidence regarding credit and saving effects. Regarding savings, rows 1 and 2

suggestive since differences in categories across studies may be related to differences in data collection methods and methodological differences for capturing incomes and expenditures.

[34] This reduction was larger than the observed in some of the multifaceted programs for the poor. For instance, in Bandiera et al. (2017), this reduction was 14% after four years for poor women in Bangladesh. However, they use a monetary poverty line, so their results are not strictly comparable with those reported here.

[35] Notice that health coverage is almost universal in Colombia because of the existence of a subsidized regime.

[36] Anderson and Eswaran (2009) propose and test a model where female autonomy depends on access to the labor market, mainly when employment opportunities are outside husbands' farms.

[37] Elsayed and Roushdy (2017) also find similar large effects in employment outcomes for women's empowerment intervention in Egypt, primarily driven by the informal sector. A key difference between their results and the ones reported in this paper is that their setting is mostly rural, whereas TTM operated in urban settings.

show that TTM increases savings for participants and other household members. In the first case, the point estimate is 7.3 percentage points, with an effect size of 31.5%. The effects are larger for the participant, with a point estimate of 21.8 percentage points, associated with an effect size of 55.5%. These results are robust to multiple comparisons. Row 5 shows that these savings are not in the formal banking system. Therefore, women save more, but they keep these savings out of the formal saving system.[38]

Regarding credit, there is no evidence that TTM is causally associated with changes in the likelihood of having a credit. As a consequence, no relationship is found between TTM and the type of lender.

Overall, these results suggest that a multifaceted intervention as TTM has significant impacts on economic empowerment. I have documented important effects on incomes, expenditures and poverty. These effects are in line with the evidence of multifaceted interventions for the poor (Banerjee et al., 2015 and Bandiera et al., 2017). These effects are mainly explained by improvements in labor market access, which is consistent with the fact that TTM mostly targeted urban women. No evidence on the realization of business activities is found, which suggests that entrepreneurship is less relevant in this setting. This can be consistent with a scenario where most TTM participants were "necessity entrepreneurs" who took advantage of TTM components to develop skills for a wage job, but this is hard to verify with the available information. The effects on savings are relevant because of the recent scholarship that shows that saving constraints are important for microenterprises' development (Dupas and Robinson, 2013). Interestingly, no effects were found for credit.[39]

4.2 Political and social empowerment

Table 7 presents the ATT impacts on decision-making and political empowerment. Rows 1 and 2 in Panel A present the effect of TTM on self-reported measures of household decision-making. In

[38] Austrian et al. (2020) and Buehren et al. (2017) also report large effects of empowerment intervention on savings. In the first case, the authors report estimated coefficients estimates of 19.3 percentage points for the treatment on the treated on the likelihood of saving (16% for the control mean at the baseline). In the second case, the authors estimate an effect of 2.8% percentage points (2% for the control mean at baseline). In the context of multifaceted programs for the poor, Banerjee et al. (2015) also find large effects for savings (156% increase with respect to the control mean).

[39] I am not the first in finding positive effects on savings but no effects on credit. Buehren et al. (2017) find a similar result in Tanzania. They rationalize this result by exploring the role of spillovers in participants' social networks.

the first case, there is an increase in the likelihood of a woman reporting having the responsibility of making decisions at the household. The effect is 5.1 percentage points, a 17.2% change with respect to the control mean. However, this effect is not robust to multiple comparisons. In the second case, no effect is found for the likelihood of both women and men sharing decision-making at the household.

Panel B in Table 7 presents the results for political empowerment. Row 3 shows that TTM led to an increase in the likelihood of voting in the 2016 elections with respect to the control group mean. The point estimate is 7.2 percentage points, representing an increase of 12.6% with respect to the control group mean.[40] The program also led to important increases in the membership in political parties (5.1 percentage points in row 4), the likelihood of being a political candidate (7.4 percentage points in row 5), and a self-reported measure of political participation (5.5 percentage points in row 6). All these effects are robust to multiple comparisons. No effect is found for a self-reported measure of women's participation in the community, although the baseline levels of this measure were already too high even in the control group (row 7).[41]

Table 8 presents results for a list experiment implemented during the data collection process to elicit the level of exposure to IPV in this setting. This an ideal approach to deal with sensitive issues like IPV where interviewed women may feel uncomfortable reporting episodes of violence.[42] Interviewed women were randomly assigned to one of two conditions. In condition 1, they are presented a list of five items about situations they may have experienced the last year including one sensitive item about whether their partners physically beat them. Condition 2 includes the same set of items, but excluding the sensitive one. Table S6 in the Online Appendix shows the exact question used to elicit IPV.

To estimate the effect of TTM on IPV, I run a linear regression of the total number of list outcomes reported on TTM participation and whether the surveyed woman was assigned to the sensitive item (condition 1), including an interaction between these two dummy variables on the

[40] Notice that voting is voluntary in Colombia.

[41] Banerjee et al. (2015) also finds large effects on political participation in their evaluation of several graduation programs for the poor. They do not find effects on voting, but document impacts on political party membership and attendance in village meetings.

[42] This method provides additional confidentiality with respect to alternative elicitation methods, which is expected to create incentives for truthful reporting. See Blair et al. (2015) for a technical discussion. Cullen (2020) compares this technique against face-to-face questions and audio computer-assisted self-interviews. Using data from Nigeria and Rwanda, she finds that IPV rates are severely underestimated under the alternative methods. Once a list experiment is used, IPV rates increase by 100% in Rwanda and 39% in Nigeria.

matched sample. The coefficient of interest is this latter interaction. TTM is found to have a positive effect on IPV. The estimated coefficient is 18.3 percentage points, significant at the 10% level. Considering a mean control response in the total number of list outcomes of 0.99 in the matched sample, this represents an increase of 18.5% in the likelihood of experiencing physical violence.[43] This result suggests the existence of backlash effects, presumably associated with the positive effects of TTM on incomes and labor market participation.[44]

Table 9 shows ATT impacts on social empowerment. The outcomes of interest are dummy variables of whether the surveyed woman belongs to a set of social organizations. A list of the 10 most common forms of social organizations in the Colombian Caribbean was included.[45] TTM is found to have large effects on the likelihood of being a member of several of these social organizations. In particular, the program increases the probability of being a member of community action boards (8.3 percentage points), producer associations (3.7 percentage points), citizen veedurias (1.9 percentage points), volunteering organizations (7.2 percentage points), sports or cultural groups (4.8 percentage points), and women organizations (14.5 percentage points). All these effects are robust to multiple comparisons. The effect sizes associated with these effects are pretty large, suggesting increases in the likelihood of being a member of these organizations of more than 100%. Hence, the program was effective in improving participants' social capital levels.

Overall, these results suggest a mixed picture. On the one hand, the program was very successful in improving political and social participation. These are critical components of empowerment as pointed out by the literature (Milazzo and Goldstein, 2019 and Duflo, 2012). Increasing women's political and social participation has been associated with a provision of public goods that better reflect women's preferences (Chattopadhyay and Duflo, 2004), raises girls' educational aspirations (Beaman et al., 2012), and increases in labor force participation (Deininger et al., 2020) and entrepreneurship (Ghani et al., 2014). Women's political participation also leads to more empower-

[43] It is important to emphasize that this result only captures a specific form of violence. Likewise, it is estimated using a linear regression for the matched sample. This particular choice is consistent with the brand of the matching literature that interpret this method as a pre-processor (Ho et al., 2007), but I am not aware of previous uses in the empowerment literature in the way presented here.

[44] Using data for 31 developing countries, Bhalotra et al. (2020) find that an increase in the probability of employment for women is associated with an 3% increase in the probability of experience IPV.

[45] I acknowledge the limitation of measuring social capital using membership in social organizations, although there are several examples in the literature. There is no consensus about what dimensions to take into account for social capital measurement but some of the proxies used in the past include trust, norms of reciprocity, engagement in public affairs, and participation in voluntary organizations.

ment (Bargain et al., 2018), and involvement in community affairs (Beath et al., 2013). On the other hand, the increase in exposure to IPV implies that TTM was not able to modify the unequal structure of power within participants' homes. These results contrast with recent evidence that conditional (Buller et al., 2018) and unconditional (Haushofer et al., 2019) cash transfers reduce IPV prevalence. However, the effect of empowerment on IPV is theoretically ambiguous (Angelucci and Heath, 2020), and these results can be consistent with extractive and status threat theories of IPV.

4.3 Control over the body and psychological empowerment

Table 10 shows the effects of TTM on several proxies of women's control over the body and psychological empowerment. In Panel A, the outcomes of interest are measures of the use of different contraceptive methods and some preventive health practices like cytology and breast exam. In Panel B the outcomes are beliefs about tolerance of violence towards women, women's role as an agent of change and a measure of self-confidence. These latter measures are interpreted as proxies of psychological empowerment.

There is no evidence of the effects of TTM on control over the body. No effect is found regarding the use of contraceptives. If something, it seems that the program reduced the use of contraceptive injections (row 7). The point estimate is -2.6 percentage points, a 65% reduction with respect to the control mean (4.1%), and robust to multiple comparisons. However, the program was effective in increasing the use of breast exams. The point estimate is 9.7 percentage points, representing an increase of 17% with respect to the control group mean. This effect is robust to multiple comparisons.

Regarding psychological empowerment, no evidence of impacts is found in any of these dimensions (Panel B). Although this does suggest the intervention was not effective in changing these beliefs, it has to be observed that the margin of improvement was small because the control group mean was already high.

These results seem to be consistent with weak or null effects of TTM on control over the body and psychological empowerment. Table S7 on the Online Appendix provides more evidence in this regard. This table presents results about the effects of TTM on attitudes towards women's rights. There is no evidence that the program affected these attitudes, although it is important to emphasize that the margin of improvement was minimal because the control mean for many of these attitudes were beyond the 95%.

However, these high levels of positive attitudes towards women's rights in this setting seem to be unrelated to changes in the allocation of responsibilities within the household. To see that, Table S8 in the Online Appendix studies the effect of TTM on time use for household chores. There is no evidence of changes in the time assigned for household chores as a consequence of the program. If something, it seems the program increased the time treated women assigned to family care (row 10). The estimated coefficient is 7.8 percentage points, representing an increase of 72% with respect to the control mean. Therefore, it seems TTM did very little or nothing in terms of changing gender inequalities within the household.

Overall, these results are consistent with the null effects of TTM on these dimensions. The evidence of the international literature in mixed in this regard. Bandiera et al. (2020) find large effects on behavior and outcomes about childbearing, marriage, and sex for girls in Uganda. On the other hand, Elsayed and Roushdy (2017), Sieverding and Elbadawy (2016), and Buehren et al. (2017) find no effect on related indicators also for young girls in Egypt and Tanzania. It is possible that the difference in ages (40 years old on average for TTM participants), and the more conservative environments in these countries, can help to explain the differences in results with Colombia. As mentioned before, the control group's baseline levels in many of these indicators were already high, which helps to explain the lack of significant differences. This issue deserves more research in the future.

5 Additional results: Consequences of empowerment

In this section, I explore some consequences of empowerment. In particular, I pay attention to the potential effects of TTM on proxies of investments and economic vulnerability. First, I study whether TTM participation is related to housing quality improvements, an important dimension of investment for poor households. Table S9 in the Online Appendix presents results for different house amenities. There is evidence in terms of better walls (row 2) and water (row 6) quality for treated women's households, but these effects are not robust to multiple comparisons.

Second, I explore the role of TTM in explaining two dimensions of vulnerability: food security and reliance on social programs. As a consequence of improvements in empowerment, less exposure to food insecurity is expected. The evidence reported in Table S10 in the Online Appendix suggests that some dimensions of food insecurity are reduced as a consequence of TTM. For instance, there is a reduction of 5.4 percentage points in the likelihood of reporting being worried about the possibility of no having food in the last 30 days (row 1), a reduction of 6.4 percentage points in the likelihood of reporting no consuming healthy foods in the last 30 days (row 3), and a reduction of 4.5 percentage points in the likelihood of no having a meal in the last 30 days (row 5). However, these effects are weak and not robust to multiple comparisons.

Regarding the dependency on social programs, one possibility is that TTM participants are less likely to depend on social programs. On the other hand, it could be possible that TTM can improve the information that participants have regarding access to social programs, leading to vulnerable participants more likely to enroll in social programs. Recall that the sample is composed of women that are considered vulnerable because they are part of the SISBEN system. Table S11 in the Online Appendix presents results regarding access to social programs. TTM increased the likelihood of being a beneficiary of the flagship conditional transfer program "Families in Action". The estimated coefficient is 5.5 percentage points (a 15.6% increase with respect to the control group mean), but it lacks of robustness when controlling for multiple comparisons. TTM is also causally associated with a reduction of 15.4 percentage points in the likelihood of receiving subsidies for displaced populations. This represents a reduction of 33.2% with respect to the control group mean. This effect is robust to multiple comparisons.

In sum, TTM seems to have mixed effects regarding the consequences of empowerment. There is some evidence in terms of better housing conditions and less exposure to vulnerability. However, these effects are weak and not robust to multiple comparisons in some cases.

6 Sensitivity analysis

This section discusses the sensitiveness of the main results to violations of the CIA assumption applying the sensitivity analysis proposed by Rosenbaum (2002) for the large set of outcomes analyzed in this study. This exercise aims to assess whether results that are statistically significant under the validity of CIA remain in this way once a degree of violation of CIA is allowed. A result would be considered robust if it remains statistically significant after this degree of CIA violation is introduced otherwise results are interpreted as very sensitive to relaxing the CIA assumption.

Tables S12-S31 in the Online Appendix show the results of this exercise. Because non-significant results are not affected by this analysis, the rest of the section will focus on those outcomes that were found statistically significant under the validity of the CIA assumption.

Tables S12 and S13 in the Online Appendix report the sensitivity analysis results for the economic empowerment outcomes originally discussed in Table 4. Lower and upper bounds for p-values for values of Γ from 1 to 2 are reported for each outcome variable. Because lower-bound p-values are always below the unconfounded p-value, the emphasis will be placed on the upper bound ones. In the case of the monthly income variables, the unconfounded p-value is 0.00. Then, for values of Γ larger than 1.4, the estimated coefficient is no longer significant. In other words, when matched treated units are 1.4 times more likely to receive the treatment due to unobserved covariates than control units, the monthly income variable is no longer statistically significant. Because this violation of CIA is large (40% higher chance of receiving treatment due to unobservables for treated units), this result suggests that the estimated effect is fairly robust.

The result for the monthly expenditure variable is less robust than the case of incomes. The original coefficient is no longer significant for values of Γ larger than 1.1, a minor violation of the CIA assumption. Not surprisingly this variable, although significant using the uncorrected p-value, was not robust to multiple comparisons. For the poverty and employed indicator variables, violations of CIA associated with 40% and 100% larger probabilities of receiving treatment due to unobservables are needed to modify the original inference about the statistical significance of these coefficients. These results suggest fairly and strongly robust estimated effects, respectively. The evidence for health and pension coverage shows that both variables are very sensitive to departures of unconfoundedness.

