Microcredit, Family Planning Programs and Contraceptive Behavior: Evidence from a Field Experiment in Ethiopia

Jaikishan Desai
The World Bank

Alessandro Tarozzi
Duke University

October 2009

Abstract

The impact of community-based family planning programs and access to credit on contraceptive use, fertility, and family size preferences has not been established conclusively in the literature. In this paper, we provide additional evidence on the possible effect of such programs by describing the results of a randomized field experiment whose main purpose was to increase the use of contraceptive methods in rural areas of Ethiopia. In this study, administrative areas of the Amhara and Oromia regions were randomly allocated to one of three intervention groups or to a fourth control group. In the first intervention group, both credit and family planning services were provided and the credit officers also provided information on family planning. Only credit or family planning services were provided in the other two intervention groups, while areas in the control group received neither type of service. Using pre and post-intervention surveys, we show that neither type of program, combined or in isolation, led to an increase in contraceptive use that is significantly greater than that observed in the control group. We conjecture that the lack of impact has much to do with the mismatch between women's preferred method (injectibles) and the contraceptives provided by community-based agents (pills and condoms).

Keywords: Ethiopia, Family Planning, Fertility, Microcredit, Randomized Controlled Trial

Acknowledgements: Jaikishan Desai was employed by Family Health International during the course of this study. We would like to thank the David and Lucile Packard Foundation for financial support and encouragement for the study, the Packard Foundation in Ethiopia for assistance with coordinating all aspects of the study, Birhan Research and Development Consultancy and Miz Hasab Research Center for conducting the two household surveys, and the four sub-grantee organizations—ACSI, ADA, OCSSCO, and ODA—who, despite several pressures, extended their cooperation in implementing their interventions according to the study design. Last but not least, we are grateful to Laura Chioda, Jed Friedman, Jonathan Robinson, seminar participants at the World Bank and NEUDC (Boston), the Editor and especially two anonymous referees for valuable comments and suggestions. The authors remain solely responsible for all remaining errors and omissions as well as for all the views and interpretations expressed throughout the paper. Desai: jdesai3@worldbank.org; The World Bank, 1818 H St NW, Washington DC 20433, Phone: (202) 458-9499, Fax: (202) 522-7112. Tarozzi: taroz@econ.duke.edu. Department of Economics, Duke University, PO Box 90097, Durham, NC 27708, Tel (919) 660-1877.

1. Introduction

Family planning programs have been adopted in several developing countries with the purpose of increasing contraceptive use, improving reproductive health, lowering fertility, and reducing rates of population growth.[1] One widely used approach has been community-based programs wherein individuals based within a community are trained to provide information on family planning and non-clinical methods like pills and condoms. The active involvement of communities is appealing but the effectiveness of these programs in increasing contraceptive use and lowering fertility has not been demonstrated in a convincing way (Bauman 1997, Freedman 1997).

In this paper, we present results of a randomized field experiment. In the study, administrative areas in the Amhara and Oromia regions of Ethiopia were randomly allocated to one of three intervention groups or to a fourth control group. In the first intervention group, both credit and family planning services were provided and the credit officers also provided information on family planning. In the other two intervention groups, only credit or family planning services were provided. Areas in the control group received neither type of service. The study was designed to determine whether linking micro-credit and family planning programs would increase contraceptive use by more than what could be accomplished by each program on its own. Pre and post-intervention surveys (in 2003 and 2006) of two independent cross-sections of approximately 6400 households were used to collect data on contraceptive use, fertility, and various other outcomes. The study areas are rural, where households are largely subsistence oriented and agriculture and livestock are the main sources of income.

Our data show that the study areas, not unlike Ethiopia as a whole, experienced considerable changes in fertility behavior and socio-economic status during the study period. For instance, we observe large improvements in schooling and in measures of economic welfare as well as significant increases in contraceptive use. However, our experimental results also show that changes in fertility behavior and preferences appear to have been largely unrelated to changes in availability of credit and family planning services. Linking credit and family planning services did not increase contraceptive use any more than what was achieved by either program on its own, and neither type of program, linked or unlinked, led to an increase in contraceptive use significantly greater than that observed in the control group. In addition, the programs appear to have been ineffective at changing reproductive behavior and preferences among women in different age groups, even when we take into account differences in pre-intervention demand for contraceptive use.

The fact that linking family planning and credit provision did not lead to a measurable impact is perhaps not surprising, because the credit programs had limited reach and information does not appear to be the major constraint for adoption of family planning in the study areas.

[1] There is not much consensus, however, on the overall effectiveness of such programs in achieving these goals. On the one hand, the most skeptical view is perhaps exemplified by Pritchett (1994), who argues that access to contraceptives and family planning programs are largely ineffective in changing reproductive behavior, instead "a low level of desired fertility appears to be both necessary and sufficient for low fertility" (p. 39). On the other hand, others have argued for an important, albeit non-dominant, role of family planning programs in reducing fertility in developing countries (see, for instance, Bongaarts 1994). An intermediate position is taken by Freedman (1997), whose literature survey concluded that, while an impact on fertility preferences has rarely been documented convincingly, several family planning services have allowed families to meet existing demand for fertility control.

Our finding that community-based family planning programs have no measurable impact on contraceptive use is more surprising because the coverage and intensity of these programs was remarkable. Our conjecture is that the main reason why these programs did not have a significant impact on contraceptive use is related to a mismatch between the method women were increasingly starting to use—injectibles—and the contraceptives provided by community-based agents (pills and condoms).

The rest of the paper is organized as follows. The next section reviews relevant literature on the relationship between family planning programs, access to credit and fertility-related outcomes. Section 3 describes the details of the intervention and the study design. Section 4 provides details on the data set, outlines the statistical methodology adopted to evaluate the programs, and provides a basic description of the study areas. This section also discusses the fact that in some areas (especially in Amhara) the actual program implementation deviated from the experimental protocol, forcing us to gauge the potential endogeneity of program placement. Our main solution to this problem is adopting two-stage least squares estimation, using the randomly assigned programs as instruments for the actual programs implemented in a given area. Next, Section 5 describes the results of the evaluation and section 6 discusses the findings and their implications as well as threats to identification. Finally, Section 7 concludes.

2. The Impact of Family Planning Programs and Microcredit on Fertility-related Outcomes

Evaluation of the impact of social programs is typically complicated by the fact that both program placement and individual participation are endogenous, that is, correlated with location or individual unobservable characteristics which in turn are correlated with the outcome of interest. Family planning programs (FPPs hereafter) are not an exception, because they are often targeted at high fertility areas and are utilized by self-selecting individuals. Pitt et al. (1993) and Gertler and Molyneaux (1994) use location specific fixed effects to evaluate the impact of FPPs in Indonesia in the 1970s and 1980s while accounting for endogenous program placement. Both studies conclude that fertility was only marginally impacted by the programs, and show that simple cross-sectional estimates would provide misleading results. Miller (2007) shows that the far-reaching Colombian FPP PROFAMILIA expanded in a near-random fashion, and then estimates that it contributed to only 10 percent of the fertility decline observed during the country's demographic transition. Other studies have found substantial effects of FPPs while accounting for the potential endogeneity of program placement and participation. Thomas and Maluccio (2001) find that mobile family planning clinics and community-based distributors led to a substantial increase in contraceptive use in Zimbabwe. Angeles et al. (1998) use a structural model to analyze fertility decisions and the choice of program location, and estimate that family planning clinics had a large impact on fertility in Tanzania. Angeles et al. (2005a) adopt an analogous estimation strategy to show that the 1985 enactment of the National Policy on Population in Peru was successful at reducing fertility. Using a stochastic dynamic model of family planning, education and fertility, Angeles et al. (2005b) re-evaluate the efficacy of the FPPs in Indonesia studied in Pitt et al. (1993) and Gertler and Molyneaux (1994). Using data from a longer time span they find substantial program effects.

A more radical approach to avoid the identification issues described above is the evaluation of FPPs by means of a true experimental design. Field experiments to evaluate the impact of family planning programs are, however, complex and expensive, which may account for their limited number to date (Bauman et al. 1994, Bauman 1997, Phillips 1999). The evaluation of an outreach family planning and health services project in the rural district of Matlab, Bangladesh is perhaps the most widely cited social experiment. This long-term study has been conducted in an area followed since 1966 with a Demographic Surveillance System. A community-based program was introduced in 70 of the 149 villages that are part of the System, and consisted of training local women who visited all reproductive-aged married women approximately every two weeks. Several studies have documented the sizeable impact of this program which, however, was unusually intensive (see, in particular Foster and Roy 1997, Sinha 2005 and Joshi and Schultz 2007). Other examples of family planning experiments are Yang et al. (1965), Freedman and Takeshita (1969), Bang (1971), Chan (1971), and Rosenfield and Limcharoen (1972). Yang et al (1965) reports results of a study completed in 1962-1964 among married couples with the wife age 15-50 in rural Korea. A family planning program involving home visits, group meetings and clinical services led to a large increase in contraceptive practices in treatment areas (from 8% to 37%), and this was greater than the increase observed in control areas (from 12% to 22%). However, the study was carried out in only two localities, with different pre-program characteristics. Freedman and Takeshita (1969) describe the results of an experiment in Taichung (Taiwan) involving about 36,000 married couples with wives age 20-29 randomly allocated to four treatment groups. They find that home visits and small group meetings increased acceptance of intra-uterine devices (IUDs), while mailings and husband involvement did not. Bang (1971) finds that randomly assigned requests of early check-up visits in Koyang (South Korea) brought evidence *against* the initial hypothesis that such program would lead to increased IUD retention. Chan (1971) finds that a home-visit program in Hong Kong aiming at reducing the number of IUD removals among IUD acceptors had only a negligible impact. Rosenfield and Limcharoen (1972) describe an experimental program that allowed auxiliary midwives (rather than physicians) to prescribe oral contraceptives in rural Thailand, leading to substantially larger acceptance rates in treatment versus control provinces.