Tables S14 and S15 in the Online Appendix present the sensitivity analysis results for labor market effects reported initially in Table 5. In general, the degree of violation of the CIA assumption needs to be large in order to modify the level of significance of the estimated coefficients. For instance, the likelihood of being part of the EAP remains significant even for values of Γ larger than 2. The number of workers in the household become insignificant only for values of Γ larger than 1.7, and the fraction of workers is still significant for values of Γ lower than 1.5. All the other labor outcomes that were found to be statistically significant in Table 5 remain significant for values of Γ of at least 1.5, suggesting that large violations of the CIA assumption are required to change the conclusion about the effectiveness of TTM on these outcomes.

Tables S16 and S17 in the Online Appendix present the results for credit and savings outcomes. In terms of savings, the effects for the indicator variables about saving at home, whether the interviewed woman saves, and whether saving takes place at home require violations of the CIA assumption for values larger than 1.3, 1.7, and 1.8 of the Γ parameter respectively to change the statistical conclusion about the effect of TTM on these outcomes. Although some conclusions regarding the sensitivity of the credit outcomes are reported in Table S17, no further discussion is pursued due to the lack of robustness to multiple comparisons for these outcomes.

Due to space constraints, no detailed discussion about the sensitivity of the effects of TTM on political and social empowerment outcomes (Tables S18-S21 in the Online Appendix) as well as the control over the body and psychological empowerment outcomes (Tables S22-S27 in the Online Appendix) is provided. In summary, those outcomes for which effects are statistically strong are typically related to larger values of Γ for results to become no longer significant due to violations of the CIA assumption. This speaks well about the validity of this paper's research design. For the sake of completeness, the sensitivity analysis for the additional results on the consequences of empowerment is also reported (Tables S28-S31 in the Online Appendix).

7 Conclusion

This paper has analyzed the long-term effects of TTM, a multifaceted intervention designed to address the multidimensional nature of women's empowerment in the Colombian Caribbean. This intervention addressed several economic, political, social, and psychological barriers via the provision of training on soft and vocational skills, self-esteem and women's rights. It also included mentoring and in-kind seed capital. Compared to the average intervention studied in the empowerment literature, this program offered a more integral approach that tackled interlinked constraints to women's well-being.

I find that TTM caused large effects on many relevant economic, political, and social dimensions of women's empowerment, but failed to affect the structure of power within the participants' households. TTM increased household incomes and labor market outcomes, causing a reduction in multidimensional poverty. The program also led to an increase in savings. Regarding political and social empowerment, the program led to more political participation and involvement in social organizations and public affairs. However, the program did not affect the use of contraceptives, self-confidence, beliefs about women's rights, or the allocation of time for household chores. Even worse, TTM is associated with a sizable positive effect on IPV. In a way, the program could not enter participants' homes despite its large effects on outside-home economic, social and political activities.

These results are consistent with related interventions that have found important economic, political, and social effects with limited or null impacts in terms of beliefs, the within-household decision making and allocation of power (Beath et al., 2013, Elsayed and Roushdy, 2017, Austrian et al., 2020, and Sieverding and Elbadawy, 2016). From a policy perspective, this implies that multifaceted programs might not be enough in settings where social norms and attitudes do not favor gender equality. Complementary interventions targeted to men can be useful in this regard. For instance, programs that tackle the support of restrictive gender norms among boys can be a profitable long-term investment (Dhar et al., 2020). On the other hand, innovations to modify adult males' beliefs about gender equality and conformity with discriminatory norms are of first order relevance. However, there is very little written about the effectiveness of this kind of approach.[46] Smarter policy design motivated by behavioral science can also modify gender norms, as some recent evidence has shown (Bursztyn et al., 2020).

In sum, women's empowerment interventions and gender-equalizing policy reforms require carefully thinking about innovative ways to incorporate men if their goal is to obtain sustainable results. Designing and evaluating such interventions constitutes an fundamental area of future research.

[46] In Colombia, the Government of Atlantico developed a TTM's sister program for men called "Transformate Tu Hombre" in 2013. The program has a similar structure as TTM and includes a training component designed to address gender stereotypes and IPV. To date, this program has not be evaluated.

References

Abadie, A. and G. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*(1), 235–267.

Abadie, A. and G. Imbens (2008). On the failure of bootstrap for matching estimators. *Econometrica 76*(6), 1537–1557.

Abadie, A. and G. Imbens (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics 29*(1), 1–11.

Abadie, A. and G. Imbens (2016). Matching on the estimated propensity score. *Econometrica 84*(2), 781–807.

Anderson, S. and M. Eswaran (2009). What determines female autonomy? Evidence from Bangladesh. *Journal of Development Economics 90*, 179–191.

Angelucci, M. and R. Heath (2020). Women empowerment programs and intimate partner violence. *AEA Papers and Proceedings 110*, 610–614.

Arceneaux, K., A. Gerber, and D. Green (2006). Comparing experimental and matching methods using a large-scale voter mobilization experiment. *Political Analysis 14*(1), 37–62.

Aslam, M. (2013). Empowering women: Education and the pathways of change. Background paper for the Education for All Global Monitoring Report.

Austrian, K., E. Soler-Hampejsek, J. Behrman, J. Digitale, N. Jackson, M. Bweupe, and P. Hewett (2020). The impact of the Adolescent Girls Empowerment Program (AGEP) on short and long-term social, economic, education and fertility outcomes: a cluster randomized controlled trial in Zambia. *BMC Public Health 20*(349), 1–15.

Baird, S. and B. Ozler (2016). Sustained effects on economic empowerment of interventions for adolescent girls: Existing evidence and knowledge gaps. CDG Background Paper.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman (2020). Women's empowerment in action: Evidence from a randomized control trial in Africa. *American Economic Journal: Applied Economics 12*(1), 210–259.

Bandiera, O., R. Burguess, N. Das, S. Gulesci, I. Rasul, and M. Sulaiman (2017). Labor markets and poverty in village economies. *Quarterly Journal of Economics 132*(2), 811–870.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Pariente, J. Shapiro, B. Thuysbaert, and C. Udry (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science 348*(6236), 1260799.

Bargain, O., D. Boutin, and H. Champeux (2018). Women's political participation and intrahousehold empowerment:. IZA Discussion Paper Series 11534.

Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science 335*(6068), 582–586.

Beath, A., F. Christia, and R. Enikolopov (2013). Empowering women through development aid: Evidence from a field experiment in Afghanistan. *American Political Science Review 107*(3), 540–557.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

Bhalotra, S., U. Kambhampati, S. Rawlings, and Z. Siddique (2020). Intimate partner violence. The influence of job opportunities for men and women. World Bank Policy Research Working Paper 9118.

Blair, G., K. Imai, and Y. Zhou (2015). Design and analysis of the randomized response technique. *Journal of the American Statistical Association 110*(511), 1304–1319.

Bott, S., A. Guedes, A. Ruiz-Celis, and J. Mendoza (2019). Intimate partner violence in the Americas: A systematic review and reanalysis of national prevalence estimates. *Pan American Journal of Public Health 43*, 1–12.

Brooks, W., K. Donovan, and T. Johnson (2018). Mentors or teachers? Microenterprise training in Kenya. *American Economic Journal: Applied Economics 10*(4), 196–221.

Buehren, N., M. Goldstein, S. Gulesci, M. Sulaiman, and V. Yam (2017). Evaluation of an adolescent development program in Tanzania. World Bank Policy Research Working Paper 7961.

Buller, A., A. Peterman, M. Ranganathan, A. Bleile, M. Hidrobo, and L. Heise (2018). A mixed-method review of cash transfers and intimate partner violence in low- and middle-income countries. *World Bank Research Observer 33*(2), 218–258.

Bursztyn, L., A. Gonzales, and D. Yanagizawa-Drott (2020). Misperceived social norms: Women working outside home in Saudi Arabia. *American Economic Review 110*(10), 2997–3029.

Buvinic, M. and R. Furst-Nichols (2016). Promoting women's economic empowerment: What works? *World Bank Research Observer 31*(1), 59–101.

Buvinic, M. and M. O'Donnell (2019). Gender matters in economic empowerment interventions: A research review. *World Bank Research Observer 34*(2), 309–346.

Campos, F., M. Frese, M. Goldstein, L. Lacavone, H. Johnson, D. McKenzie, and M. Mensmann (2017). Teaching personal initiative beats traditional training in boosting small businesses in West Africa. *Science 357*(6357), 1287–1290.

Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica 72*(5), 1409–1443.

CNC (2017). Informe de resultados y recomendaciones de la evaluacion de Transformate Tu Mujer. Producto 4. Centro Nacional de Consultoria.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second ed.). Lawrence Erlbaum Associates.

Colson, K., K. Rudolph, S. Zimmerman, D. Goin, E. Stuart, M. van der Laan, and J. Ahern (2016). Optimizing matching and analysis combinations for estimating causal effects. *Nature Scientific Reports 6*(23222).

Croke, K., M. Goldstein, and A. Holla (2017). Can job training decrease women's self-defeating biases? Experimental evidence from Nigeria. World Bank Policy Research Working Paper 8141.

Crump, R., Hotz, G. Imbens, and O. Mitnik (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika 96*(1), 187–199.

Cullen, C. (2020). Methods matters. Underreporting of intimate partner violence in Nigeria and Rwanda. World Bank Policy Research Working Paper 9274.

de Mel, S., D. McKenzie, and C. Woodruff (2008). Returns to capital in microenterprises: Evidence from a field experiment. *Quarterly Journal of Economics 123*(4), 1329–1372.

de Mel, S., D. McKenzie, and C. Woodruff (2012). One-time transfers of cash or capital have long-lasting effects of microenterprises in Sri Lanka. *Science 335*(6071), 962–966.

Deininger, K., S. Jin, H. Nagarajan, and S. Singh (2020). Political reservation and female labor force participation in rural India. World Bank Policy Research Working Paper 9350.

Dhar, D., T. Jain, and S. Jayachandran (2020). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. Mimeo.

Diamond, A. and J. Sekhon (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics 95*(3), 932–945.

Doepke, M. and M. Tertilt (2011). Does female empowerment promote economic development? World Bank Policy Research Working Paper 5714.

Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic Literature 50*(4), 1051–1079.

Dupas, P. and J. Robinson (2013). Savings constraints and microenterprise development: Evidence from a field experiment in Kenya. *American Economic Journal: Applied Economics 5*(1), 163–192.

ECLAC (2020). Gender equality observatory for Latin America and the Caribbean. https://oig.cepal.org/.

Elsayed, A. and R. Roushdy (2017). Empowering women under social constraints: Evidence from a field intervention in rural Egypt. IZA Discussion Paper Series 11240.

Fafchamps, M., D. McKenzie, S. Quinn, and C. Woodruff (2014). Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana. *Journal of Development Economics 106*, 211–226.

Fonseca, A. (2019). Informe de empoderamiento economico de las mujeres en Colombia. Situacion actual y recomendaciones de politica. Equidad de la Mujer, Gobierno de Colombia.

Fox, L. and C. Romero (2017). In the mind, the household, or the market? Concepts and measurement of women's economic empowerment. World Bank Policy Research Working Paper 8079.

Ghani, E., W. Kerr, and S. O'Connell (2014). Political reservations and women's entrepreneurship in India. *Journal of Development Economics 108*, 138–153.

Haushofer, J., C. Ringdal, J. Shapiro, and X. Wang (2019). Income changes and intimate partner violence: Evidence from unconditional cash transfers in Kenya. NBER Working Paper 25627.

Heath, R. and S. Jayachandran (2018). The causes and consequences of increased female education and labor force participation in developing countries. In S. Averett, L. Argys, and S. Hoffman (Eds.), *Oxford Handbook of Women and the Economy*. Oxford University Press.

Ho, D., K. Imai, G. King, and E. Stuart (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis 15*(3), 199–236.

IDEA, PNUD, and ONU MUJERES (2019). Colombia: La hora de la paridad. IDEA and PNUD and ONU MUJERES.

Imai, K., G. King, and E. Stuart (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society. Series A 172*(Issue 2), 481–502.

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics 86*(1), 4–29.

Imbens, G. (2015). Matching methods in practice: Three examples. *The Journal of Human Resources 50*(2), 373–419.

Imbens, G. and D. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences:An Introduction.* Cambridge University Press.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics 7*, 63–88.

Jayachandran, S. (2020). Social norms as a barrier to women's employment in developing countries. Northwestern University.

Keele, L. (2010). An overview of rbounds: An r package for Rosenbaum bounds sensitivity analysis with matched data. Mimeo.

Klasen, S. (2019). What explains uneven female labor force participation levels and trends in developing countries? *World Bank Research Observer 34*(2), 161–197.

LaFortune, J., J. Riutort, and J. Tessada (2018). Role models or individual consulting? The impact of personalizing micro-entrepreneurship training. *American Economic Journal: Applied Economics 10*(4), 222–245.

McKenzie, D. (2017). How effective are active labor market policies in developing countries? A critical review of recent evidence. *World Bank Research Observer 32*(2), 127–154.

McKenzie, D., S. Stillman, and J. Gibson (2010). How important is selection? Experimental vs. non-experimental measures of the income gains from migration. *Journal of the European Economic Association 8*(4), 913–945.

Milazzo, A. and M. Goldstein (2019). Governance and women's economic and political participation: Power inequalities, formal constraints and norms. *World Bank Research Observer 34*(1), 34–64.

Montalvao, J., M. Frese, M. Goldstein, and T. Kilic (2017). Soft skills for hard constraints: Evidence from high-achieving female farmers. World Bank Policy Research Working Paper 8095.

ONU MUJERES (2018). *El progreso de las mujeres en Colombia. Transformar la economia para realizar los derechos.* ONU MUJERES.

ONU MUJERES (2020). Boletin estadistico: Empoderamiento economico de las mujeres en Colombia. ONU MUJERES.

Prennushi, G. and A. Gupta (2014). Women's empowerment and socio-economic outcomes. Impacts of the Andra Pradesh rural poverty reduction program. World Bank Policy Research Working Paper 6841.

Rosenbaum, P. (2002). *Observational Studies.* Springer.

Rosenbaum, P. and D. Rubin (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician 39*(1), 33–38.

Roy, T., R. Acharya, and A. Roy (2016). *Statistical Survey Design and Evaluating Impact.* Cambridge University Press.

Sanchez, M. (2018). Empoderamiento educativo de las mujeres. Situacion actual y lineas de politica. Equidad de la Mujer, Gobierno de Colombia.