Experimental evaluations of FPPs in Africa are rare. Most of them have been carried out in West Africa and almost all of them rely on a very small number of primary stage units. As a result, even though they are very valuable as case studies, it is too early to draw more general inferences on the basis of these results.[2] Omu et al. (1989) report results from an experiment in Nigeria. More than one thousand high-parity women admitted for prenatal care at the University of Benin Teaching Hospital during a 19-month period were randomly assigned to receive either standard family planning information or individualized counseling sessions on family planning methods and the health risks associated with high parity. The fraction of women who were using contraception six weeks after partum was 71 percent among women in the treatment group and only 51 percent among the controls. In the Navrongo study, carried out in a poor and relatively remote area of Northern Ghana, four groups of neighboring communities were randomly assigned to different experimental arms (Binka et al. 1995, Debpur et al 2002, Phillips et al 2006). In the first intervention arm, trained nurses were assigned to live

[2] See Phillips et al. (1999) for a broader overview of community-based distribution of family planning in Africa.

and work in communities. They provided health services and made door–to-door visits to provide family planning services, including contraception methods. In a second arm, analogous tasks were assigned to local volunteers, and community mobilization was emphasized to reduce the social cost of adopting contraception. A third group of communities saw the introduction of both interventions, while a fourth received neither. In areas with both interventions, mean fertility was reduced by one birth in three years, a 15 percent reduction. In The Gambia a community-based information campaign in one primary health care circuit increased contraceptive use (relative to a control area), but improved access to fertility control methods in a third circuit did not lead to any additional increase in contraceptive use (Luck et al. 2000). Katz et al (1998) discuss the results of interventions in five sub-districts of rural Mali. A nongovernmental organization's primary health care system saw the introduction of a community-based distribution program in two sub-districts. Information campaigns were introduced in two other sub-districts and a fifth sub district served as a control group. A survey completed one and half year after the introduction of these interventions showed increased knowledge and use of contraceptives in the sub-districts where the community-based distribution was initiated, but no additional increase was observed in districts with information campaigns.

Knowledge and availability of contraceptive methods are of course only two of the many factors entering fertility decisions. At least as important are socio-economic factors such as gender-specific human capital, which can modify the cost-benefit calculus of contraceptive use and childbearing (see Schultz 1997 and Schultz 2005 for overviews). Some have argued that credit programs which encourage borrowing by women can increase the opportunity cost of women's time, potentially increase their control over household resources, and thus empower them enough to express their fertility preferences (Schuler and Hashemi 1994, Hashemi et al. 1996, Schuler et al. 1997, but see also Mayoux 1999 for an opposite argument). Some authors have argued that fertility in poor areas may decrease among women with access to micro-credit, that is, to small collateral-free and group-based loans that allow relaxing budget constraints and may facilitate the creation of small economic enterprises. The theoretical predictions are, however, unclear. On the one hand, if access to credit enhances women's economic opportunities, it may reduce desired fertility by increasing their opportunity cost of time. Where women's desired fertility is lower than men's, access to credit may also improve the woman's bargaining power and increase contraceptive behavior. Empowering effects may also derive from the interactions with other women and social workers that usually accompany micro-credit operations. On the other hand, if children are "normal goods", fertility may actually increase if access to credit results in an increase in income. Overall, the direction of any causal impact of micro-credit on fertility is unclear, as it depends on income and substitution effects of opposite signs.

A number of studies have discussed these considerations and attempted to empirically test the existence and magnitude of a causal link between access to credit and fertility choices. The identification of the causal link has to overcome the usual selection concerns: micro-credit programs are usually non-randomly placed, and both eligibility and actual participation in the programs are likely correlated with characteristics that also directly affect fertility choices. Using different estimation strategies, several authors have found that in Bangladesh, in the 1990s, participation in credit programs increased contraceptive use and in some cases reduced

fertility (Schuler and Hashemi 1994, Amin et al. 1995, Hashemi et al. 1996, Schuler et al. 1997, Steele et al. 2001). Pitt et al. (1999) find instead no evidence of an impact when program participation is measured as a function of the amount borrowed. Buttenheim (2006) uses panel data from Indonesia and uses community and household fixed effects to control for endogenous program placement and participation. She finds that the presence of credit programs increases contraceptive use while borrowing does not. To the best of our knowledge, the link between micro-credit and fertility choices has not been studied rigorously outside of Bangladesh (with the exception of Buttenheim 2006) and it has never been analyzed using a randomized field experiment.

One further aspect which has not been addressed in the literature is whether micro-finance programs can represent useful entry points for family planning services. The regular contact that credit officers have with their clients offers a convenient avenue for providing information on family planning methods, and the group monitoring element of micro-financing programs has the potential for building a support mechanism for adoption of a new practice. One of the purposes of the evaluation described in this paper was precisely this, to analyze whether such an accessory service offered by loan officers could lead to an increase in contraceptive adoption over and above that achievable by a community-based FPP on its own.

3. Program Interventions and Study Design

In Ethiopia, the Packard Foundation provides grants and technical assistance to micro-credit programs and family planning programs in the Amhara and Oromia regions (see Figure 1 for a map). In Amhara, the Program supports the credit activities of the Amhara Credit and Savings Institute (ACSI) and the community-based family planning programs of the Amhara Development Association (ADA). In Oromia, support is provided to the credit activities of the Oromia Credit and Savings Share Company (OCSSCO) and the family planning programs of the Oromia Development Association (ODA).

The credit programs target poor households, and "emphasize" women borrowers, though no specific activities or criteria are used to seek out this target group. Each organization has a specific set of criteria that it uses to select borrowers, of which creditworthiness, viable business plan and poverty are the more salient ones. There is no collateral requirement, instead borrowers form small groups, and take on collective responsibility for repayment of loans. Loans are made for a year at interest rates that reflect market conditions. Credit officers help fill out loan applications, and also monitor the groups through monthly and bi-weekly meetings with all clients. Borrowers are expected to make regular deposits and repayments, and in recent years the repayment rate has been reported to exceed 95 percent.

The family planning programs have a community-based orientation. Residents from the communities are trained, provided a uniform, and receive a fee for their services. They make house-to-house visits to provide information and pills and condoms. They also provide referrals for clinic-based services like injectibles, the main method in use in these regions, but they do not provide injectibles. The organizations also organize other events to provide information on family planning, reproductive health, sexually transmitted diseases, including HIV/AIDS.

The family planning programs have been in operation for several years and have steadily extended their coverage. In this process they have continuously sought to improve the quality

of services they provide and to identify new service delivery options. Given the Packard Foundation's continued support for micro-credit in the same regions, one option that evoked interest was the linking of the family planning programs with the credit programs. The underlying rationale was that the monthly meetings of the borrowers and credit officers present an (additional) opportunity for providing information on family planning and motivating clients to adopt contraception. One of the objectives of the present study is to evaluate this form of linkage.

The evaluation was conducted by Family Health International with the purpose of determining whether linking the credit programs of ACSI and OCSSCO with the family planning activities of ADA and ODA could lead to a measurable increase in contraceptive use over and above that achievable by each type of program on its own. An experimental design was used to randomly allocate administrative areas to four groups with different combinations of credit and family planning services. One intervention group received both credit and family planning services with the credit officers also providing information on family planning. A second intervention group received only family planning services and a third received only credit services. The fourth arm was a control group that received neither type of program services. Change in contraceptive use (and related variables) was then measured with pre and post-intervention household surveys.

The randomly allocated administrative areas are *kebeles*, or peasant associations (PAs). Administratively, Ethiopia is divided (from large to small) into regions, zones, *woredas*, *kebeles*, and villages and so a *kebele* is a cluster of villages. In 2002, the four organizations identified 133 PAs as areas where they intended to start activities in the coming years. Fifty-five of these are in the Amhara region, and 78 are in the Oromia region. In Amhara, the PAs are from Bugna, Gidan, Meket Delanta, Metema, Chilga, Alefa Takusa, and Lay Armachiho *woredas* which are in the North Wollo and North Gonder zones. In Oromia, the PAs are from Mendi, Harru, Nejo, Ayra Guliso, Sayo, Anfilo, Metu and Chora *woredas*, which fall in the West Wollega and Illubabor zones. Population data from the most recent census were used to randomly allocate PAs to the four groups, and the allocation was stratified by region.

Unfortunately, the randomized design was not always followed by the implementing agencies, although deviations remained relatively limited, especially in Oromia. The study protocol was followed in 37 of 55 PAs in Amhara and 66 of 78 PAs in Oromia (see Table 1). Eight PAs already had functioning programs at the time of randomization, of which two were interrupted at the start of the study, and two other PAs were merged by local authorities during the study period. We analyze non-compliance in detail in the next section, which also describes the data and the main estimation strategy.