Schreiner, M. (2014). The process of poverty-scoring analysis. Available at http://www.simplepovertyscorecard.com/.

Sekhon, J. and W. Mebane (1998). Genetic optimization using derivatives: Theory and application to nonlinear models. *Political Analysis 7*, 189–203.

Sen, A. (1999). *Development as Freedom.* Alfred Knopf.

Sieverding, M. and A. Elbadawy (2016). Empowering adolescent girls in socially conservative settings: Impacts and lessons learned from the Ishraq program in rural upper Egypt. *Studies in Family Planning 47*(2), 129–144.

SMEG (2020). Transformate Tu Mujer. Secretaria de las Mujeres y la Equidad de Genero de la Gobernacion del Atlantico. https://www.atlantico.gov.co/index.php/secretarias/mujeres-y-equidad.

Smith, J. and P. Todd (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics 125*(Issues 1–2), 305–353.

Upadhyay, U., J. Gipson, E. Ciaraldi, M. Withers, A. Fraser, N. Prata, S. Lewis, and M. Huchko (2014). Women's empowerment and fertility: A review of the literature. *Social Science and Medicine 115*, 111–120.

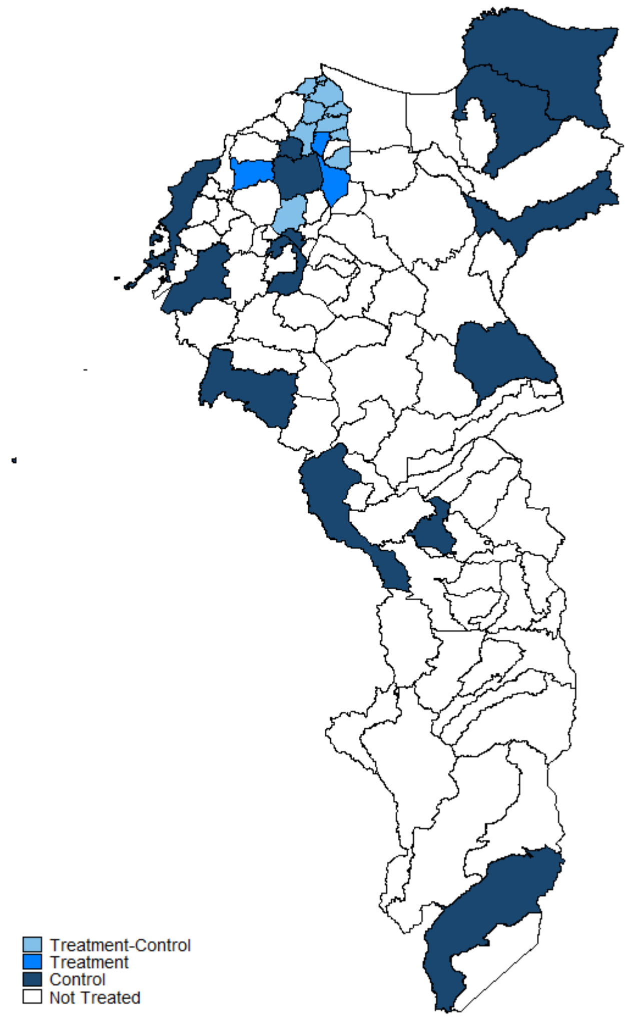
Vyas, S. and C. Watts (2009). How does economic empowerment affect women's risk of intimate partner violence in low and middle income countries? A systematic review of published evidence. *Journal of International Development 21*(5), 577–602.

Woodruff, C. and D. McKenzie (2012). Improving the productivity and earnings of women owned and/or managed enterprises in developing countries: What works? Available at http://www.womeneconroadmap.org/.

World Bank (2012). *Gender equality and development.* World Development Report.

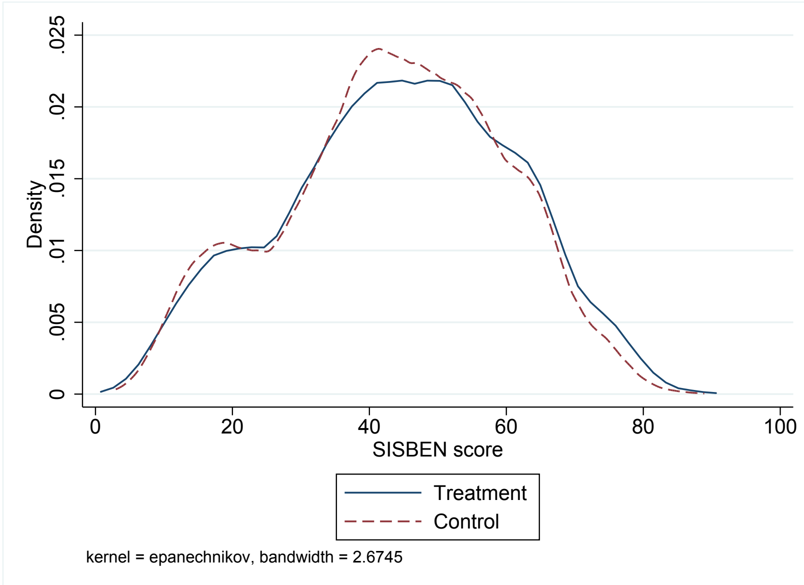
World Economic Forum (2019). *Global Gender Gap Report 2020.* World Economic Forum.

Figure 1: Municipalities in the sampling frame



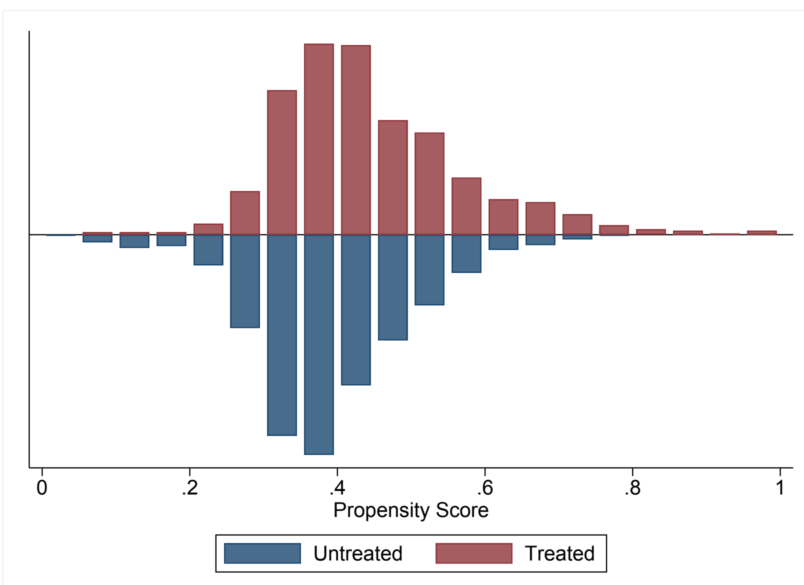
Note: Author's elaboration. "Treatment" refers to municipalities where only programs' participants were identified for the sampling frame. "Control" refers to municipalities where only control units were identified. "Treatment-control" refers to municipalities where both participants and non-participants were identified. The map covers the departments of Atlantico, Bolivar and Magdalena.

Figure 2: Kernel density for the SISBEN score by treatment status in the matched sampling frame



Note: Author's elaboration. The SISBEN score is a welfare-index used by the Colombian government for the selection of social programs' beneficiaries. Each household receives a score from 0 to 100 (from poorest to richest). The score is calculated using 24 variables across four dimensions: health, education, housing, and vulnerability.

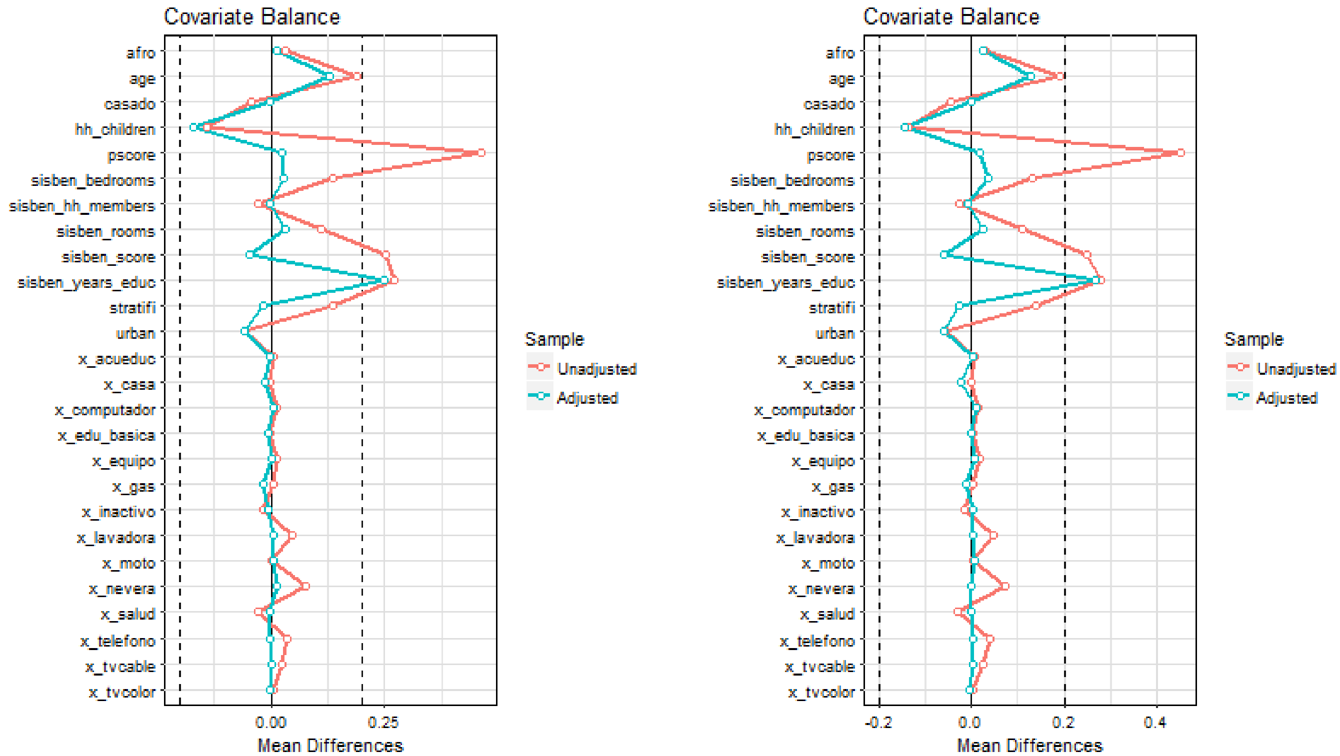
Figure 3: Assessing overlap in the sample-based propensity score estimated using Imbens and Rubin's (2015) algorithm



Note: Author's elaboration. The propensity score was estimated using the methodology suggested by Imbens and Rubin (2015). The variables used in the estimation are described in Table S3 in the Online Appendix.

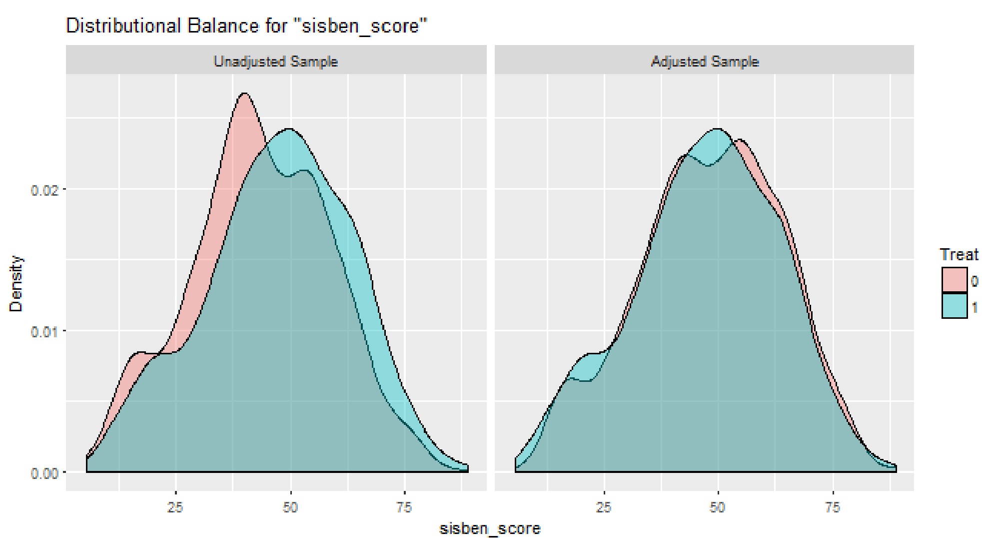
Figure 4: Assessing balance before and after matching

(a) Full sample (b) Trimmed sample



Note: Author's elaboration. The variables used in this exercise are described in tables S1, S2 and S3 in the Online Appendix.

Figure 5: Assessing distributional balance of the SISBEN score before and after matching



Note: Author's elaboration. The figure displays the difference in the distribution of the SISBEN score before and after matching for the trimmed sample.

Table 1. Descriptive and balance statistics of sampling frame

Variable	Treated Mean (1)	Treated Variance (2)	Control Mean (3)	Control Variance (4)	Standardized Difference (5)	Variance ratio (6)	t-test p-value (7)	Wilconxon p-value (8)	Fisher p-value (9)
House	0.852	0.126	0.861	0.120	-0.024	1.051	0.118	0.118	0.120
Gas	0.831	0.141	0.838	0.136	-0.020	1.037	0.202	0.202	0.203
Telephone	0.127	0.111	0.108	0.096	0.058	1.148	0.000	0.000	0.000
Aqueduct	0.909	0.082	0.908	0.083	0.004	0.988	0.790	0.790	0.806
Fridge	0.583	0.243	0.571	0.245	0.025	0.992	0.119	0.119	0.123
Washing machine	0.190	0.154	0.184	0.150	0.015	1.024	0.348	0.348	0.353
Television	0.831	0.141	0.827	0.143	0.009	0.984	0.548	0.548	0.560
Cable TV	0.081	0.075	0.087	0.079	-0.021	0.940	0.190	0.190	0.195
Computer	0.037	0.035	0.034	0.033	0.017	1.088	0.286	0.286	0.282
Stereo	0.176	0.145	0.178	0.146	-0.005	0.992	0.759	0.759	0.773
Motorcycle	0.043	0.041	0.042	0.040	0.006	1.027	0.715	0.715	0.724
Health	0.755	0.185	0.761	0.182	-0.014	1.018	0.361	0.362	0.367
Basic education	0.923	0.071	0.910	0.082	0.047	0.868	0.003	0.003	0.003
Inactive	0.901	0.089	0.904	0.087	-0.010	1.028	0.516	0.516	0.524
Propensity score	0.074	0.012	0.053	0.006	0.222	2.049	0.000	0.000	NA
SISBEN score	44.321	280.082	43.388	260.017	0.057	1.077	0.000	0.001	NA
Married	0.726	0.199	0.721	0.201	0.009	0.991	0.565	0.565	0.574
Age	38.609	99.778	38.161	115.363	0.043	0.865	0.007	0.000	NA
Sampling frame size	5,656		14,145						

Note: This table reports balance statistics for the sampling frame based on SISBEN's records for 2011. Columns 1 and 3 present the mean, and columns 2 and 4 present the variance of the pre-treatment variables for treated and control groups. Column 5 shows the standardized difference in means and column 6 presents the variance ratio between treated and control groups. Columns 7, 8 and 9 show the p-value for a t-test, Wilconxon test and Fisher test of differences in means for treated and control groups respectively.