4. Data and methods

A baseline survey was conducted between the months of January and April in 2003 and preceded the start of sub-grantee programs in the study areas. The survey covered 6440 households, and was spread over 356 villages in the 133 PAs where the family planning and credit organizations intended to expand in the following years. The population in each PA includes a relatively small number of households, ranging from 109 to 1377. Sampling was designed to select approximately 3200 households with at least one woman of age 15 to 49 years

in each of the two regions. Within each region, PAs were randomly allocated to the four study cells so as to yield 800 to 810 households in each cell in each region. In the selected PAs, interview teams obtained a complete listing of all villages along with estimates of the number of households in each village. If a PA had more than 400 households, then at random three villages were selected for interviewing. If the PA had fewer than 400 households, then two villages were selected at random. Within the selected villages, a complete enumeration of households was undertaken, and a random sample of households was selected for interviewing. The sample is not self-weighted and therefore sampling weights are required to produce unbiased estimates of population statistics.

A follow-up survey was completed during the months of April to July, 2006. The survey was conducted in the same villages as the baseline survey, but a new sample of households was drawn using the same procedures used in the baseline survey. It is therefore important to note that the two surveys constitute a *panel of villages*, but *not* a panel of households. Difficulties in accessing some areas resulted in the survey teams not being able to cover all villages in one PA and one village in another PA. As a result, the follow-up sample has only 6275 households.

During the final household survey, a community questionnaire was used to collect village-level information on demographics, income sources, infrastructure, access to facilities, and availability (and timing) of family planning services; unfortunately, similar information was not collected at the time of the baseline survey. During the study period monthly service statistics data were obtained from the *woreda* offices of the four sub-grantee organizations. These data allow us to determine when program services were first introduced in a PA, what services were provided and how many clients were served. At the end of the study the same *woreda* offices were visited again, and data was collected on provision of program services in all PAs in the *woreda*, This information allows us to gauge program coverage in a *woreda*, in the PAs that make up the study area, and in PAs not in the study area.

Table 2 presents selected summary statistics from the baseline survey for the four assigned study groups along with tests of randomization. For each variable we test the null hypothesis of equal means across the study groups, taking into account the clustered nature of the design. These test results show that in each region the four study groups are well-balanced along almost all examined dimensions. The main outcome of interest is contraceptive use, and in both regions we cannot reject the null hypothesis of equal pre-intervention usage rates across the four arms, at standard significance levels. In Amhara, the null of equality is rejected at the 10 percent level for three variables: intention to use family planning in the future, desired number of children and proportion of households who borrowed from revolving credit associations. In Oromia the null cannot be rejected for any variable.

Table 2 also highlights clear demographic and socio-economic differences between the two regions. Fertility was high in both regions in 2003 but distinctly higher in the study areas in Amhara, where women married earlier, began childbearing sooner, had more births, and wanted to have more children than their counterparts in Oromia. For example, mean desired family size was 4.83 in Amhara and 0.5 less in Oromia. Relative to Oromia, twice as many women were married at age 25 in Amhara (74 percent versus 37 percent) and a higher fraction had already had children (53 versus 42 percent).

There was minimal contraceptive use in both regions in 2002 with the current contraceptive prevalence rate (CPR) at 3.5 percent in the Amhara study areas and 7.4 percent

in Oromia. Intentions to use were also higher in Oromia where 71 percent of non-users said they intended to use contraceptives in the future; the corresponding figure for Amhara is 46 percent. While contraceptive use and intentions to use were lower in Amhara than in Oromia, awareness of contraceptives was higher: 85 percent of women in the Amhara study areas were able to list at least one form of contraceptive and 58 percent were aware of the existence of pills or injectibles. In Oromia 78 percent identified at least one contraceptive and only 45 percent were aware of pills or injectibles. These latter results, however, are not necessarily at odds with the lower contraceptive use observed in Amhara because factors other than awareness affect use, and these (like desired family size) point towards lower motivation to use contraception.

The study areas in the two regions are also substantially different in their economic structure, schooling levels and religious affiliation of inhabitants. In both regions, almost all households engage in crop cultivation, but households in Oromia are much more likely (48 percent *vs.* 3 percent in Amhara) to cultivate coffee (a cash crop), even though a smaller fraction sell crops (37 percent vs. 54 percent in Amhara). Three quarters of households in Amhara are engaged in livestock maintenance and the value of their livestock assets is, on average, around twice that of households in Oromia, where livestock care only engages about half of the sample. In both regions 18 percent of households reported taking or repaying a loan in the 12 months before the interview, and borrowing by women, as well as borrowing from revolving credit associations, was infrequent. Education levels are much higher in Oromia. For example, in Amhara only 11 percent of women 15 to 49 years of age had ever attended school, while in Oromia the corresponding figure was 47 percent. Almost all households in Amhara describe themselves as Christian Orthodox, while only one-third of respondents do so in Oromia. In the latter region, religious affiliation is more heterogeneous and also includes 51 percent of Christian Protestants and 13 percent of Muslims. More than half of the household heads ever attended school in Oromia while only 10.5 percent did in Amhara.

These differences between the two regions present an interesting possibility for comparing the impact of essentially similar credit and family planning programs in different settings but we do not pursue this in this paper. Instead we analyze the two regions separately, largely because different organizations implemented the interventions in each region, and for our results to have programmatic value it is important that they be region, and, therefore, organization-specific.

4.1 Estimation Strategy

To estimate the impact of the programs on different outcomes we use a difference-in-differences (DD) approach. Let y_{pit} denote an outcome for individual (or household) i from peasant association p at time t and let D_t denote a binary variable equal to one in the post-intervention period (that is, when $t = 1$). Let also Cr_p, FP_p and $CrFP_p$ denote binary variables equal to one for PAs where the intervention introduces respectively micro-credit, family planning or both services (with loan officers also providing contraceptive information). Then, if treatment is assigned randomly, the causal impact of each intervention is measured by the coefficients α_1, α_2 and α_3 in the following regression:

$$y_{pit} = \beta_0 + \beta_1 CrFP_p + \beta_2 Cr_p + \beta_3 FP_p + \alpha_0 D_t + \alpha_1(D_t \times CrFP_p) + \alpha_2(D_t \times Cr_p) + \alpha_3(D_t \times FP_p) + \varepsilon_{pit} \quad (1)$$

This equation cannot be estimated in DD form, because we do not have a panel of households, so that the household-specific differences $\Delta y_{pi} = y_{pi1} - y_{pi0}$ cannot be calculated. We therefore transform equation (1) in DD form by calculating survey-specific PA-level means of both sides and take first differences, which leads to

$$\Delta y_p = \alpha_0 + \alpha_1 CrFP_p + \alpha_2 Cr_p + \alpha_3 FP_p + u_p, \tag{2}$$

where Δy_p is the change over time in the PA-specific mean of y_{pit}, and where the error u_p is the corresponding mean change of ε_{pit}.[3] Because the left-hand side variable in equation (2) is a mean, we estimate the model using weighted regression, with weights proportional to the number of observations in the PA.[4]

With random assignment of treatment, the parameters of interest can be estimated consistently using Ordinary Least Squares (OLS). However, the analysis of the results is complicated by the imperfect compliance of the sub-grantees with the randomization protocol. We first analyze whether non-compliance was associated with specific characteristics of the communities involved. In particular, we are interested in testing whether family planning was more likely to be implemented in areas which differed systematically from others in observed measures of demand for contraception or in supply availability. For this purpose, we first use data from the 2003 baseline household survey to calculate PA-specific statistics (recall that the randomization was done at the PA level).[5] We calculate fertility-related statistics including only women of childbearing age (15-49). As indicators of current fertility we included the fraction of women who are pregnant and the median number of live births (we prefer medians to means because the former are not as responsive as the latter to outliers). Demand for contraception is measured via desired family size, the fraction of latest births not desired by either the woman or her husband, and the proportion of women who ever used pills or injectibles (these being the only two methods used by a non-negligible number of women at baseline). As measures of availability of family planning services we use the fraction of women who obtained contraceptives from a location less than one hour away, the median distance from the nearest health center and the fraction who knew where contraceptives could be obtained, or who had been visited by a FP worker in the previous 12 months. We acknowledge that our use of the terms "demand" and "supply" to label these variables is approximate, because several of these indicators are likely the result of the interaction between demand and supply. For instance, contraceptive use is only possible if supply is able to meet an existing demand for controlling fertility. We also include indicators of economic activity and pre-intervention exposure to micro-credit (in the form of revolving credit associations) and other forms of formal loans.

Since the programs are operated by separate organizations in Amhara and Oromia, we analyze program non-compliance separately for the two regions. In each region, we regress a

[3] Alternatively, equation (1) could be estimated directly using PA-fixed effects. An earlier draft of this paper used this latter approach, which led to almost identical results.

[4] More specifically, we calculate the weights as the mean number of observations from a given PA in the baseline and in the follow up. These numbers are not always identical, although they are always very similar (the correlation is approximately 0.97 in both regions). Note also that the dependent variable in model (2) is a *weighted* PA-specific mean, where all observations are weighted using the village-specific sampling weights.

[5] A table including means, standard deviations, minimum and maximum for all the PA-specific characteristics is available upon request.

dummy variable equal to one if the PA actually received FPP, on dummy variables that equal one if FPP or micro-credit (MC) was assigned to the PA, and on all the PA-specific characteristics listed above. Since observations are at the PA level we do not cluster standard errors which are, in any case, heteroskedasticity robust. We also estimate the corresponding equation to study compliance with the introduction of micro-credit. When a PA received both interventions, the corresponding variables in the two regressions are both set equal to one.