Table 2. Descriptive and balance statistics of trimmed sample

Variable	Treated Mean (1)	Treated Variance (2)	Control Mean (3)	Control Variance (4)	Standardized difference (5)	Variance ratio (6)	t-test p-value (7)	Wilconxon p-value (8)	Fisher p-value (9)
House	0.880	0.106	0.882	0.104	-0.005	1.012	0.922	0.922	0.940
Gas	0.865	0.117	0.861	0.120	0.011	0.919	0.829	0.829	0.887
Telephone	0.134	0.116	0.097	0.087	0.117	0.978	0.016	0.016	0.016
Aqueduct	0.928	0.067	0.921	0.073	0.026	1.330	0.592	0.592	0.645
Fridge	0.630	0.233	0.558	0.247	0.148	0.945	0.003	0.003	0.003
Washing machine	0.207	0.165	0.162	0.136	0.118	1.213	0.015	0.016	0.018
Television	0.848	0.129	0.844	0.132	0.010	0.982	0.844	0.844	0.892
Cable TV	0.094	0.086	0.071	0.066	0.085	1.298	0.078	0.078	0.088
Computer	0.038	0.037	0.024	0.023	0.083	1.586	0.082	0.082	0.085
Stereo	0.189	0.153	0.173	0.143	0.043	1.074	0.382	0.382	0.407
Motorcycle	0.048	0.046	0.046	0.044	0.008	1.033	0.877	0.877	0.908
Health	0.760	0.183	0.792	0.165	-0.076	1.107	0.119	0.119	0.125
Basic education	0.938	0.058	0.937	0.059	0.004	0.985	0.929	0.929	1.000
Inactive	0.883	0.104	0.900	0.090	-0.056	1.153	0.248	0.248	0.267
Pscore IR	0.444	0.015	0.390	0.011	0.487	1.370	0.000	0.000	NA
SISBEN score	47.237	255.020	43.245	236.638	0.255	1.078	0.000	0.000	NA
Married	0.708	0.207	0.753	0.186	-0.102	1.113	0.036	0.036	0.040
Afro	0.237	0.181	0.208	0.165	0.069	1.098	0.155	1.155	0.157
Sample size	709		1,014						

Note: This table reports balance statistics for the sample collected for the retrospective evaluation of TTM in 2017. The sample was trimmed using the propensity score following Crump et al. (2009). Columns 1 and 3 present the mean, and columns 2 and 4 present the variance of the pre-treatment variables for treated and control groups. Column 5 shows the standardized difference in means and column 6 presents the variance ratio between treated and control groups. Columns 7, 8 and 9 show the p-value for a t-test, Wilconxon test and Fisher test of differences in means for treated and control groups respectively.

Table 3. Descriptive and balance statistics of trimmed sample - Before and after matching

Variable	**Before Matching**					**After Matching**				
	Mean treatment (1)	**Mean control (2)**	**Standardized difference (3)**	**Var. ratio (4)**	**t-test p-value (5)**	**Mean treatment (6)**	**Mean control (7)**	**Standardized difference (8)**	**Var. ratio (9)**	**t-test p-value (10)**
House	0.880	0.882	-0.475	1.012	0.923	0.880	0.903	-7.087	1.206	0.049
Gas	0.865	0.861	1.066	0.978	0.828	0.865	0.876	-3.433	1.080	0.454
Telephone	0.134	0.097	10.955	1.330	0.018	0.134	0.132	0.483	1.011	0.916
Aqueduct	0.928	0.921	2.693	0.919	0.589	0.928	0.926	0.818	0.974	0.860
Fridge	0.630	0.558	14.964	0.945	0.003	0.630	0.631	-0.146	1.001	0.930
Washing machine	0.207	0.162	11.240	1.213	0.017	0.207	0.204	0.869	1.013	0.286
Color TV	0.848	0.844	0.971	0.982	0.843	0.848	0.852	-1.307	1.026	0.623
Cable TV	0.094	0.071	8.026	1.298	0.085	0.094	0.092	0.723	1.021	0.584
Computer	0.038	0.024	7.525	1.586	0.095	0.038	0.030	4.050	1.246	0.044
Stereo	0.189	0.173	4.190	1.074	0.385	0.189	0.182	1.740	1.029	0.634
Motorcycle	0.048	0.046	0.750	1.033	0.877	0.048	0.042	2.639	1.127	0.102
Married	0.708	0.753	-9.981	1.113	0.037	0.708	0.708	0.000	1.000	1.000
Health	0.760	0.792	-7.417	1.107	0.122	0.760	0.760	0.000	1.000	1.000
Basic education	0.938	0.937	0.438	0.985	0.929	0.938	0.938	0.000	1.000	1.000
Inactive	0.883	0.900	-5.427	1.153	0.254	0.883	0.881	0.658	0.985	0.891
Afro	0.237	0.208	6.784	1.098	0.158	0.237	0.211	6.022	1.085	0.196
SISBEN score	47.237	43.245	25.001	1.078	0.000	47.237	48.198	-6.014	1.039	0.044
Propensity score	0.445	0.390	45.281	1.370	0.000	0.445	0.442	1.675	1.112	0.044

Note: This table reports balance statistics before and after matching for the sample collected for the retrospective evaluation of TTM in 2017. The sample was trimmed using the propensity score following Crump et al. (2009). Columns 1, 2, 6 and 7 present the mean of the pre-treatment variables for treated and control groups before and after matching. Columns 3 and 8 show the standardized difference in means before and after matching multiplied by 100. Columns 4 and 9 present the variance ratio between treated and control groups before and after matching. Columns 5 and 10 show the p-value for a t-test of differences in means for treated and control groups before and after matching.

Table 4. Effects of TTM on economic empowerment
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
Panel A: Main outcomes								
1. Monthly income	144.800	45.730	15.084	3.032	0.002	0.014	1	0.316
2. Monthly expenditure	378.400	30.306	14.135	2.144	0.032	0.018	0	0.080
3. Poverty	0.402	-0.107	0.029	-3.765	0.000	0.009	1	-0.267
4. Employed	0.414	0.216	0.029	7.324	0.000	0.005	1	0.521
5. Health coverage	0.985	-0.021	0.011	-1.894	0.058	0.023	0	-0.021
6. Pension coverage	0.115	-0.023	0.018	-1.240	0.215	0.036	0	-0.199
Panel B: Assets								
7. Fridge	0.858	0.033	0.018	1.776	0.076	0.027	0	0.038
8. DVD	0.238	0.039	0.027	1.455	0.146	0.032	0	0.164
9. Motorcycle	0.156	-0.012	0.021	-0.583	0.560	0.050	0	-0.079
10. Car	0.019	-0.006	0.010	-0.611	0.541	0.045	0	-0.331
11. Washing machine	0.559	-0.026	0.030	-0.880	0.379	0.041	0	-0.047
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1457

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table 5. Effects of TTM on labor market outcomes
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
1. Belonging to the EAP	0.629	0.182	0.026	6.930	0.000	0.010	1	0.289
2. Number of workers in the HH	1.568	0.313	0.061	5.101	0.000	0.030	1	0.200
3. Fraction of workers in the HH	0.380	0.079	0.014	5.575	0.000	0.020	1	0.208
4. Numer of formal workers in the HH	0.725	-0.014	0.050	-0.279	0.780	0.045	0	-0.019
5. Fraction of formal workers in the HH	0.170	-0.001	0.013	-0.106	0.915	0.050	0	-0.008
6. Unpaid family worker	0.516	-0.190	0.030	-6.408	0.000	0.015	1	-0.368
7. Realization of paid activity	0.495	0.226	0.028	7.996	0.000	0.005	1	0.456
8. Realization of business activity	0.043	-0.004	0.013	-0.334	0.738	0.040	0	-0.100
9. Job search	0.091	-0.039	0.015	-2.563	0.010	0.035	1	-0.431
10. Work hours	19.040	7.527	1.469	5.126	0.000	0.025	1	0.395
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1457

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table 6. Effects of TTM on credit and saving outcomes
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
Panel A: Savings								
1. Household savings	0.233	0.073	0.027	2.759	0.006	0.017	1	0.315
2. Savings (interviewed woman)	0.218	0.121	0.027	4.550	0.000	0.003	1	0.555
3. Savings in banks or corporations	0.020	0.008	0.010	0.785	0.432	0.033	0	0.376
4. Savings in microcredit institution	0.006	-0.001	0.005	-0.119	0.905	0.047	0	-0.094
5. Saving at home	0.188	0.112	0.025	4.497	0.000	0.006	1	0.596
6. Savings in informal schemes	0.003	0.004	0.004	0.906	0.365	0.031	0	1.201
Panel B: Credit								
7. Debt holdings	0.394	0.037	0.030	1.229	0.219	0.025	0	0.094
8. Lender- Bank	0.197	0.038	0.026	1.487	0.137	0.022	0	0.194
9. Lender- Microcredit	0.029	0.018	0.011	1.651	0.099	0.019	0	0.635
10. Lender-Producers	0.006	-0.001	0.005	-0.222	0.824	0.039	0	-0.197
11. Lender- Suppliers	0.053	0.003	0.015	0.189	0.850	0.042	0	0.052
12. Lender- Lenders	0.107	0.003	0.019	0.148	0.882	0.044	0	0.026
13. Lender - Friends	0.069	-0.018	0.016	-1.138	0.255	0.028	0	-0.260
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1457

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table 7. **Effects of TTM on decision-making and political empowerment**
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
Panel A: Decision-making at home								
1. Woman	0.294	0.051	0.027	1.903	0.057	0.025	0	0.172
2. Both	0.556	-0.042	0.029	-1.459	0.145	0.040	0	-0.076
Panel B: Political outcomes								
3. Vote in 2016	0.570	0.072	0.029	2.462	0.014	0.020	1	0.126
4. Membership in political parties	0.057	0.051	0.015	3.300	0.001	0.010	1	0.894
5. Candidate	0.043	0.074	0.017	4.377	0.000	0.005	1	1.719
6. Political participation	0.864	0.055	0.018	3.068	0.002	0.015	1	0.063
7. Women's participation in the community	0.942	0.022	0.013	1.724	0.085	0.030	0	0.023
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1457

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table 8. Effect of TTM on intimate partner violence
List experiment

Variable	Estimate	S.E.	t-value	Pr(> |t|)
Treatment	-0.035	0.071	-0.499	0.618
Condition 1	-0.141	0.078	-1.810	0.071
Treatment X Condition 1	0.183	0.099	1.836	0.067
Matched observations				709

Note: This table reports the results of a list experiment to elicit sensitive behaviors, in this case the exposure to intimate partner violence (IPV). Surveyed women were randomly exposed to one of two conditions: Condition 1 is a five-item list containing a sensitive question about IPV and condition 2 is the same list but excluding this sensitive question. The original question used in the list experiment is shown in Table S6 in the Online Appendix. A regression model with interactions between treatment status and condition 1 in the list experiment was estimated using the matched sample. The dependent variable is the total number of list outcomes reported.

Table 9. Effects of TTM on social capital
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
1. Community action board	0.082	0.083	0.020	4.169	0.000	0.020	1	1.015
2. Producers' associations	0.009	0.037	0.010	3.654	0.000	0.025	1	4.151
3. Citizen veedurias	0.004	0.019	0.007	2.560	0.010	0.030	1	4.717
4. Displaced population organizations	0.056	-0.007	0.014	-0.499	0.618	0.045	0	-0.121
5. Volunteering	0.038	0.072	0.015	4.762	0.000	0.010	1	1.883
6. Parents' groups	0.088	-0.003	0.017	-0.175	0.861	0.050	0	-0.035
7. Sports or cultural groups	0.008	0.048	0.011	4.491	0.000	0.015	1	5.986
8. Women groups	0.037	0.145	0.019	7.453	0.000	0.005	1	3.909
9. Youth groups	0.006	0.006	0.006	1.015	0.310	0.040	0	1.041
10. Syndicates	0.006	0.007	0.005	1.382	0.167	0.035	0	1.176
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1453

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table 10. Effects of TTM on control over body and psychological empowerment
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
Panel A: Control over body								
1. Use of contraceptive methods	0.434	-0.031	0.031	-1.004	0.315	0.022	0	-0.071
2. Use of natural method	0.002	0.001	0.003	0.218	0.827	0.050	0	0.334
3. Use of pills	0.027	-0.012	0.010	-1.285	0.199	0.017	0	-0.453
4. Use of condom	0.008	-0.005	0.006	-0.803	0.422	0.033	0	-0.631
5. Use of intrauterine device	0.007	-0.003	0.005	-0.546	0.585	0.044	0	-0.391
6. Female sterilization	0.319	0.022	0.029	0.784	0.433	0.039	0	0.070
7. Use of contraceptive injections	0.041	-0.026	0.010	-2.571	0.010	0.011	1	-0.645
8. Cytology	0.904	0.015	0.017	0.849	0.396	0.028	0	0.016
9. Breast exam	0.571	0.097	0.028	3.406	0.001	0.006	1	0.170
Panel B: Pyschological empowerment								
10. Disagree with violence against women	0.956	-0.024	0.015	-1.597	0.110	0.035	0	-0.025
11. Woman as agent of change	0.869	-0.002	0.021	-0.075	0.940	0.050	0	-0.002
12. Self-confidence	0.760	0.036	0.026	1.375	0.169	0.045	0	0.047
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1449

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Online Appendix to Accompany "Empowering women through multifaceted interventions: Long-term evidence from a double matching design"

Stanislao Maldonado*

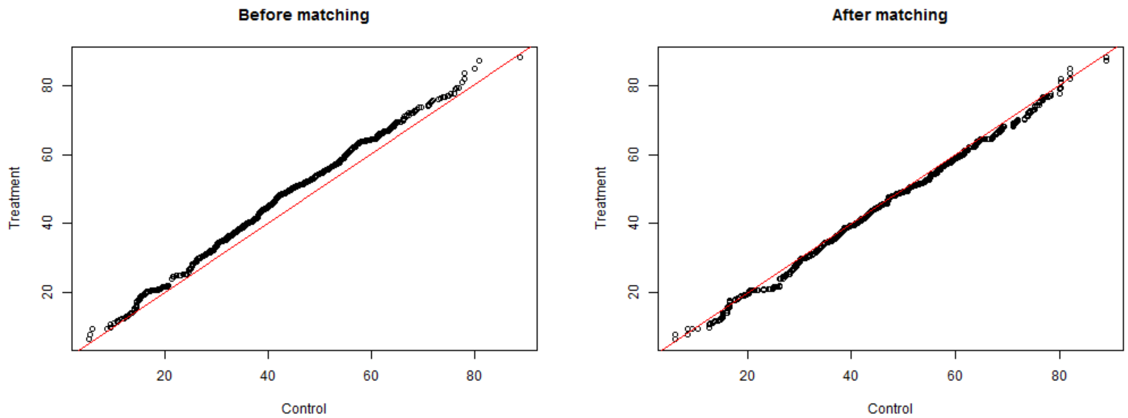
Universidad del Rosario

This version: September 14, 2020
For the latest version of this paper, click here.

*Address: Department of Economics, Calle 12C N0 4-59, Universidad del Rosario, Bogota, Colombia, E-mail: stanislao.maldonado@urosario.edu.co. Web: http://stanislaomaldonado.org/

Figure S1: Quantile-quantile plot for SISBEN score in the trimmed sample

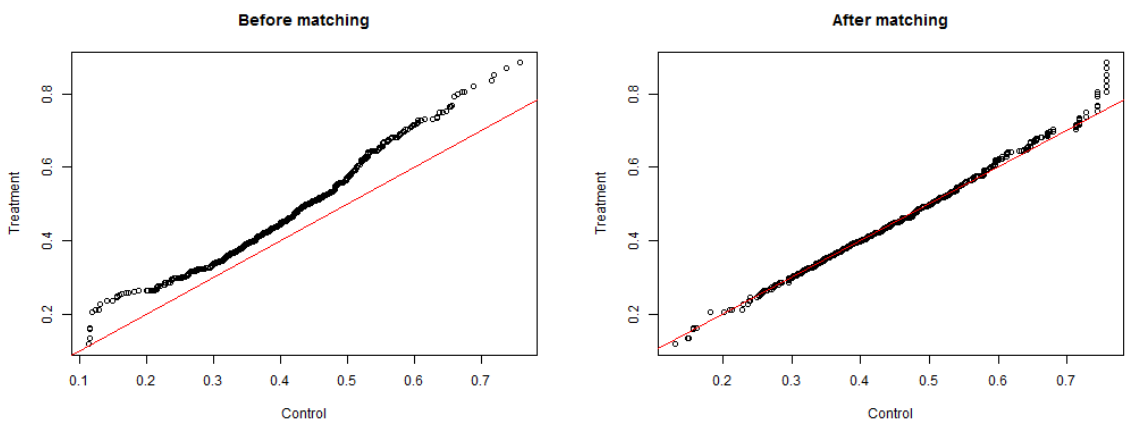
(a) (b)



Note: Author's elaboration. Quantile-quantile plot to evaluate balance before and after matching using the trimmed sample. Panel (a) presents the results before matching. Panel (b) presents the results after matching.

Figure S2: Quantile-quantile plot for the propensity score in the trimmed sample

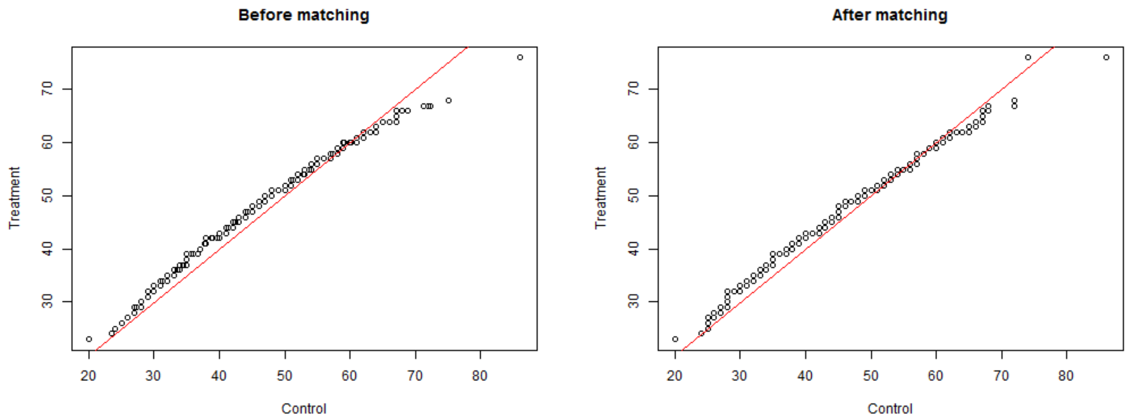
(a) (b)



Note: Author's elaboration. Quantile-quantile plot to evaluate balance before and after matching using the trimmed sample. Panel (a) presents the results before matching. Panel (b) presents the results after matching.

Figure S3: Quantile-quantile plot for participant's age in the trimmed sample

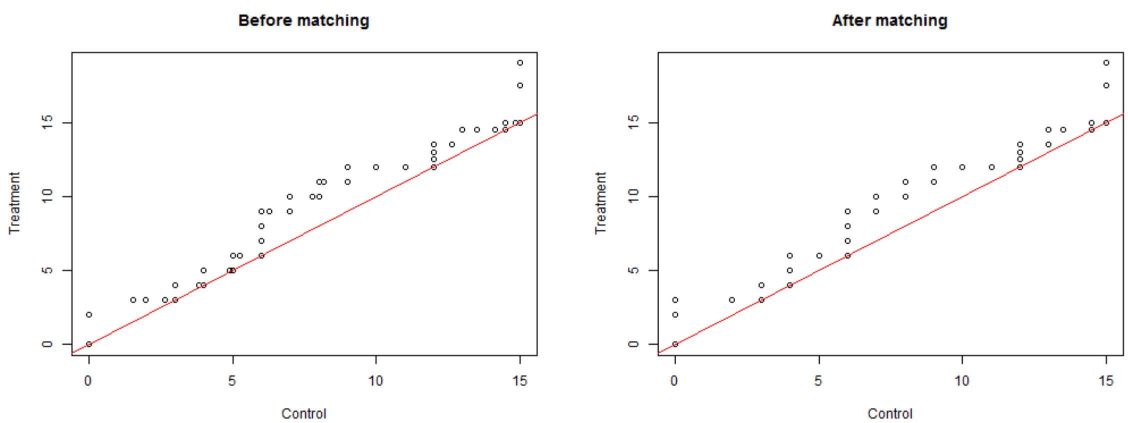
(a) (b)



Note: Author's elaboration. Quantile-quantile plot to evaluate balance before and after matching using the trimmed sample. Panel (a) presents the results before matching. Panel (b) presents the results after matching.

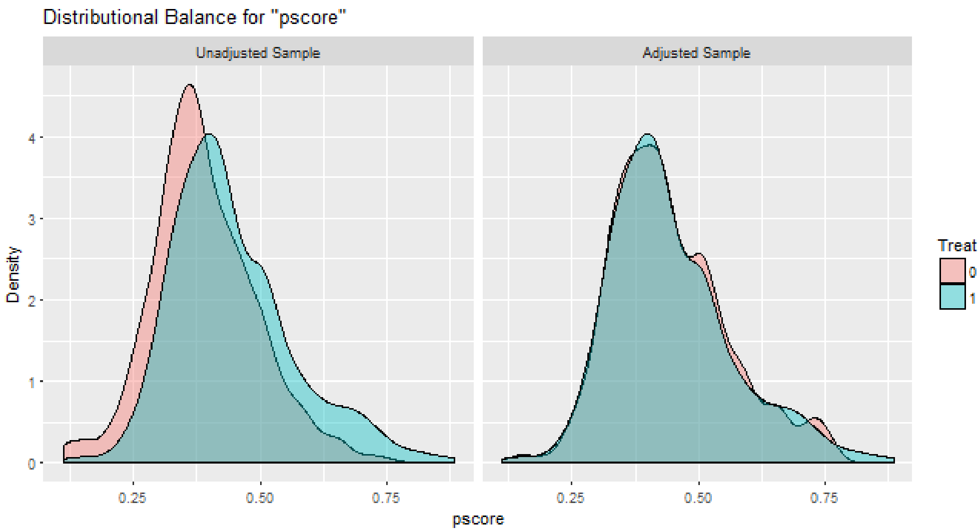
Figure S4: Quantile-quantile plot for participant's years of education in the trimmed sample

(a) (b)



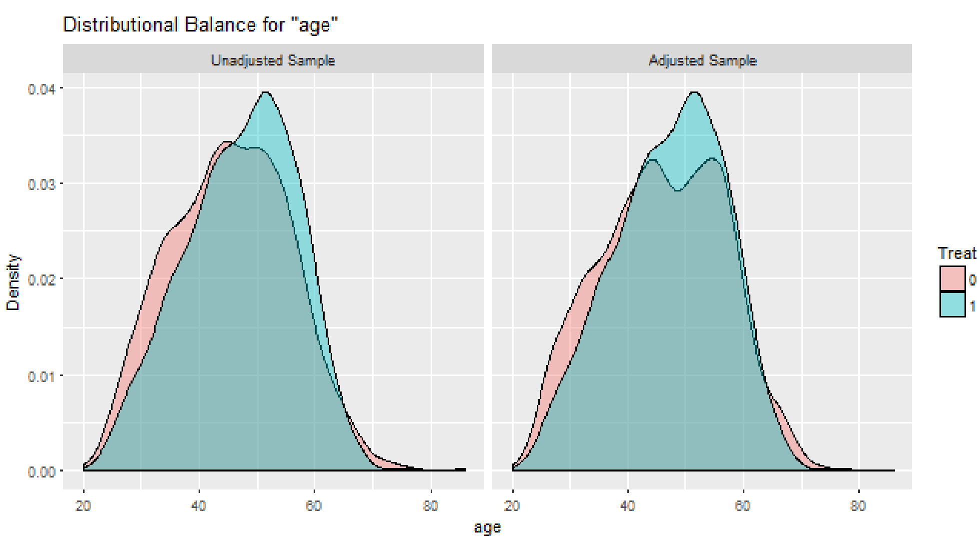
Note: Author's elaboration. Quantile-quantile plot to evaluate balance before and after matching using the trimmed sample. Panel (a) presents the results before matching. Panel (b) presents the results after matching.

Figure S5: Assessing distributional balance of the propensity score before and after matching



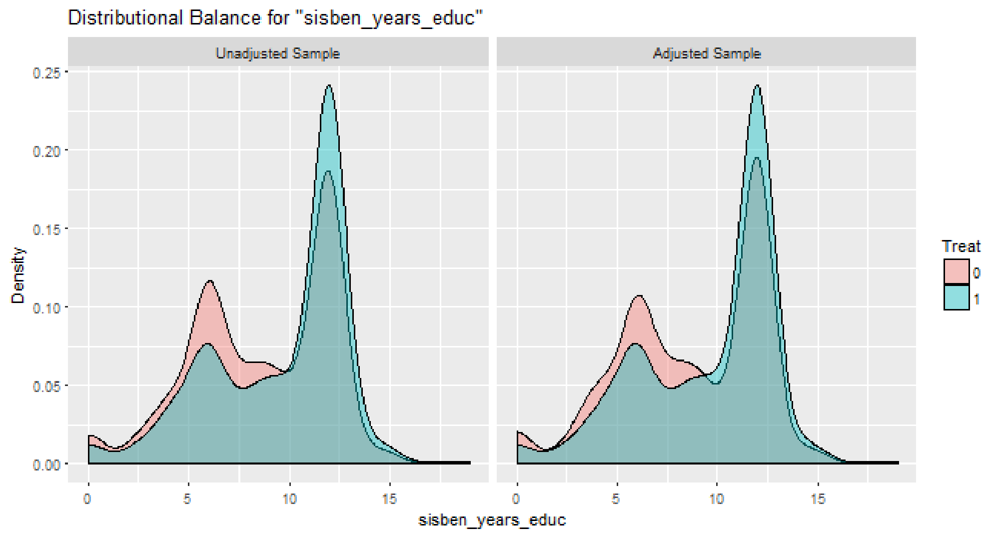
Note: Author's elaboration. The figure displays the difference in the distribution of the pre-treatment variable of interest before and after matching for the trimmed sample.

Figure S6: Assessing distributional balance of participant's age before and after matching



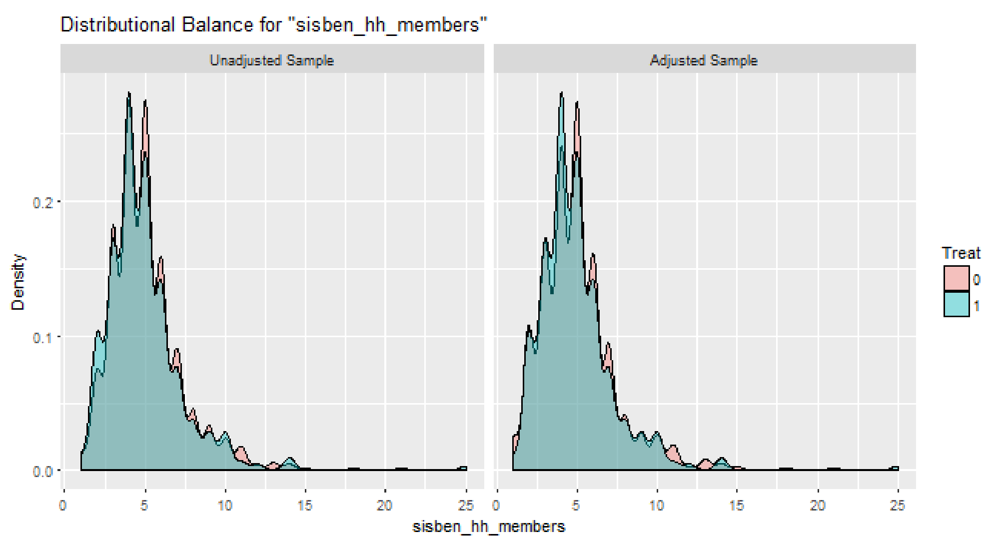
Note: Author's elaboration. The figure displays the difference in the distribution of the pre-treatment variable of interest before and after matching for the trimmed sample.