The results are displayed in Table 3. We comment first on the results for Amhara, where compliance, as we documented above, was more problematic. The only predictor of actual FP implementation that is individually statistically significant at standard levels is the randomly assigned FP treatment (p-value = 0.000). The joint null hypothesis that the coefficients for all the 17 PA-specific characteristics equal zero is rejected at standard levels (p-value = 0.0039) However, the individual coefficients do not follow any consistent pattern that would suggest that actual treatment status was associated with observables related to existing demand (p-value = 0.83) or supply of family planning (p-value = 0.46), or to the economic environment. Naturally we cannot exclude the possibility that non-compliance was associated with unobserved factors possibly correlated with the potential response of fertility-related outcomes to FPPs. When we examine correlates of correct assignment of FPPs in Oromia the results are consistent with the absence of endogenous placement (see column 3 in Table 3). First, the coefficient for randomly assigned FP treatment is very close to one (0.960). Second, none of the other regressors is significant at standard levels and in this case the joint null that all PA-specific characteristics do not enter the regression cannot be rejected (p-value = 0.9998).

When we look at assignment of micro-credit, we again find no clear indication of endogenous program placement in Oromia (column 4), but a different, though inconclusive, result holds in Amhara. In Oromia the joint null hypothesis that all PA characteristics together equal zero cannot be rejected (p-value = 0.710), and none of the coefficients is statistically significant. In Amhara, on the other hand, the null hypothesis that all PA-specific characteristics are (together) equal to zero is rejected (p-value = 0.00). Examination of statistically significant coefficients shows that MC was 26 percent more likely to be introduced in PAs where a FPP had been randomly assigned as well, everything else being the same. We also find that the introduction of MC was positively associated with the presence of households borrowing from banks or cooperatives and negatively associated with realized fertility, past or present use of injectibles, past visits by FP workers and distance from the nearest health center. However, these significant coefficients do not fit into an easily interpretable pattern that would provide a reasonable explanation for how the associated factors might have informed decisions on MC program placement.

Overall, the results in Table 3 show that the deviations from protocol of program implementation do not appear to be associated in any transparent way to PA-level indicators likely correlated with unobserved levels or trends in demand or supply of contraception, especially in Oromia. Note also that OLS estimation of the DD model in equation (2) will adequately account for endogenous program placement if this depended only on time-invariant characteristics at the PA level, which are differenced out in the DD framework. At the same time, correlation with unobserved and time-variant characteristics cannot be ruled out, in which case a simple difference-in-differences approach would not uncover the causal relationship between interventions and outcomes of interest.

In order to measure the causal impact of treatment, we adopt an instrumental variable approach, using treatment assignment, which was random and therefore exogenous, as an instrument for actual treatment, which is potentially endogenous due to imperfect compliance with the study design. In other words, we estimate equation (2) above with two-stage least squares (2SLS) using three binary indicators for the randomly assigned treatment groups as instruments.[6] As a consequence, the number of instruments is the same as the number of endogenous variables, so that equation (2) is exactly identified. One limitation of this is that we cannot perform tests of overidentification, but this is not a serious concern because the instruments, by construction, are random, and thus the tests are not necessary..

We estimate all regressions using linear models, and we adjust standard errors for heteroskedasticity. Note that no adjustment for clustering is necessary, because we estimate the regressions at the cluster (that is, at the PA) level. In an attempt to increase precision, we also estimate models where we include a set of pre-intervention controls. Given the relatively small number of observations (54 in Amhara and 78 in Oromia), we chose a small set of controls, thus avoiding excessive reduction in degrees of freedom. Specifically, we estimate models where we add, as controls, the same 17 PA-level characteristics we have included in the analysis of compliance with the experimental protocol (see Table 3). The DD estimates, by construction, eliminate all PA-specific pre-intervention and/or time-invariant characteristics. This means that such estimates already wipe out residual variance associated with time-invariant PA-level characteristics. Still, the inclusion of covariates may be interpreted as allowing for the model to include time-trends associated with pre-intervention PA characteristics, which may contribute to increase the precision of the estimates.[7]

5. Basic Results

The study areas witnessed substantial demographic and economic change during the three years of this study, though the patterns of change were different in the two regions. In Table 4, we briefly examine some of these aggregate changes before turning to an examination of the impact of the interventions. As we will show, while the changes are remarkable, particularly so given the relatively short period of time between the two surveys, the evidence suggests that they were not necessarily associated with the interventions.

In Amhara, contraceptive use increased by 9 percentage points, and the percentage of non-users who said they intend to use contraception in the future increased from 46 to 65 percent. Awareness of contraceptives, which was already high at 85 percent in 2003 increased to 97 percent and the percentage of women who had heard of pills and injectibles, the two most commonly used methods, increased from 58 percent to 80 percent. The increase in

[6] Alternatively, equation (1) could be estimated with 2SLS and PA fixed effects. This estimation strategy, which we adopted in an earlier version of the paper, leads to almost identical results.

[7] We note, however, that in small samples the inclusion of pre-intervention characteristics does not necessarily increase the precision of the estimates, even in situations where the randomization was done carefully. On the one hand, the R^2 of the regression will, by construction, increase. However, in small samples the orthogonality between assigned treatment and the residuals usually does not hold exactly, with the consequence that the inclusion of other covariates is not guaranteed to decrease the standard errors. We also note that although the inclusion of additional covariates is often advocated by randomized controlled trials practitioners (see, e.g. Duflo, Glennerster and Kremer, 2008), the practice has been criticized by others as an ex-post adjustment not justified by randomization (see Freedman, 2008, Deaton, 2009).

contraceptive use does not appear to have had much effect on fertility: the total fertility rate (TFR) actually went up by 0.5 births, from 5.5 to 6 births per woman, a change significant at the 5 percent level. The number of births women had in the three years before the interview also increased from 0.51 to 0.55 (significant at the 1 percent level). Desired family size was essentially unchanged and women, on average, continued to want almost 5 children.

In the Oromia study areas, demographic change was more marked. Contraceptive use went up by 14 percentage points amongst all women (from 7 to 21 percent), and even more amongst currently married women, for whom the increase was threefold; in 2006 almost a third were using contraceptives. There was little change in intention to use contraception in the future among non-users, but this was already high (at 71 percent) in 2003. Awareness increased and, as in Amhara, was almost universal at 97 percent; awareness of pills and injectibles increased from 45 to 78 percent. There was a small drop in fertility in most age groups and the result is a drop in the Total Fertility Rate from 5.1 to 4.8. The change is, however, not statistically significant at standard levels. Desired family size also dropped by 0.5 births and, on average, women in the study areas in Oromia wanted only 4 children.

In both regions there is underlying momentum for further change because younger cohorts, who have lower desired family size and high levels of awareness of contraceptive methods, are delaying marriage and the start of childbearing. For example, in the study areas the fraction of women married by age 25 decreased from 74 to 63 percent in Amhara, and from 37 to 30 percent in Oromia. The share of women who already had children at 25 also declined, from 53 to 48 percent in Amhara, and from 42 to 28 percent in Oromia. Furthermore, there has been a large increase in schooling in recent years and as a result a substantially larger percentage of younger cohorts have attended school. As younger, better educated women, with high levels of awareness of contraception and lower desired family size, move into childbearing years, contraceptive use is likely to increase further and also result in lower fertility.

Table 5 shows that remarkable changes also took place in relation to economic indicators such as credit uptake, market participation, and livestock holdings. For example, the percentage of households that took a loan in the twelve months before the interview increased from 18 to 44 percent in Amhara, and from 18 to 37 percent in Oromia. Even though much of the borrowing in 2006 was still being undertaken by males, borrowing by women increased substantially, from 3 to 10 percent in Amhara and from 2 to 14 percent in Oromia. There was a significant increase in the value of livestock holdings which, in real terms (2003 prices) doubled in Amhara and almost trebled in Oromia. The large increases in the number of animals owned by households show that the increase in livestock value was not merely due to increases in the relative price of animals. In Oromia, household income sources became more diversified with larger percentages of households deriving income from services, trade, and manufacturing and production. In both regions, but particularly so in Oromia, there was an increase in cash crop cultivation and marketing of crops. What is most striking, though, is the large increase in school attendance. In the primary school age group (6 to 10 years), in Amhara attendance more than doubled, going up from 17.3 to 41.7 percent. In Oromia attendance in these ages increased from 36.2 to 45 percent. Among 11-14 year olds, school attendance increased by 27.4 percentage points in Amhara and by 12.5 percentage points in Oromia. Similarly, in the 15 to 18 age

group, there was a 23 percentage points increase in Amhara and a 14.8 percentage points increase in Oromia.[8]

Overall, Tables 4 and 5 reveal a scenario of remarkable change in the study areas. This is consistent with the important changes documented between 2000 and 2005 in Ethiopia in two nationally representative Demographic and Health Surveys (Macro International Inc. 2007). For example, data from these surveys show a 10 percent decrease in the fraction of women 15 to 49 with no schooling in Amhara, and 16 percent decline in Oromia (see Table 2.2 in the report). The increase in contraception among currently married women of childbearing age in the DHS data is also remarkable, from 6.6 to 15.7 percent in Amhara, and from 4.3 to 12.9 percent in Oromia (see Table 6.2 of the report).

5.1 Basic Estimates of Program Impact

Turning to the focus of this paper, Table 6 presents estimates of the impact of interventions on contraceptive use. All estimates should be interpreted as measuring an intent to treat, interpreted as the average impact of *exposure* to the programs in the community, rather than the average impact of *using* any of the studied interventions.[9] Three sets of estimates are presented for each region. Columns 1 and 4 display the results from an Ordinary Least Squares estimation of equation (2) where exposure to interventions is defined by dummy variables for actual exposure to intervention. These estimates do not identify the causal program impacts if program placement is systematically correlated with unobserved differences in trends across the assigned treatment groups. For this reason, we also estimate a second model using 2SLS (columns 2 and 5) with the assigned exposure dummies as instruments for actual exposure. As shown in Table 3, such instruments are strongly correlated with actual exposure in both regions. The instruments are also arguably exogenous, because their being randomly determined implies that their only correlation with the dependent variable should be through actual treatment. Finally, columns 3 and 6 show the results of 2SLS estimation with the inclusion of pre-intervention PA-level characteristics as controls.