Figure S7: Assessing distributional balance of participant's years of education before and after matching



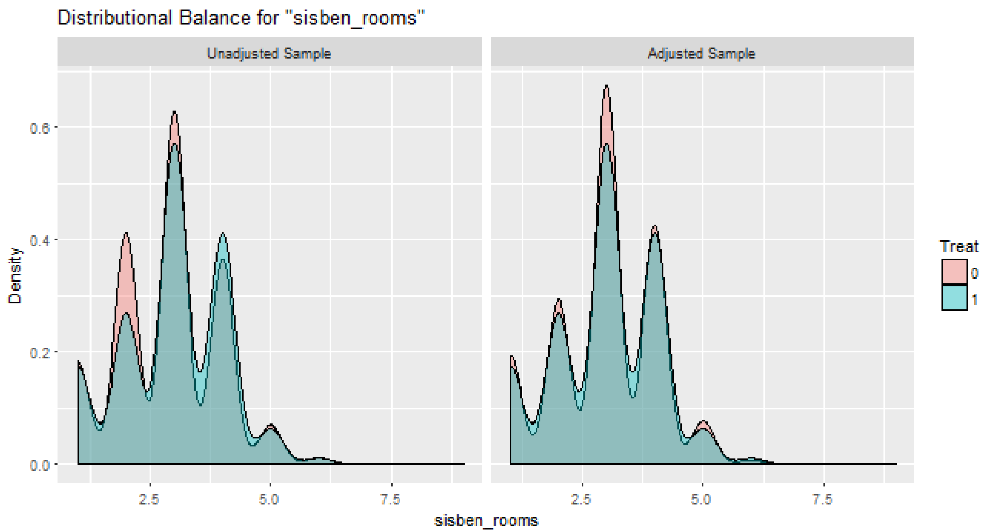
Note: Author's elaboration. The figure displays the difference in the distribution of the pre-treatment variable of interest before and after matching for the trimmed sample.

Figure S8: Assessing distributional balance of participant's number of household members before and after matching



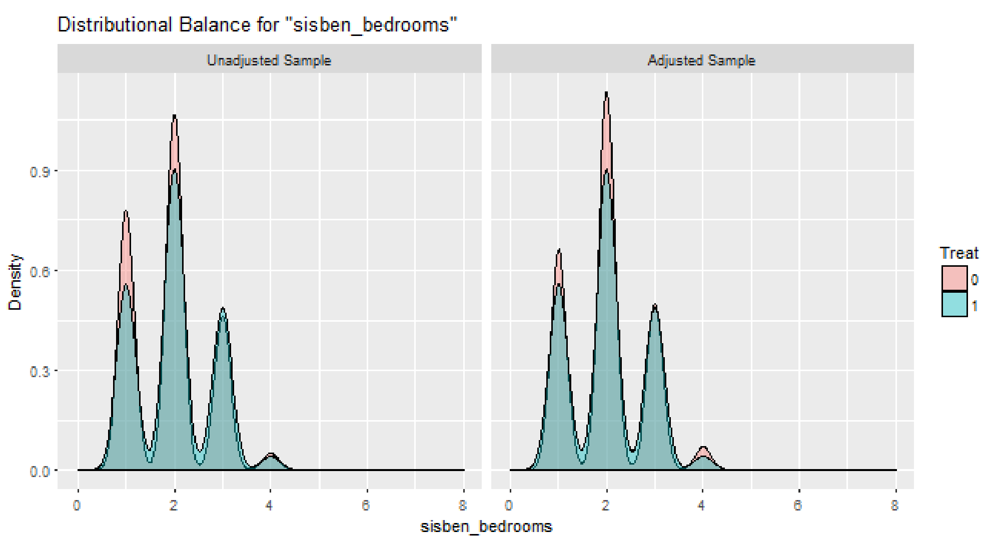
Note: Author's elaboration. The figure displays the difference in the distribution of the pre-treatment variable of interest before and after matching for the trimmed sample.

Figure S9: Assessing distributional balance of participant's number of rooms in the house before and after matching



Note: Author's elaboration. The figure displays the difference in the distribution of the pre-treatment variable of interest before and after matching for the trimmed sample.

Figure S10: Assessing distributional balance of participant's number of bedrooms in the house before and after matching



Note: Author's elaboration. The figure displays the difference in the distribution of the pre-treatment variable of interest before and after matching for the trimmed sample.

Table S1. Full list of variables in the sampling frame

Variables	Description
Household	
House	If dwelling is a house or an apartment.
Walls	If the material of dwelling walls are block, brick, stone, etc
Electric power	If the dwelling has public electric power services.
Sewerage	If the dwelling has public sewer services.
Gas	If the dwelling has public gas services.
Telephone	If the dwelling has public telephone services.
Aqueduct	If the dwelling has public aqueduct services.
Assets	
Fridge	If household has fridge.
Washing machine	If household has washing machine.
Television	If household has television
TV-cable	If household has TV-cable.
Water heater	If household has water heater.
Microwave	If household has microwave.
Air conditioning	If household has air conditioning.
Computer	If household has computer.
Stereo	If household has stereo.
Motorcycle	If household has motorcycle.
Car	If household has car.
Tractor	If household has tractor.
Socio-demographic characteristics	
Married	If woman is married.
Disability	If woman has any disability.
Health	If woman is covered by health insurance.
Without education	If woman doesn't have any educational level.
Basic education	If woman reached a basic level of education.
Inactive	If woman doesn't have an activity in the last month.
SISBEN score	Real part of SISBEN score
Square SISBEN score	SISBEN score squared

Note: Author's elaboration based on SISBEN data for 2011. This table reports the definitions and descriptions of the variables used for the construction of the sampling frame.

Table S2. Full list of variables for the estimation of the propensity score in the sampling frame

Variables	Description
Household	
House	If dwelling is a house or an apartment.
Gas	If the dwelling has public gas services.
Telephone	If the dwelling has public telephone services.
Aqueduct	If the dwelling has public aqueduct services.
Assets	
Fridge	If household has fridge.
Washing machine	If household has washing machine.
Television	If household has TV-color.
TV-cable	If household has TV-cable.
Computer	If household has computer.
Stereo	If household has stereo.
Motorcycle	If household has motorcycle.
Socio-demographic characteristics	
Married	If woman is married.
Age	Woman age
Health	If woman is covered by health insurance.
Basic education	If woman reached a basic level of education.
Inactive	If woman doesn't have an activity in the last month.
SISBEN score	Real part of SISBEN score
Square SISBEN score	SISBEN score squared

Note: Author's elaboration on SISBEN data for 2011. This table reports the definitions and descriptions of the variables used for the estimation of the propensity score in the sampling frame.

Table S3. Full list of variables for the estimation of the propensity score in the sample

Variables	Description
Household	
House	If dwelling is a house or an apartment.
Gas	If the dwelling has public gas services.
Telephone	If the dwelling has public telephone services.
Aqueduct	If the dwelling has public aqueduct services.
Assets	
Fridge	If household has fridge.
Washing machine	If household has washing machine.
Television	If household has TV-color.
TV-cable	If household has TV-cable.
Computer	If household has computer.
Stereo	If household has stereo.
Motorcycle	If household has motorcycle.
Socio-demographic characteristics	
Married	If woman is married.
Afro	If woman recognizes herself as black or afro
Health	If woman is covered by health insurance.
Basic education	If woman reached a basic level of education.
Inactive	If woman doesn't have an activity in the last month.
SISBEN score	Real part of SISBEN score

Note: Author's elaboration based on survey data collected in 2017 for the retrospective evaluation of "Transformate Tu Mujer". This table reports the definitions and descriptions of the variables used for the estimation of the propensity score in the sample.

Table S4. Estimated parameters of propensity score using Imbens and Rubin's (2015) algorithm

Variable	Estimate	S.E	Z	p-value
Panel A: Linear terms				
Intercept	11.437	381.402	0.030	0.976
House	0.331	1.406	0.240	0.814
Aqueduct	-14.383	381.403	-0.040	0.970
Gas	0.400	0.359	1.110	0.265
Telephone	-0.072	0.185	-0.390	0.698
Afro	0.192	0.578	0.330	0.740
Basic education	-13.052	381.402	-0.030	0.973
Health	-0.190	0.121	-1.560	0.118
Married	-0.289	0.116	-2.490	0.013
Inactive	-0.252	0.183	-1.370	0.170
SISBEN Score	0.126	0.035	3.600	0.000
Panel B: Additional linear terms				
Fridge	0.202	0.115	1.760	0.079
Television	-0.220	0.152	-1.440	0.150
Panel C: Second order terms				
SISBEN Score X House	-0.067	0.024	-2.780	0.005
Aqueduct X House	3.087	0.823	3.750	0.000
SISBEN Score X Aqueduct	-0.080	0.029	-2.750	0.006
Basic education X Aqueduct	14.698	381.403	0.040	0.969
Inactive X Afro	0.898	0.429	2.090	0.036
Afro X Aqueduct	-0.911	0.445	-2.050	0.040
Afro X Telephone	0.876	0.412	2.130	0.033
SISBEN Score X SISBEN Score	0.000	0.000	1.870	0.062
Basic education X House	-1.693	1.114	-1.520	0.129
Gas X House	-0.676	0.400	-1.690	0.091

Note: Author's elaboration. This table reports the estimated coefficients, standard errors, z-statistics and p-values for the estimation of the propensity score following the algorithm described in Imbens and Rubin (2015). Panel A includes the original variables proposed by the author to enter linearly in the estimation of the algorithm. Panel B shows the additional linear variables selected by the algorithm from a subset of household-level asset variables. Panel C presents the interactions selected by the algorithm from the former linear variables.

Table S5. Full list of variables for multivariate genetic matching in the sample

Assets	
Fridge	If household has fridge.
Washing machine	If household has washing machine.
Television	If household has TV-color.
TV-cable	If household has TV-cable.
Computer	If household has computer.
Stereo	If household has stereo.
Motorcycle	If household has motorcycle.
Socio-demographic characteristics	
Married	If woman is married.
Basic education	If woman reached a basic level of education.
SISBEN score	Real part of SISBEN score
Propensity score	Imbens and Rubin's (2015) propensity score

Note: Author's elaboration. This table reports the names and descriptions of the variables used in the implementation of the multivariate genetic matching in the sample.

Table S6. List experiment on intimate partner violence

Text: Next, I will read some situations that happen to people with some frequency. Please tell me how many of those situations have happened to you in the last year. You should not tell me WHICH these situations are, but simply HOW MANY of them apply to you.

Condition 1	
Item 1	Attended social events such as parties and meetings with friends.
Item 2	Take care of a friend, neighbor or acquaintance who was ill.
Item 3	Physically beaten by your partner.
Item 4	Asked for a loan to buy a house.
Item 5	Went on vacation to another city.
Condition 2	
Item 1	Attended social events such as parties and meetings with friends.
Item 2	Take care of a friend, neighbor or acquaintance who was ill.
Item 3	NOT INCLUDED.
Item 4	Asked for a loan to buy a house.
Item 5	Went on vacation to another city.

Note: Author's elaboration. This table reproduces question 15 from the survey questionnaire for the retrospective evaluation of "Transformate Tu Mujer". Conditions 1 and 2 were randomized across surveyed women being the only difference the presence of item 3 regarding exposure to physical violence in the previous year.

Table S7. Effects on attitudes towards women rights
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
Decide:								
1. Kind of family	0.963	-0.007	0.011	-0.664	0.507	0.035	0	-0.008
2. Be a mother	0.992	-0.004	0.007	-0.600	0.548	0.040	0	-0.004
3. Number of children	0.977	-0.003	0.009	-0.337	0.736	0.050	0	-0.003
4. Use contraceptive methods	0.953	0.016	0.012	1.288	0.198	0.025	0	0.017
5. Interrupt pregnancy in case of rape	0.535	-0.036	0.030	-1.217	0.224	0.030	0	-0.068
6. Enjoy sexuality with freedom and without violence	0.976	-0.018	0.010	-1.759	0.079	0.005	0	-0.019
7. Choose couple	0.994	-0.009	0.006	-1.497	0.135	0.015	0	-0.010
8. Prevent pregnancy	0.964	-0.006	0.012	-0.506	0.613	0.045	0	-0.006
9. Express sexual orientation	0.927	-0.024	0.018	-1.327	0.185	0.020	0	-0.026
10. Refuse to have sex with partner	0.862	-0.037	0.022	-1.701	0.089	0.010	0	-0.043
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1456

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table S8. Effects of TTM on time use in household chores
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
1. Hand-washing	0.579	0.019	0.029	0.654	0.513	0.025	0	0.033
2. Machine-washing	0.685	0.007	0.028	0.252	0.801	0.046	0	0.010
3. Cooking	0.789	0.009	0.024	0.397	0.691	0.033	0	0.012
4. House cleaning	0.676	0.017	0.028	0.613	0.540	0.029	0	0.026
5. Dish-washing	0.702	-0.010	0.027	-0.368	0.713	0.038	0	-0.014
6. Shopping	0.666	0.029	0.029	1.024	0.306	0.021	0	0.044
7. Children care	0.385	-0.043	0.029	-1.493	0.135	0.017	0	-0.111
8. Participation in children's activities	0.287	0.010	0.027	0.353	0.724	0.042	0	0.034
9. Appointment with children's doctor	0.381	0.007	0.029	0.251	0.802	0.050	0	0.019
10. Care of family members	0.108	0.078	0.021	3.622	0.000	0.004	1	0.720
11. Productive activities in own farm	0.006	-0.008	0.005	-1.572	0.116	0.013	0	-1.282
12. Productive activities in third party farms	0.001	0.008	0.004	1.828	0.068	0.008	0	8.012
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1456

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table S9. Effects of TTM on housing quality
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
1. Home-ownership	0.567	0.037	0.029	1.259	0.208	0.042	0	0.065
2. Noble materials walls	0.951	0.020	0.008	2.539	0.011	0.008	0	0.021
3. Noble materials floors	0.962	0.014	0.009	1.572	0.116	0.025	0	0.015
4. Toilet	0.971	0.011	0.008	1.361	0.174	0.033	0	0.012
5. Gas stove	0.960	0.012	0.011	1.070	0.285	0.050	0	0.012
6. Piped water connection	0.936	-0.035	0.017	-2.107	0.035	0.017	0	-0.038
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1452

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table S10. Effects of TTM on food security
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
1. Worry about lack of food	0.843	-0.054	0.024	-2.287	0.022	0.010	0	-0.064
2. Food scarcity	0.494	0.020	0.030	0.654	0.513	0.050	0	0.040
3. Scarcity of healthy meals	0.555	-0.064	0.030	-2.148	0.032	0.020	0	-0.116
4. Lack of food variety	0.567	-0.043	0.030	-1.433	0.152	0.040	0	-0.075
5. No having a meal in a day	0.369	-0.048	0.028	-1.671	0.095	0.030	0	-0.129
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1452

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table S11. Effects of TTM on the use of social programs
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Mean Control (1)	Estimate (2)	S.E. (3)	t-stat (4)	p-value (5)	Benjamini and Hochberg (6)	Reject Ho (7)	Effect Size (8)
1. Families in Action	0.348	0.054	0.027	1.985	0.047	0.021	0	0.156
2. Red Unidos	0.035	0.008	0.012	0.704	0.481	0.036	0	0.237
3. Adulto Mayor	0.082	0.038	0.016	2.355	0.019	0.014	0	0.467
4. Ser Pilo Paga	0.004	-0.003	0.004	-0.693	0.489	0.043	0	-0.644
5. Youth in Action	0.036	-0.002	0.011	-0.190	0.850	0.050	0	-0.057
6. Free housing	0.015	-0.007	0.007	-1.019	0.308	0.029	0	-0.445
7. Subsidies for displaced populations	0.464	-0.154	0.029	-5.362	0.000	0.007	1	-0.332
Observations								1723
Treated observations								709
Matched observations								709
Matched observations (unweighted)								1452

Note: This table reports the results of implementing the bias-adjusted nearest neighbor matching estimator proposed by Abadie and Imbens (2006, 2011) to estimate the average treatment effect on the treated (ATT). Balance was achieved using the multivariate genetic matching algorithm proposed by Diamond and Sekhon (2013) using the variables described in Table S5 in the Online Appendix. The genetic matching algorithm was based on a population size of 1,000 and a number of neighbors equal to 2. The sample was trimmed following Crump et al. (2009). Multiple outcomes are corrected by adjusting for the false discovery rate as suggested by Benjamini and Hochberg (1995). In this procedure, the unadjusted p-values $P_{(i)}$ are used to create a rank i for a number of m outcomes. The proportion q of the rejected null hypotheses that are erroneously rejected is set at 5%. Dependent variables are reported in rows. Column 1 presents the outcome mean for the control group and column 2 the causal effect estimate. Columns 3, 4 and 5 report the standard error, the t-statistics and the associated unadjusted p-value respectively. Column 6 shows the Benjamini and Hochberg factor ($\frac{i}{m}q$) to adjust for multiple outcomes and column 7 informs whether the null hypothesis is rejected after this adjustment ($P_{(i)} \leq \frac{i}{m}q$). Column 8 reports the effect size.