Even though our focus is on the two-stage least squares specifications, the OLS results are, on the whole, very close to those obtained using instrumental variable estimation, suggesting that deviations from the study protocol in program placement were not strongly correlated with unobserved location-specific differences in outcome trends. We also test formally the null hypothesis that actual program placement is exogenous, using a test that is robust to the presence of heteroskedasticity of unknown form. Under the null, both the 2SLS in Table 6 and an alternative 2SLS estimator where *both* actual *and* randomly assigned treatment are used as instruments is consistent.[10]

The results in Table 3 suggest that the dummies for assigned treatment (the instruments) are strongly associated with actual treatment (the endogenous variables). However, we also formally test the null hypothesis that the estimated equations are identified, by using a

[8] The changes in socio-economic status described in Table 5 were overall relatively similar across treatment groups, with the exception of borrowing which, not surprisingly, shows significantly larger increases in areas where micro-credit was introduced. For almost all other variables, we cannot reject the null hypothesis that the changes were the same across the four (actual) treatment groups. The detailed results are available upon request from the authors.

[9] See, for instance, Heckman et al. 1999, p. 1903.

[10] For a description of the test, which is identical to a Hausman test under conditional homoskedasticity, see Hayashi (2000, p. 200-201) or Baum et al. (2007, p. 16).

Kleibergen-Paap *rk LM* statistic, which tests the null hypothesis of *underidentification* (Kleibergen and Paap 2006).[11] We do not provide results for formal tests of instrument strength. Because in our model we have three instruments and three endogenous variables, we cannot use the "usual" rule-of-thumb F-test, where the value of the first-stage F-test for the exclusion restrictions is compared with a threshold equal to 10, below which weak instruments should be suspected. Such test is only appropriate when there is only one endogenous variable (see Section 4.2 in Stock, Wright and Yogo 2002). [12]

The sample is restricted to currently married women because in rural Ethiopia contraceptive use essentially occurs only within marital unions; in the sample, 95 percent of contraceptive users are those who are currently married. Results for all eligible women (15 to 49 years of age) are not substantively different and are available upon request. The estimated intercepts in the models without added controls (columns 1, 2, 4 and 5) are large and significant at the 1 percent level, indicating that contraceptive use increased in the control group between 2003 (baseline survey) and 2006 (follow-up survey) by about 12 percentage points in Amhara and by twice as much in Oromia.[13] This is consistent with the aggregate results in Table 5. However, in both regions and in all regressions, almost all the coefficients that measure intervention impacts are small, not significant at standard levels, and in most cases *negative*, suggesting that, if anything, the increase in contraceptive use in the intervention groups was slightly *smaller* than in control groups. The only two exceptions arise for the impact of micro-credit. The null of no impact is rejected at the 5 percent level in Amhara when we estimate the model with OLS and at the 10 percent level in the 2SLS estimates with added controls for Oromia. In both cases the point estimate is *negative* (–0.066 and –0.076 respectively). The null of no differential change in contraceptive use between the linked group and the groups with only family planning services is never rejected at standard confidence levels (row (3)). The p-values in row (4) indicate that the null of exogeneity of program placement is never rejected at the 1 or 5 percent level, although it is rejected at the 10 percent level in Oromia (column 6). Note finally that the test of underidentification is always rejected at standard levels (row (5)), which is not surprising given that assigned treatment is a strong predictor of actual treatment in both regions.

Table 7 reports the 2SLS results for other correlates of contraceptive use, namely, the intention to use family planning in the future (among current non-users), the number of contraceptive methods the woman has heard of, the number of births in the previous three years and desired family size.[14] The number of methods known to the respondent (which is at most 11), includes methods mentioned by the respondent either spontaneously or after prompting. As

[11] A version of the test robust to the presence of heteroskedasticity or clustering, can be performed by using the Stata command ivreg2. This test can be considered as a generalization of the Anderson canonical correlation rank statistic to the non-i.i.d. case. The null hypothesis is that the smallest canonical correlation is zero, in which case the equation is not identified. A rejection of the null hypothesis indicates instead that the excluded instruments are relevant. The results of the first-stage regressions are available upon request from the authors.

[12] In the presence of more than one endogenous variable, multivariate versions of the test have been developed which evaluate the first stage for all endogenous variables jointly. Critical values for such tests have been developed only for specific combinations of the number of instruments and endogenous variables (see Tables 1 and 2 in Stock and Yogo 2002). Unfortunately, such critical values do not exist for a case such as ours, where there are three endogenous variables and three excluded.

[13] In the model with pre-intervention controls estimated in columns 3 and 6 the intercept does not have a meaningful interpretation.

[14] The OLS estimates, which are not reported for brevity, are available upon request from the authors.

in Table 6, we estimate model (2) both with and without additional controls and we perform the same set of tests. In all cases, the null of underidentification is rejected at the 1 percent level. In Amhara, the null hypothesis of no program impact cannot be rejected at standard significance levels in all but one case. The only exception is the estimate for the introduction of micro-credit only, which is estimated to have *increased* desired family size by almost one child and is significant at the 10 percent level (column 7), although even this coefficient is no longer significant once we include pre-intervention controls (column 8). In Oromia too we find that most estimates are close to zero and not statistically significant at standard significance levels, although the picture is more complex. When we include pre-intervention covariates, the introduction of micro-credit predicts a 9 percent *decline* in the intention to use contraception (significant at the 10 percent) and an *increase* of 0.38 in desired family size (significant at the 5 percent level). The predicted impact on awareness and births in the previous three years is instead negative but small and not significant. Overall, these results are consistent with micro-credit leading to an increase in the demand for children.

All estimates of the impact on number of births in the previous 3 years are negative, and even though in some cases they are statistically significant, the magnitude is always small, ranging from –0.106 to –0.166. When we look at estimated program impacts on desired family size we find instead positive and, in most cases, statistically significant results. In the model with added control (column 8), the already mentioned 0.4 predicted increase associated with micro-credit is complemented by a similar increase in FP areas (0.42, significant at the 5 percent level) and a lower increase (0.27, significant at the 10 percent level) in PAs where both services were introduced.

Overall, then, the impact of the credit and family planning programs on contraception and other fertility-related outcomes appears to have been limited and, in some cases, even of sign *opposite* to what was expected. In the next section, we turn to examine possible explanations and their implications, as well as alternative estimates of program impacts.

6. Discussion

What explains the limited impact revealed by our estimates? A first important drawback of the intent-to-treat analysis described in the previous section is that the parameter being estimated is a measure of the impact of the programs on mean outcomes *for the whole targeted population*. This implies that the intent-to-treat estimates will be a weighted average of likely heterogeneous responses, with weights proportional to the prevalence of each response type in the population. In particular, the introduction of FPPs should be expected to be more effective in communities with a higher degree of latent demand for contraception. In section 6.1 we analyze the hypothesis that program impact differed among women in different age groups or among communities where indices of demand for contraception are dissimilar. In addition, we analyze further how the results may be explained by the design of the study (Section 6.2) and by the coverage and content of the interventions (Section 6.3). As a final step, in Section 6.4 we also look at patterns of migration in the study areas to see if the observed results could be due to selective migration related to program placement.

6.1 Heterogeneity in Program Response

Do these results hide significant differences in the impact of the program on contraception for women of different age cohorts? On the one hand, larger impacts can be expected among younger women, who are on average better educated and who may be more willing to accept methods of fertility control largely ignored in the past by older women. On the other hand, older women may be more likely to adopt because they are, on average, closer to achieving their desired family size. To explore the hypothesis that different age groups may have been impacted differently by the programs, we re-estimate equation (2) separately for different cohorts, defined on the basis of women's age at baseline, that is, in 2003. We then calculate different 2SLS-DD estimates for women of age below 20, 20 to 25, 25 to 30, 30 to 35, 35 to 40 and older than 40.

The four graphs in Figure 2 show the estimates for Amhara of the coefficients α_0, α_1, α_2, and α_3 (as defined in equation 2) together with 95% confidence bands, plotted against the cohort. Figure 3 shows the corresponding coefficients for Oromia. The coefficients do not show definite patterns in either region. In Amhara, the coefficients α_0 (graph A) show that the largest increase in contraceptive use took place among older women, while for women younger than 25 the change is small and not statistically significant at the 5% level. However, the difference-in-differences estimates show that in each of the three intervention groups the change in contraception remains close to that observed in control areas, and the 95% confidence bands include zero for all cohorts and in each intervention group, with only two cases where the bands barely cross zero. Note also that the point estimates of the impacts are in most cases *negative*. In Oromia, the change in contraception is large (above 20 percentage points) and very similar across different age groups, with the exception of women 35 or above, for whom the change is somewhat smaller but not statistically significant at the 5 percent level. Once again the estimated program impacts (graphs B, C and D) do not display clear patterns and are mostly close to zero and not statistically significant; the only exception is women in the 35 to 40 cohort in areas with FP only, as the estimated DD is large (0.27) and positive.

Overall, the findings described in figures 2 and 3 suggest that contraceptive adoption increased especially among younger women (panels A) but they also suggest that the lack of program impacts was shared by all demographic groups (panels B, C and D). On the other hand, as we have mentioned, age is not a satisfactory indicator of demand for contraceptives, because while younger women may be more open to experiment with contraceptive technology, they are less likely to have achieved their desired family size. At the same time, gauging whether the interventions had some efficacy among women with latent demand for family planning methods is important, because given the relatively duration of the study, one would expect increases in contraceptive use due to easier access to contraception to derive mostly from the satisfaction of existing demand. for this reason, to probe further whether the program impacts were related to pre-intervention demand for contraception, we re-estimate the basic model in (2) interacting treatment status with indicators of demand.

Measures of demand would be ideally constructed at the individual level, but since our data are not an individual-level panel, we have very limited pre-intervention information on demand for contraception for women interviewed in the follow up survey. One solution to this problem is to use baseline information to identify peasant associations (PAs) where pre-program demand for contraception appears to be stronger. For this purpose, we have constructed a

measure of demand for contraception using baseline information about the desire to delay or avoid past pregnancies. The survey questionnaire asks all pregnant women whether they and/or the child's father desired the pregnancy, and it also separately asks whether the mother would have preferred to delay the pregnancy. Analogous questions are asked about the previous pregnancy (if any). Our measure of demand for contraception is then constructed as the fraction of ever-pregnant women of child-bearing age in the PA who (or whose partner) would have preferred to avoid or delay the current pregnancy (if any) or the previous one (if any). Admittedly, this measure of demand is not perfect. First, wanting to delay or "avoid" a pregnancy may signal but does not imply a demand for contraception: instead of using contraceptives such as condoms, pills or injectibles, couples may decide to regulate their fertility with abstinence, or they may differ in their attitudes towards pregnancy risk. Second, the measure is subject to some extent to post-hoc rationalization, that is, respondents may avoid describing a pregnancy or its timing as unwanted because they prefer to see their fertility outcomes as the result of rational choices, or they may feel uneasy about describing the birth or conception of a child as undesired. In addition, the indicator may to some extent reflect not only demand, but also the cost and availability of contraception (Pritchett 1994). With these caveats, we will refer to the index constructed as described above as to "demand" and we estimate the following modification of equation (2), where the variable D_p denotes the index, measured on a 0 to 1 scale:

$$\Delta y_p = \alpha_0 + \alpha_1 CrFP_p + \alpha_2 Cr_p + \alpha_3 FP_p + \alpha_4 D_p + \alpha_5 D_p \times CrFPp + \alpha_6 D_p \times Cr_p + \alpha_7 D_p \times FP_p + u_p. \qquad (3)$$

In Amhara, the values of the PA-specific measure of demand have mean 0.19 and range from zero to 38 percent (Table 8). In Oromia, consistently with the descriptive statistics presented in Table 1, demand for contraception is significantly higher, with the median equal to 0.36 (almost twice as large as in Amhara) and the values ranging from two to 81 percent. We then estimate equation (3) with 2SLS using, as outcomes, both contraceptive use and the other outcomes analyzed in Table 7. We treat the three actual program dummy variables and their interactions with demand as six endogenous variables and we use the randomized assigned program dummies and their interaction with demand as the six instruments. We also include demand itself in the model, but because demand is measured before the intervention, we treat it as exogenous. As usual, the DD model is estimated with observations at the PA level. The results are displayed in Tables 9 (Amhara) and 10 (Oromia). For each regression, we also test the joint null hypothesis that the introduction of FP interacted with demand has no impact on the outcome, either in isolation or when introduced with micro-credit (that is, the null is $\alpha_5=\alpha_7=0$).

In both regions, the null of no interactions between demand and program indicators is never rejected at standard significance levels. The same is true for the joint null that $\alpha_5=\alpha_7=0$. At the same time, the results for Amhara should be interpreted with caution, because the Kleibergen-Paap tests (in rows labeled (3)) indicate that the null of underidentification is not rejected at standard levels, and this suggests the existence of collinearity problems in the first stage regressions. Indeed, this is confirmed by the second-stage estimates, where several of the coefficients and their standard errors are remarkably large in magnitude. Note, however, that

the same problem does not arise in Oromia, where the null of underidentification is always rejected at the 5 percent level.

Overall, we are left to conclude that the FPPs were ineffective in changing contraceptive use even after we taken into account differences in pre-existing demand. So, in the next subsections we move to considering other possible explanations for these results.

6.2 Study Design and Contamination

The design of the study called for randomly allocating administrative areas (the PAs) to the four study group, and for the most part the process followed protocol. Sub-grantee organizations provided the list of PAs to be randomized and this consisted of areas they intended to start programs in. Randomization was undertaken by Family Health International and the randomized list was communicated to the implementing organizations in both regions. It turns out that the list was not entirely error-free because programs were already functional in eight of the 133 PAs. Depending on the allocation, these were either continued or interrupted. We have analyzed the results excluding these areas, but all the conclusions remain essentially unchanged.

As already indicated, implementation of the interventions deviated from the study protocol in a number of PAs (see Table 1). This occurred either because of pressure from local authorities or due to organizational decisions related to availability of services from other organizations, or inaccessibility of PAs. Note, however, that we do not find evidence of important systematic correlations between program implementation and pre-intervention community characteristics (see Table 3). We also note that the DD estimates control, by construction, for any pre-intervention difference in PA characteristics, as long as these enter the outcome equation in an additively separable form. In addition, the 2SLS estimates should take care of the possible endogeneity of program placement, which may still generate inconsistent estimates of program impacts if program placement was correlated with group specific unobserved trends. At the same time, the estimates that include pre-treatment characteristics also implicitly control for differences in trends explained by such added variables. On the other hand, neither DD nor 2SLS estimates can overcome the confounding impact that may derive, first, from spillovers across treatment groups and, second, from the fact that non-program FPPs may have endogenously reacted to the introduction of program FP services in a way that could have attenuated any impact of our study interventions. We address these two issues in turn.

The possibility of spillovers across study areas is especially relevant for the impact of family planning services, for which information was a central component. To the extent that PAs in the two groups that did not receive family planning services from ADA or ODA (the control group and the credit only group) bordered those with family planning services from the sub-grantees, there could have been spillover of information and infusion of the idea of family limitation. This possibility cannot be ruled out because the program office is organized at the level of the *woreda*, the larger administrative area within which the study PAs are based, and each *woreda* was blanketed by family planning services from ADA and ODA. For example, program functioning data show that in the study *woredas*, on average, 70 to 73 percent of all PAs had family planning services from one of these two organizations. Also, we find that all woredas which included at least one control PA also included at least one PA where FP services were introduced. While this certainly raises the possibility that information available in PAs

with family planning programs spread to other PAs, its impact is unlikely to have been large for at least two reasons. Firstly, personal contact with community-based reproductive health agents, and their motivational impact, was available only in the designated PAs. Secondly, injectibles are the preferred method for most women in this region, and the health agents only provided pills and condoms (an issue we return to later).

Another possible source of study contamination is the presence of non-program providers of FP services. The four intervention groups are defined in terms of exposure to the credit and family planning services provided by the sub-grantee organizations. There was no control over the health services provided by government facilities and providers, and only limited ability to influence the actions of other organizations. The expectation was that the initial randomization of PAs would yield a random distribution of government health facilities and services from other organizations, so that the services provided by the sub-grantees could be viewed as being "additional." However, the presence of additional programs may have contributed to attenuate any cross-group differences in the impact of the interventions implemented by the sub-grantees, if such other programs were disproportionately introduced after 2003 in communities not impacted by ADA or ODA.

Data from a community questionnaire administered at the same time as the post-intervention survey suggest that in about one-half of all surveyed villages family planning services were available from non-ADA/ODA sources. In Oromia villages, these other sources are primarily public sector providers like the "Health Post," "Health Center," or "Health Worker," and in Amhara non-governmental organizations play an important role (see Table 11). In Oromia, 46 to 55 percent of villages have non-ODA family planning services, but the differences between the four study arms are not statistically significant; a chi-square test of association is not able to reject the null hypothesis that the presence of non-intervention programs is the same across all treatment arms (p-value 0.535). However, in the Amhara region the presence of non-intervention programs is not similar across the four study arms and the control villages are much more likely to have services from other sources (58 percent vs. 21 to 50 percent in the other three groups); a chi-square test of association strongly rejects the null of equal distributions (p-value = 0.000).

These data suggest that in the case of Oromia it is reasonable to assume that the presence of other non-intervention services did not compromise the study design. In the Amhara study areas, it is possible that the availability of family planning from non-public sources compensated for the lack of ADA services, though data also indicate that this is likely only in the control group villages. No such compensatory placement is evident for the "credit only" group which also did not receive services from ADA. Indeed, the "credit only'' group is least likely to have received services from all other sources (public and non-public) and yet we observe no difference in change in contraceptive use between any of the other groups and this group. Note also that the two intervention groups in Amhara (Both and FP) are *more* likely to have had public providers, who are the primary source for injectibles, the method of choice. Discussions with organization officials also indicate that placement of public providers - who provide a wide range of health services besides family planning - is unlikely to have had anything to do with the operation of the ADA program. In fact, if anything, the unequal presence of these providers should have produced an upward bias to any difference between the family planning intervention groups and the groups not receiving these services. Unfortunately

we do not have baseline data on the presence of these other programs so it is not possible to provide a definitive answer to this issue, but given the role played by public providers and their disproportionately higher presence in the intervention villages, it seems reasonable to conclude that placement of other services probably did not have a major bearing on the study results. We next turn our attention to the coverage and content of the interventions.