Table S12. Rosenbaum's sensitivity tests for economic effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Monthly income		Monthly expenditure		Poverty		Employed		Health coverage	
Unconfounded p-value	0.000		0.001		0.000		0.000		0.001	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.001
1.1	0.000	0.000	0.000	0.074	0.000	0.000	0.000	0.000	0.000	0.003
1.2	0.000	0.000	0.000	0.493	0.000	0.000	0.000	0.000	0.000	0.009
1.3	0.000	0.001	0.000	0.903	0.000	0.007	0.000	0.000	0.000	0.024
1.4	0.000	0.017	0.000	0.994	0.000	0.067	0.000	0.000	0.000	0.050
1.5	0.000	0.149	0.000	1.000	0.000	0.267	0.000	0.000	0.000	0.092
1.6	0.000	0.487	0.000	1.000	0.000	0.575	0.000	0.000	0.000	0.151
1.7	0.000	0.819	0.000	1.000	0.000	0.829	0.000	0.000	0.000	0.225
1.8	0.000	0.964	0.000	1.000	0.000	0.952	0.000	0.000	0.000	0.309
1.9	0.000	0.996	0.000	1.000	0.000	0.990	0.000	0.000	0.000	0.400
2	0.000	1.000	0.000	1.000	0.000	0.999	0.000	0.002	0.000	0.490

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S13. Rosenbaum's sensitivity tests for economic effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Pension coverage**		**Fridge**		**DVD**		**Motorcycle**		**Car**		**Washing machine**	
Unconfounded p-value	0.0226		0.0035		0.0047		0.2957		0.220		0.061	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1	0.023	0.023	0.004	0.004	0.005	0.005	0.296	0.296	0.220	0.220	0.061	0.061
1.1	0.003	0.112	0.000	0.029	0.000	0.069	0.077	0.637	0.114	0.367	0.002	0.386
1.2	0.000	0.308	0.000	0.120	0.000	0.322	0.013	0.877	0.055	0.521	0.000	0.805
1.3	0.000	0.561	0.000	0.305	0.000	0.682	0.001	0.972	0.025	0.661	0.000	0.972
1.4	0.000	0.776	0.000	0.542	0.000	0.909	0.000	0.995	0.011	0.773	0.000	0.998
1.5	0.000	0.907	0.000	0.750	0.000	0.984	0.000	0.999	0.005	0.855	0.000	1.000
1.6	0.000	0.967	0.000	0.885	0.000	0.998	0.000	1.000	0.002	0.911	0.000	1.000
1.7	0.000	0.990	0.000	0.955	0.000	1.000	0.000	1.000	0.001	0.947	0.000	1.000
1.8	0.000	0.997	0.000	0.985	0.000	1.000	0.000	1.000	0.000	0.969	0.000	1.000
1.9	0.000	0.999	0.000	0.995	0.000	1.000	0.000	1.000	0.000	0.983	0.000	1.000
2	0.000	1.000	0.000	0.999	0.000	1.000	0.000	1.000	0.000	0.990	0.000	1.000

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S14. Rosenbaum's sensitivity test of labor effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Belonging to the EAP**		**Number workers in the HH**		**Fraction workers in the HH**		**Number formal workers in the HH**		**Fraction formal workers in the HH**	
Unconfounded p-value	0.000		0.000		0.000		0.228		0.220	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.000	0.000	0.000	0.000	0.000	0.000	0.228	0.228	0.220	0.220
1.1	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.706	0.017	0.717
1.2	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.957	0.000	0.964
1.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.998	0.000	0.998
1.4	0.000	0.000	0.000	0.000	0.000	0.001	0.000	1.000	0.000	1.000
1.5	0.000	0.000	0.000	0.001	0.000	0.019	0.000	1.000	0.000	1.000
1.6	0.000	0.000	0.000	0.016	0.000	0.133	0.000	1.000	0.000	1.000
1.7	0.000	0.000	0.000	0.095	0.000	0.423	0.000	1.000	0.000	1.000
1.8	0.000	0.000	0.000	0.298	0.000	0.748	0.000	1.000	0.000	1.000
1.9	0.000	0.000	0.000	0.581	0.000	0.931	0.000	1.000	0.000	1.000
2.0	0.000	0.001	0.000	0.817	0.000	0.988	0.000	1.000	0.000	1.000

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S15. Rosenbaum's sensitivity test of labor effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	Unpaid family worker		Realization of paid activity		Realization of bussines activity		Job search		Work hours	
Unconfounded p-value	0.000		0.000		0.256		0.000		0.000	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.000	0.000	0.000	0.000	0.256	0.256	0.000	0.000	0.000	0.000
1.1	0.000	0.000	0.000	0.000	0.122	0.442	0.000	0.000	0.000	0.000
1.2	0.000	0.000	0.000	0.000	0.052	0.625	0.000	0.003	0.000	0.000
1.3	0.000	0.000	0.000	0.000	0.020	0.772	0.000	0.013	0.000	0.000
1.4	0.000	0.000	0.000	0.000	0.007	0.873	0.000	0.042	0.000	0.000
1.5	0.000	0.000	0.000	0.000	0.002	0.934	0.000	0.103	0.000	0.007
1.6	0.000	0.001	0.000	0.000	0.001	0.967	0.000	0.203	0.000	0.068
1.7	0.000	0.007	0.000	0.000	0.000	0.985	0.000	0.335	0.000	0.277
1.8	0.000	0.044	0.000	0.000	0.000	0.993	0.000	0.481	0.000	0.601
1.9	0.000	0.157	0.000	0.000	0.000	0.997	0.000	0.622	0.000	0.856
2.0	0.000	0.364	0.000	0.000	0.000	0.999	0.000	0.741	0.000	0.966

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S16. Rosenbaum's sensitivity tests of credit and saving effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

	Household Savings		**Savings (interviewed woman)**		**Saving in banks or corporations**		**Savings in microcredit institution**		**Saving at home**	
Unconfounded p-value	0.000		0.000		0.097		0.500		0.000	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.000	0.000	0.000	0.000	0.097	0.097	0.500	0.500	0.000	0.000
1.1	0.000	0.000	0.000	0.000	0.044	0.186	0.421	0.579	0.000	0.000
1.2	0.000	0.007	0.000	0.000	0.019	0.300	0.352	0.648	0.000	0.000
1.3	0.000	0.065	0.000	0.000	0.008	0.426	0.292	0.708	0.000	0.000
1.4	0.000	0.258	0.000	0.000	0.003	0.550	0.241	0.759	0.000	0.000
1.5	0.000	0.560	0.000	0.001	0.001	0.661	0.199	0.801	0.000	0.001
1.6	0.000	0.815	0.000	0.012	0.001	0.753	0.164	0.836	0.000	0.006
1.7	0.000	0.944	0.000	0.057	0.000	0.826	0.135	0.865	0.000	0.029
1.8	0.000	0.988	0.000	0.173	0.000	0.880	0.111	0.889	0.000	0.097
1.9	0.000	0.998	0.000	0.368	0.000	0.919	0.091	0.909	0.000	0.231
2.0	0.000	1.000	0.000	0.591	0.000	0.946	0.075	0.925	0.000	0.419

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S17. Rosenbaum's sensitivity tests of credit and saving effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Savings in informal scheme**		**Debt holdings**		**Lender - Bank**		**Lender - Microcredit**		**Lender - Producers**	
Unconfounded p-value	0.073		0.028		0.005		0.003		0.416	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.073	0.073	0.028	0.028	0.005	0.005	0.003	0.003	0.416	0.416
1.1	0.053	0.099	0.001	0.259	0.000	0.068	0.001	0.012	0.331	0.505
1.2	0.038	0.128	0.000	0.695	0.000	0.302	0.000	0.033	0.260	0.587
1.3	0.028	0.159	0.000	0.942	0.000	0.645	0.000	0.073	0.202	0.659
1.4	0.021	0.192	0.000	0.995	0.000	0.884	0.000	0.134	0.156	0.720
1.5	0.015	0.225	0.000	1.000	0.000	0.975	0.000	0.218	0.121	0.772
1.6	0.011	0.260	0.000	1.000	0.000	0.996	0.000	0.317	0.093	0.815
1.7	0.009	0.294	0.000	1.000	0.000	1.000	0.000	0.423	0.071	0.850
1.8	0.007	0.328	0.000	1.000	0.000	1.000	0.000	0.529	0.055	0.879
1.9	0.005	0.361	0.000	1.000	0.000	1.000	0.000	0.626	0.042	0.902
2.0	0.004	0.393	0.000	1.000	0.000	1.000	0.000	0.712	0.033	0.921

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S18. Rosenbaum's sensitivity tests for empowerment effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Woman**		**Both**		**Vote in 2016**		**Membership in political parties**		**Candidate**	
Unconfounded p-value	0.001		0.011		0.000		0.000		0.000	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.001	0.001	0.010	0.010	0.000	0.000	0.000	0.000	0.000	0.000
1.1	0.000	0.018	0.000	0.138	0.000	0.002	0.000	0.000	0.000	0.000
1.2	0.000	0.142	0.000	0.508	0.000	0.043	0.000	0.000	0.000	0.000
1.3	0.000	0.448	0.000	0.850	0.000	0.250	0.000	0.000	0.000	0.000
1.4	0.000	0.770	0.000	0.976	0.000	0.613	0.000	0.001	0.000	0.000
1.5	0.000	0.938	0.000	0.998	0.000	0.880	0.000	0.003	0.000	0.000
1.6	0.000	0.989	0.000	1.000	0.000	0.978	0.000	0.011	0.000	0.000
1.7	0.000	0.999	0.000	1.000	0.000	0.997	0.000	0.029	0.000	0.000
1.8	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.063	0.000	0.000
1.9	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.117	0.000	0.001
2.0	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.193	0.000	0.003

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S19. Rosenbaum's sensitivity tests for empowerment effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Political Participation**		**Women's participation in the community**		**Disagree with violence against women**		**Woman as agent of change**		**Self-confidence**	
Unconfounded p-value	0.000		0.004		0.005		0.396		0.396	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.000	0.000	0.004	0.004	0.005	0.005	0.396	0.396	0.396	0.396
1.1	0.000	0.000	0.001	0.015	0.001	0.024	0.122	0.738	0.122	0.738
1.2	0.000	0.000	0.000	0.046	0.000	0.078	0.023	0.928	0.023	0.928
1.3	0.000	0.002	0.000	0.108	0.000	0.180	0.003	0.987	0.003	0.987
1.4	0.000	0.011	0.000	0.203	0.000	0.328	0.000	0.998	0.000	0.998
1.5	0.000	0.040	0.000	0.326	0.000	0.496	0.000	1.000	0.000	1.000
1.6	0.000	0.108	0.000	0.461	0.000	0.653	0.000	1.000	0.000	1.000
1.7	0.000	0.223	0.000	0.591	0.000	0.780	0.000	1.000	0.000	1.000
1.8	0.000	0.377	0.000	0.705	0.000	0.870	0.000	1.000	0.000	1.000
1.9	0.000	0.542	0.000	0.797	0.000	0.928	0.000	1.000	0.000	1.000
2.0	0.000	0.693	0.000	0.865	0.000	0.962	0.000	1.000	0.000	1.000

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S20. Rosenbaum's sensitivity tests of social capital
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Community action board**		**Producers' associations**		**Citizen veedurias**		**Displaced population organizations**		**Volunteering**	
Unconfounded p-value	0.000		0.000		0.000		0.435		0.000	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.00	0.000	0.000	0.000	0.000	0.000	0.000	0.435	0.435	0.000	0.000
1.10	0.000	0.000	0.000	0.000	0.000	0.000	0.228	0.662	0.000	0.000
1.20	0.000	0.000	0.000	0.000	0.000	0.000	0.101	0.829	0.000	0.000
1.30	0.000	0.000	0.000	0.000	0.000	0.000	0.039	0.925	0.000	0.000
1.40	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.970	0.000	0.000
1.50	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.989	0.000	0.000
1.60	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.996	0.000	0.000
1.70	0.000	0.006	0.000	0.000	0.000	0.001	0.000	0.999	0.000	0.000
1.80	0.000	0.022	0.000	0.000	0.000	0.001	0.000	1.000	0.000	0.000
1.90	0.000	0.058	0.000	0.000	0.000	0.002	0.000	1.000	0.000	0.000
2.00	0.000	0.124	0.000	0.000	0.000	0.003	0.000	1.000	0.000	0.000