6.3 Coverage and content of interventions

If programs did not reach sufficiently large numbers of individuals or if the types of services they provided were not consistent with what women want or what holds them back from adopting family planning, then the interventions might not have any demonstrable impact on contraceptive use. We examine these issues separately for credit and family planning interventions, because the functioning of the former is relevant to the question of linking while the latter might help explain the lack of any impact of family planning programs.

The micro-credit programs operated by ACSI and OCSSCO are necessarily limited in their coverage because, like several other micro-credit programs, they employ selection criteria that restrict lending to certain types of individuals and households. Service statistics show that, on average, the credit-programs serve between 112 and 125 clients per PA per month, and while 60 to 70 percent of these clients are female, these account for no more than 20 to 25 percent of the adult population of a PA. The number of credit clients is higher in Amhara, and even though in both regions the number of clients (per month) increased over the two-year period covered by these data, at the end of this period these still make up no more than 28 percent of households in the credit intervention PAs. In both regions, coverage of households in the linked PAs (at 25 percent) is significantly lower than that in the unlinked, credit only, PAs (34 percent). This means that even if the linking of credit and family planning services were to lead to higher contraceptive use amongst borrowers (from the credit program) this might not get reflected in a group-wide measure of contraceptive prevalence. Of course it is an open question as to whether this type of linking even has any effect on the subset of the population that borrows from the credit program.

We next examine data on borrowing to see if there is any relationship between borrowing, awareness of family planning methods, and contraceptive use. Our intent is not to establish a causal relationship between these variables, but to see if there is any association between participation in the credit intervention, contraceptive awareness, and contraceptive use. Borrowing and contraceptive use are both individual decisions, and as such affected by individual characteristics, only some of which are observed in our data set. Factors such as "entrepreneurship", quality of schooling, risk aversion or attitudes towards modern contraceptive methods are all likely to affect both outcomes but they are all inherently hard to measure. Establishing a causal relationship between borrowing and contraceptive use requires identifying at least one variable that affects borrowing but not contraceptive use. Such a requirement does not seem to hold for any of the variables in our data, so even though exposure to a credit program is randomized in our experiment we cannot identify the causal impact of borrowing on fertility-related choices.

Table 12 presents data on contraceptive awareness and use amongst women from households that did not take any loans in the 12 months preceding the follow up survey, those that took loans from the sub-grantee credit organizations (ACSI and OCSSCO) and those that

took loans from other sources. On the whole those who took a loan are somewhat more likely to be aware of family planning methods, but the differences are minor and not statistically significant (test results not shown but available on request). Here, awareness is measured by the number of family planning methods mentioned by a woman (spontaneously and on being prompted). Contraceptive use displays greater differences between the different types of households and there is some indication that contraceptive use is higher amongst women from households that are engaged in the credit market. However, there is no difference between those who borrowed from ACSI/OCSSCO and those who borrowed from other sources, suggesting that the information provided by credit officers did not necessarily lead to their clients having appreciably higher levels of awareness and contraceptive use. What this means is that not only do the credit programs reach a sub-set of households, but the type of family planning service they provide (information) is largely redundant. It is, therefore, not surprising that linking the credit and family planning programs of ACSI and ADA and OCSSCO and ODA, in this particular way, had no impact on contraceptive use.

Turning to the coverage and content of the family planning intervention there are few concerns with coverage because most programs (87 percent) started within 12 months of the baseline survey, and were in operation for at least 24 of the 36-month study period. We also find that program duration has little bearing on levels of contraceptive use (Family and Health International 2007). In addition, service statistics data from the *woreda* offices show that the programs covered at least 50 percent of eligible households in the initial nine-month period (August 2004 to April 2005) and almost 60 percent in the later twelve-month period (Family and Health International, 2007). Interestingly, in both regions and over both time periods, the rate of household coverage was much greater in the PAs that received both credit and family planning programs. This was not part of the study design, but is an important finding nevertheless because even with a more intensive effort in the linked group contraceptive prevalence increased by the same amount in all groups.

The content of these planning programs is more likely to be the reason for their limited impact on contraceptive use. By all accounts the information provision activities of the community-based agents were remarkable, but these do not seem to have translated into significantly higher levels of awareness of women in intervention PAs. This might well be a reflection of the limitation of the survey instrument and the questions that we are using to measure knowledge and awareness, but it is important to remember that awareness was already high before these programs were introduced (Table 2), so limited awareness was not the main barrier to adoption of family planning.

A bigger shortcoming of the programs might have been that the contraceptives provided by the community-based agents, that is, pills and condoms, were *not* the ones women were increasingly turning to by 2006. Figure 4 shows that in 2003 the method mix was dominated by injectibles and pills, with injectibles making up a larger share in Amhara and pills a larger share in Oromia, although with contraceptive prevalence only 3 percent in Amhara and 7 percent in Oromia (in 2003) these shares do not translate into large numbers of users. Over the next three years the method mix shifted towards injectibles. By 2006, almost 80 percent of women using contraceptives in Amhara were using injectibles, and in Oromia the share of injectibles was almost 62 percent. Since women have to go to a health center or clinic for an injectible, location of these facilities, more than the efforts of community-based agents, might at least partly

account for differences in contraceptive use across communities. Indeed, Figure 5 shows that there is a clear correlation between contraceptive use among currently married women and distance to the nearest health center. Of course, such correlation does not necessarily indicate that a causal relation exists, because women who live at different distances from health centers might also differ along several other characteristics, such as attitudes towards contraceptives or schooling levels.

6.4 Migration

As in virtually any program evaluation, results can be biased by the presence of selective attrition. In our case, it is important to establish if the results we have described are likely to be driven by selective migration. Recall that the two surveys that constitute our dataset are repeated cross sections from the same list of villages. This, unfortunately, does not allow us to evaluate the extent of migration away from the sample villages. On the other hand, the follow up survey includes a random sample from the complete listing of households who resided in the selected villages at the time of the field work. In this post-intervention sample, only 80 of 6,275 respondents (that is, 1.28 percent) report having lived in their village for less than 4 years. Information on the reason for migration is only available for 46 of these households, but in no case is availability of family planning or microcredit indicated as a reason for relocation. Overall, the data suggests that the extent of migration in the study areas between the two surveys was very limited, and unrelated to the interventions. Note also that any bias due to migration would be likely to bias *upwards* the impact of the interventions, because we would expect relocation to be mostly towards areas where the programs have been introduced, and from households who intend to make use of the programs themselves.

7. Conclusions

In conclusion, the results of this study show that, in the study areas, linking credit and family planning services did not increase contraceptive use any more than what was achieved by either program on its own. More importantly, neither type of program, linked or unlinked, led to an increase in contraceptive use that is significantly greater than that observed in the control group. We also do not find systematic differences in program impact among women of different ages, or in areas where pre-intervention measures of demand for contraception were higher.

In interpreting these findings, and in assessing whether they can be extended to other geographical and institutional frameworks, it is important to recognize the specifics of the interventions and the study locations; in other words the usual *caveats* about external validity of experimental results should be kept in mind. In the programs evaluated in this paper, the linking of credit and family planning services took the specific form of credit officers providing information on family planning to their clients, and the FP services offered relied on using community-based reproductive health agents to inform and motivate potential users, provide non-clinical contraceptives (pills and condoms) and referrals for clinical methods. It is also possible that fertility behavior did not respond to the intervention because the study period (three years) was not long enough to allow reproductive behavior and especially demand for contraception to change. On the other hand, as we have documented, even in such a relatively short period of time the study areas did, as a whole, experience a large increase in contraceptive

use, although such change does not appear to be associated with the programs we have evaluated. However, contraceptive use has increasingly taken the form of injectibles, a clinic-based method which surely has been made possible by its availability in local health facilities. This latter observation still strongly points to the potential importance of FP service provision in changing contraceptive behavior.

Our finding that linking credit and family planning services does not have incremental benefits for contraceptive use is quite robust. The lack of differences in change in contraceptive use is not an isolated finding but it is confirmed by lack of statistically significant differences in current fertility, contraceptive awareness, intentions to use contraception, and other relevant demographic variables. We hypothesize that the reason why linking has such a limited impact is because the credit programs reach only one-quarter of all adults, and it only provides them with information, which is important, but probably not the main constraint. If linking were to take a form that altered the incentive structure for contraceptive use, say by offering credit on better terms to women, or to contraceptive users, it might have a greater impact, although our data are silent about this possibility. The data show higher contraceptive use in households that are engaged in the credit market, but given that such correlation cannot be interpreted causally, it is not clear whether this result should be interpreted as suggesting that an expansion of credit access would lead to an increase in contraceptive use. Indeed, the 2SLS results reported in Table 6 show that contraceptive use in PAs where credit was expanded saw relative *declines* in contraceptive use, even though the estimates are small in magnitude and not statistically significant at the 5 percent level.

Our second finding, that the family planning programs of ADA and ODA had no measurable impact on contraceptive use is perhaps more surprising. Besides the relatively short time period between pre and post-intervention surveys, we have hypothesized that the most likely cause for the lack of impact is the fact that community health agents were able to supply condoms and pills but not injectibles, despite the latter being the most commonly used contraceptive method in the area, especially in the post-intervention year. We have also argued that spillovers from intervention to control areas (or from neighboring non-study PAs where ADA or ODA programs were already in place) may have further attenuated any impact. On the other hand, these concerns are mitigated by the fact that information about contraceptive was already widespread in study areas and by the fact that ADA/ODA-trained health agents did not supply injectibles. Similarly, we do not find clear evidence (especially in Oromia) that the lack of impacts on fertility behavior and preferences may have been caused by the entry of alternative family planning services offered by public or private structures other than ADA or ODA.