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S21. Rosenbaum's sensitivity tests of social capital
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Parents' groups**		**Sports or cultural groups**		**Women groups**		**Youth groups**		**Syndicates**	
Unconfounded p-value	0.423		0.000		0.000		0.061		0.010	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.00	0.423	0.423	0.000	0.000	0.000	0.000	0.061	0.061	0.010	0.010
1.10	0.176	0.706	0.000	0.000	0.000	0.000	0.036	0.097	0.005	0.016
1.20	0.055	0.887	0.000	0.000	0.000	0.000	0.022	0.142	0.003	0.026
1.30	0.013	0.966	0.000	0.000	0.000	0.000	0.013	0.193	0.002	0.038
1.40	0.003	0.992	0.000	0.000	0.000	0.000	0.008	0.250	0.001	0.052
1.50	0.000	0.998	0.000	0.000	0.000	0.000	0.005	0.309	0.001	0.070
1.60	0.000	1.000	0.000	0.000	0.000	0.000	0.003	0.369	0.000	0.089
1.70	0.000	1.000	0.000	0.000	0.000	0.000	0.002	0.428	0.000	0.112
1.80	0.000	1.000	0.000	0.000	0.000	0.000	0.001	0.485	0.000	0.136
1.90	0.000	1.000	0.000	0.000	0.000	0.000	0.001	0.539	0.000	0.161
2.00	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.589	0.000	0.188

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S22. Rosenbaum's sensitivity test health and sexual rights
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

	Use of contraceptive methods		Use of natural method		Use of pill		Use of condom	
Unconfounded p-value	0.075		0.500		0.014		0.108	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.075	0.075	0.500	0.500	0.014	0.014	0.108	0.108
1.1	0.003	0.439	0.448	0.552	0.005	0.038	0.066	0.166
1.2	0.000	0.847	0.401	0.599	0.001	0.081	0.039	0.235
1.3	0.000	0.982	0.360	0.640	0.000	0.146	0.024	0.310
1.4	0.000	0.999	0.323	0.677	0.000	0.232	0.014	0.386
1.5	0.000	1.000	0.290	0.710	0.000	0.331	0.008	0.462
1.6	0.000	1.000	0.261	0.739	0.000	0.435	0.005	0.533
1.7	0.000	1.000	0.235	0.765	0.000	0.537	0.003	0.599
1.8	0.000	1.000	0.212	0.788	0.000	0.631	0.002	0.659
1.9	0.000	1.000	0.191	0.809	0.000	0.713	0.001	0.711
2.0	0.000	1.000	0.173	0.827	0.000	0.781	0.001	0.757

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S23. Rosenbaum's sensitivity test health and sexual rights
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Intrauterine device**		**Female sterilization**		**Use of contraceptive injections**		**Cytology**		**Breast exam**	
Unconfounded p-value	0.252		0.102		0.000		0.201		0.000	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.252	0.252	0.102	0.102	0.000	0.000	0.201	0.201	0.000	0.000
1.1	0.188	0.325	0.007	0.472	0.000	0.001	0.057	0.460	0.000	0.000
1.2	0.140	0.398	0.000	0.847	0.000	0.005	0.012	0.717	0.000	0.003
1.3	0.103	0.469	0.000	0.979	0.000	0.014	0.002	0.883	0.000	0.039
1.4	0.076	0.535	0.000	0.998	0.000	0.031	0.000	0.961	0.000	0.204
1.5	0.057	0.596	0.000	1.000	0.000	0.061	0.000	0.989	0.000	0.517
1.6	0.042	0.650	0.000	1.000	0.000	0.105	0.000	0.997	0.000	0.803
1.7	0.031	0.698	0.000	1.000	0.000	0.164	0.000	0.999	0.000	0.946
1.8	0.023	0.741	0.000	1.000	0.000	0.235	0.000	1.000	0.000	0.990
1.9	0.017	0.778	0.000	1.000	0.000	0.316	0.000	1.000	0.000	0.999
2.0	0.013	0.809	0.000	1.000	0.000	0.401	0.000	1.000	0.000	1.000

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S24. Rosenbaum's sensitivity tests of women's rights effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Decide Kind of family**		**Decide Be a mother**		**Decide Number of children**		**Decide use of contraceptive methods**		**Decide interrupt pregnancy in case of rape**	
Unconfounded p-value	0.460		0.298		0.449		0.048		0.048	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.460	0.460	0.298	0.298	0.449	0.449	0.048	0.048	0.048	0.048
1.1	0.283	0.645	0.212	0.398	0.309	0.595	0.015	0.125	0.002	0.340
1.2	0.157	0.789	0.147	0.496	0.201	0.719	0.004	0.248	0.000	0.768
1.3	0.081	0.885	0.101	0.586	0.125	0.813	0.001	0.401	0.000	0.963
1.4	0.039	0.941	0.069	0.665	0.076	0.880	0.000	0.558	0.000	0.997
1.5	0.018	0.972	0.046	0.732	0.044	0.925	0.000	0.696	0.000	1.000
1.6	0.008	0.987	0.031	0.788	0.026	0.955	0.000	0.804	0.000	1.000
1.7	0.003	0.994	0.021	0.834	0.015	0.973	0.000	0.881	0.000	1.000
1.8	0.001	0.997	0.014	0.871	0.008	0.984	0.000	0.930	0.000	1.000
1.9	0.001	0.999	0.009	0.900	0.005	0.991	0.000	0.961	0.000	1.000
2.0	0.000	1.000	0.006	0.922	0.003	0.994	0.000	0.979	0.000	1.000

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S25. Rosenbaum's sensitivity tests of women's rights effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Decide enjoy sexuality with freedom**		**Decide choose couple**		**Decide prevent pregnancy**		**Decide express sexual orientation**		**Decide refusing to have sex with your partner**	
Unconfounded p-value	0.035		0.021		0.500		0.044		0.004	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.035	0.035	0.021	0.021	0.500	0.500	0.044	0.044	0.004	0.004
1.1	0.012	0.085	0.011	0.039	0.309	0.691	0.007	0.172	0.000	0.044
1.2	0.004	0.168	0.006	0.063	0.170	0.830	0.001	0.399	0.000	0.195
1.3	0.001	0.278	0.003	0.095	0.085	0.915	0.000	0.646	0.000	0.466
1.4	0.000	0.403	0.002	0.133	0.040	0.960	0.000	0.831	0.000	0.735
1.5	0.000	0.529	0.001	0.176	0.017	0.983	0.000	0.933	0.000	0.901
1.6	0.000	0.644	0.000	0.224	0.007	0.993	0.000	0.977	0.000	0.972
1.7	0.000	0.741	0.000	0.275	0.003	0.997	0.000	0.993	0.000	0.993
1.8	0.000	0.818	0.000	0.327	0.001	0.999	0.000	0.998	0.000	0.999
1.9	0.000	0.876	0.000	0.380	0.000	1.000	0.000	1.000	0.000	1.000
2.0	0.000	0.917	0.000	0.432	0.000	1.000	0.000	1.000	0.000	1.000

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S26. Rosenbaum's sensitivity test of time use effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Hand-washing**		**Machine-washing**		**Cooking**		**House cleaning**		**Dish-washing**		**Shopping**	
Unconfounded p-value	0.0266		0.500		0.1907		0.1654		0.3833		0.1569	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.027	0.027	0.500	0.500	0.191	0.191	0.165	0.165	0.383	0.383	0.157	0.157
1.1	0.001	0.241	0.119	0.881	0.032	0.542	0.016	0.581	0.078	0.796	0.014	0.570
1.2	0.000	0.663	0.012	0.988	0.003	0.841	0.001	0.899	0.007	0.968	0.001	0.895
1.3	0.000	0.927	0.001	0.999	0.000	0.966	0.000	0.988	0.000	0.997	0.000	0.988
1.4	0.000	0.992	0.000	1.000	0.000	0.995	0.000	0.999	0.000	1.000	0.000	0.999
1.5	0.000	0.999	0.000	1.000	0.000	0.999	0.000	1.000	0.000	1.000	0.000	1.000
1.6	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
1.7	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
1.8	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
1.9	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
2.0	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S27. Rosenbaum's sensitivity test of time use effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Children care**		**Participation in children's activities**		**Appointment with children's doctor**		**Care of family members**		**Productive activities in own farm**		**Productive activities third party farms**	
Unconfounded p-value	0.0029		0.4681		0.4234		0.000		0.0059		0.0017	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.003	0.003	0.468	0.468	0.423	0.423	0.000	0.000	0.006	0.006	0.002	0.002
1.1	0.000	0.064	0.102	0.866	0.077	0.851	0.000	0.000	0.003	0.011	0.001	0.003
1.2	0.000	0.342	0.009	0.986	0.005	0.985	0.000	0.000	0.002	0.017	0.001	0.004
1.3	0.000	0.733	0.000	0.999	0.000	0.999	0.000	0.000	0.001	0.026	0.000	0.007
1.4	0.000	0.942	0.000	1.000	0.000	1.000	0.000	0.003	0.001	0.037	0.000	0.009
1.5	0.000	0.993	0.000	1.000	0.000	1.000	0.000	0.016	0.000	0.051	0.000	0.013
1.6	0.000	0.999	0.000	1.000	0.000	1.000	0.000	0.058	0.000	0.067	0.000	0.017
1.7	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.154	0.000	0.085	0.000	0.021
1.8	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.307	0.000	0.106	0.000	0.026
1.9	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.492	0.000	0.128	0.000	0.032
2.0	0.000	1.000	0.000	1.000	0.000	1.000	0.000	0.669	0.000	0.152	0.000	0.039

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S28. Rosenbaum's sensitivity tests of housing quality effects
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Home-ownership**		**Noble material walls**		**Noble material floors**		**Toilet**		**Gas stove**		**Piped water connection**	
Unconfounded p-value	0.032		0.000		0.014		0.140		0.240		0.000	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.032	0.032	0.000	0.000	0.014	0.014	0.140	0.140	0.240	0.240	0.000	0.000
1.1	0.001	0.265	0.000	0.001	0.005	0.034	0.076	0.234	0.119	0.407	0.000	0.000
1.2	0.000	0.689	0.000	0.001	0.002	0.071	0.039	0.344	0.054	0.578	0.000	0.002
1.3	0.000	0.936	0.000	0.003	0.001	0.126	0.020	0.458	0.022	0.723	0.000	0.011
1.4	0.000	0.993	0.000	0.007	0.000	0.199	0.010	0.566	0.009	0.831	0.000	0.039
1.5	0.000	1.000	0.000	0.013	0.000	0.285	0.005	0.663	0.003	0.902	0.000	0.102
1.6	0.000	1.000	0.000	0.023	0.000	0.378	0.002	0.744	0.001	0.946	0.000	0.208
1.7	0.000	1.000	0.000	0.036	0.000	0.473	0.001	0.810	0.000	0.972	0.000	0.350
1.8	0.000	1.000	0.000	0.054	0.000	0.563	0.001	0.861	0.000	0.985	0.000	0.507
1.9	0.000	1.000	0.000	0.076	0.000	0.644	0.000	0.900	0.000	0.993	0.000	0.654
2.0	0.000	1.000	0.000	0.104	0.000	0.716	0.000	0.928	0.000	0.996	0.000	0.774

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S29. Rosenbaum's sensitivity tests of food security
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Worry about food**		**Food scarcity**		**Scarcity of healthy meals**		**Lack of food variety**		**No having a meal in a day**	
Unconfounded p-value	0.000		0.071		0.001		0.023		0.013	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.000	0.000	0.071	0.071	0.001	0.001	0.023	0.023	0.013	0.013
1.1	0.000	0.004	0.003	0.418	0.000	0.028	0.001	0.230	0.000	0.148
1.2	0.000	0.040	0.000	0.828	0.000	0.226	0.000	0.659	0.000	0.516
1.3	0.000	0.180	0.000	0.977	0.000	0.619	0.000	0.929	0.000	0.850
1.4	0.000	0.440	0.000	0.999	0.000	0.899	0.000	0.993	0.000	0.975
1.5	0.000	0.712	0.000	1.000	0.000	0.985	0.000	1.000	0.000	0.998
1.6	0.000	0.889	0.000	1.000	0.000	0.999	0.000	1.000	0.000	1.000
1.7	0.000	0.967	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
1.8	0.000	0.992	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
1.9	0.000	0.998	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
2.0	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S30. Rosenbaum's sensitivity test for use of social programs
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Families in Action**		**Red Unidos**		**Adulto Mayor**		**Ser Pilo Paga**	
Unconfounded p-value	0.000		0.167		0.000		0.274	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.000	0.000	0.167	0.167	0.000	0.000	0.274	0.274
1.1	0.000	0.003	0.072	0.318	0.000	0.000	0.223	0.331
1.2	0.000	0.046	0.028	0.491	0.000	0.004	0.182	0.385
1.3	0.000	0.234	0.010	0.651	0.000	0.016	0.148	0.438
1.4	0.000	0.565	0.003	0.779	0.000	0.053	0.121	0.487
1.5	0.000	0.838	0.001	0.869	0.000	0.127	0.099	0.533
1.6	0.000	0.960	0.000	0.927	0.000	0.244	0.082	0.575
1.7	0.000	0.993	0.000	0.961	0.000	0.390	0.067	0.614
1.8	0.000	0.999	0.000	0.980	0.000	0.545	0.056	0.650
1.9	0.000	1.000	0.000	0.990	0.000	0.684	0.046	0.682
2.0	0.000	1.000	0.000	0.995	0.000	0.795	0.039	0.711

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).

Table S31. Rosenbaum's sensitivity test for use of social programs
Bias-adjusted nearest neighbor matching estimator for the ATT in the trimmed sample

Variable	**Youth in Action**		**Free housing**		**Subsidies for displaced populations**	
Unconfounded p-value	0.459		0.045		0.000	
Γ	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
1.0	0.459	0.459	0.045	0.045	0.000	0.000
1.1	0.286	0.641	0.024	0.078	0.000	0.000
1.2	0.161	0.783	0.013	0.123	0.000	0.000
1.3	0.084	0.879	0.007	0.177	0.000	0.000
1.4	0.041	0.937	0.004	0.239	0.000	0.000
1.5	0.019	0.969	0.002	0.306	0.000	0.000
1.6	0.009	0.985	0.001	0.374	0.000	0.001
1.7	0.004	0.993	0.001	0.442	0.000	0.008
1.8	0.002	0.997	0.000	0.507	0.000	0.045
1.9	0.001	0.999	0.000	0.568	0.000	0.154
2.0	0.000	0.999	0.000	0.624	0.000	0.349

Note: This table reports the results of implementing the sensitivity analysis proposed by Rosenbaum (2002) to compute bounds for p-values under different degrees of violation of the CIA assumption. This degree is captured by the parameter Γ. A value of $\Gamma = \Gamma_0$ implies that a matched unit in the treatment group is Γ_0 times more likely to receive treatment than a matched unit in the control group after controlling for a set X of observable characteristics. P-values under the validity of the CIA assumption are reported below each outcome variable. Lower and upper bounds for p-values for different values of Γ are reported for each outcome variable. A set of values for Γ from 1 to 2 are chosen following Keele (2010).