Overall, given women's preference for injectibles, and the importance of location of the health center for provision of injectibles, one obvious modification of the family planning programs operated by ADA and ODA is to train their community-based reproductive health agents to provide injectibles. As it turns out, independently of this evaluation, the Ethiopian government has recently adopted exactly this type of approach and started placing trained village health workers in each PA.

References

Amin, R., Hill, R. B., and Li, Y. (1995). "Poor Women's Participation in Credit-Based Self Employment: The Impact on Their Empowerment, Fertility, Contraceptive Use, and Fertility Desire in Rural Bangladesh." *Pakistan Development Review*, 34(2), 93-119.

Angeles, G., D. Guilkey, and T. Mroz (1998). "Purposive Program Placement and the Estimation of Family Planning Program Effects in Tanzania." *Journal of the American Statistical Association* 93(443), 884-899.

Angeles, G., D. Guilkey, and T. Mroz (2005a). "The determinants of fertility in rural Peru: Program effects in the early years of the national family planning program." *Journal of Population Economics* 18(2), 367-389.

Angeles, G., D. Guilkey, and T. Mroz (2005b). "The Effects of Education and Family Planning Programs on Fertility in Indonesia." *Economic Development and Cultural Change* 54(1), 165-201.

Bang, Sook (1971). "KOREA: The Relationship Between IUD Retention and Check-up Visits." *Studies in Family Planning* 2(5), 110-112.

Baum, Christopher F. and Mark E. Schaffer and Steven Stillman (2007). "Enhanced routines for instrumental variables/generalized method of moments estimation and testing." *Stata Journal* 7(4), 465-506.

Bauman, Karl E. (1997). "The effectiveness of family planning programs evaluated with true experimental designs". *American Journal of Public Health* 87 (4), 666–669.

Bauman, Karl E., Claire I. Viadro, Amy O. Tsui (1994). "Use of True Experimental Designs for Family Planning Program Evaluation: Merits, Problems and Solutions." *International Family Planning Perspectives* 20(3), 108-113.

Binka, Fred N., Alex Nazzar and James F. Phillips (1995). "The Navrongo Community Health and Family Planning Project". *Studies in Family Planning* 26(3), 121-139.

Bongaarts, John (1994). "The Impact of Population Policies: Comment". *Population and Development Review* 20(3), 616-620.

Buttenheim, Alison (2006). "Microfinance Programs and Contraceptive Use: Evidence from Indonesia." Working Paper, California Center for Population Research, CCPR-020-06.

Chan KC. (1971). "Hong Kong: report of the IUD Reassurance Project." *Studies in Family Planning*. 2(11), 225-233.

Deaton, Angus (2009), "Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development", RPDS Working Paper, Princeton University.

Debpuur, C., J. F. Phillips, E. F. Jackson, A. Nazzar, P. Ngom, and F. N. Binka (2002). The impact of the Navrongo project on contraceptive knowledge and use, reproductive preferences, and fertility. *Studies in Family Planning* 33 (2), 141–164.

Duflo, Esther, Rachel Glennerster and Michael Kremer (2008), "Using randomization in development economics research: a toolkit," Chapter 61 in T. Paul Schultz and John Strauss, eds., *Handbook of Development Economics*, Vol. 4, Elsevier.

Family Health International (2007). *Linking Access to Credit and Family Planning Services in Ethiopia. Final Report*. Prepared for the David and Lucile Packard Foundation Population Program in Ethiopia.

Foster, Andrew and Nikhil Roy (1997). "The Dynamics of Education and Fertility: Evidence from a Family Planning Experiment." Economics Department Working Paper, University of Pennsylvania.

Freedman, David (2008), "On regression adjustments to experimental data," *Advances in Applied Mathematics*, 40, 180–93.

Freedman, R. (1997). "Do family planning programs affect fertility preferences? A literature review". *Studies in Family Planning* 28 (1), 113.

Freedman, R. and JY Takeshita (1969). *Family Planning in Taiwan: An Experiment in Social Change*. Princeton, NJ: Princeton University Press.

Gertler, Paul and John Molyneaux (1994). "How Economic Development and Family Planning Programs Combined to Reduce Indonesian Fertility." *Demography* 31 (1), 33-63.

Goldberg, H. I., M. McNeil, and A. Spitz (1989). "Contraceptive use and fertility decline in Chogoria, Kenya". *Studies in Family Planning* 20 (1), 17–25.

Hashemi, S. M., S. R. Schuler, and A. P. Riley (1996). "Rural credit programs and women's empowerment in Bangladesh". *World Development* 24 (4), 635–653.

Hayashi, Fumyo (2000). *Econometrics*. 1st ed. Princeton, NJ: Princeton University Press.

Heckman, J., R. LaLonde, and J. Smith (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics, Vol. 3A*. Amsterdam, The Netherlands: Elsevier Science.

Katz, K., C. West, F. Doumbia, and F. Kané (1998). "Increasing access to family planning services in rural Mali through community-based distribution". *International Family Planning Perspectives* 24 (3), 104–110.

Kleibergen, F., and R. Paap. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 127(1): 97–126.

Luck, M., E. Jarju, M. D. Nell, and M. O. George (2000). "Mobilizing demand for contraception in rural Gambia". *Studies in Family Planning* 31 (4), 325–335.

Macro International Inc. (2007). *Trends in Demographic and Reproductive Health Indicators in Ethiopia*. Calverton, Maryland, USA: Macro International Inc..

Mayoux, L. (1999). "Questioning virtuous spirals: Microfinance and women's empowerment in Africa". *Journal of International Development* 11, 957–984.

Miller, Grant (2007). "Contraception as Development? New Evidence from Family Planning in Colombia." Working Paper, Stanford Medical School and NBER.

Omu, Alexander E., Sharon S. Weir, Barbara Janowitz, Deborah L. Covington, Peter R. Lamptey and Nadine N. Burton (1989). "The Effect of Counseling on Sterilization Acceptance by High-Parity Women in Nigeria." *International Family Planning Perspectives* 15(2), 66-71.

Phillips, James, Wendy Greene and Elizabeth Jackson (1999). Lessons from community-based distribution of family planning in Africa. Policy Research Division Working Paper 121, The Population Council, New York, NY.

Phillips, J., A. Bawah, and F. Binka (2006). Accelerating reproductive and child health programme impact with community-based services: the Navrongo experiment in Ghana. *Bulletin of the World Health Organization* 84 (12), 949–955.

Pitt, M., Khandker S., S.-M. Mckernan and M. Abdul Latif (1999). "Credit programs for the poor and reproductive behavior in low-income countries: Are the reported causal relationships the result of heterogeneity bias?" *Demography* 36 (1), 1–21.

Pitt, M., M. Rosenzweig and D. Gibbons (1993). "The determinants and consequences of the placement of government programs in Indonesia". *World Bank Economic Review* 7 (3), 319–348.

Pritchett, L. (1994). "Desired Fertility and the Impact of Population Policies." *Population and Development Review* 20 (1), 1-55.

Rosenfield, Allan G. and C. Limcharoen (1972). "Auxiliary midwife prescription of oral contraceptives: an experimental project in Thailand." *American Journal of Obstetrics and Gynecology* 114(7), 942-949.

Schuler, S. R. and S. M. Hashemi (1994). "Credit programs, women's empowerment, and contraceptive use in rural Bangladesh". *Studies in Family Planning* 25 (2), 65–76.

Schuler, S. R., S. M. Hashemi, and A. P. Riley (1997). "The influence of women's changing roles and status in Bangladesh's fertility transition: Evidence from a study of credit programs and contraceptive use". *World Development* 25 (4), 563–575.

Schultz, T. P. (1997). Demand for children in low income countries. In M. R. Rosenzweig and O. Stark (Eds.), *Handbook of Population and Family Economics*. Amsterdam: Elsevier Science.

Schultz, T. P. (2005). "Fertility and income". Yale University, Economic Growth Center Discussion Paper no. 925.

Sinha, Nistha (2005). "Fertility, ChildWork, and Schooling Consequences of Family Planning Programs: Evidence from and Experiment in Rural Bangladesh." *Economic Development and Cultural Change* 54 (1), 97-128.

Snyder L, Diop-Sidibé N, B. L. (2003). "A meta-analysis of the effectiveness of family planning campaigns in developing countries". Paper presented at the 2003 Convention of the International Communication Association, ay 23-27 2003, San Diego, CA.

Steele, F., Amin, S., & Naved, R. T. (2001). "Savings/Credit Group Formation and Change in Contraception." *Demography,* 38(2), 267-282.

Stock, J, J Wright and M Yogo (2002), "A survey of weak instruments and weak identification in generalized method of moments", *Journal of Business and Economic Statistics*, 20(4) 518-529.

Stock, J and M Yogo (2002), "Testing for weak instruments in linear IV regression", NBER Technical Working Paper 284.

Thomas, D. and J. Maluccio (2001). "Fertility, Contraceptive Choice, and Public Policy in Zimbabwe." *The World Bank Economic Review* 10 (1), 189-222.

Yang, Jae Mo, Sook Bang, Myung Ho Kim and Man Gap Lee (1965). "Fertility and Family Planning in Rural Korea." *Population Studies* 18(3), 237-250.

Figure 1

Map of Ethiopia. The study areas are Peasant Associations in the states of Amhara and Oromia

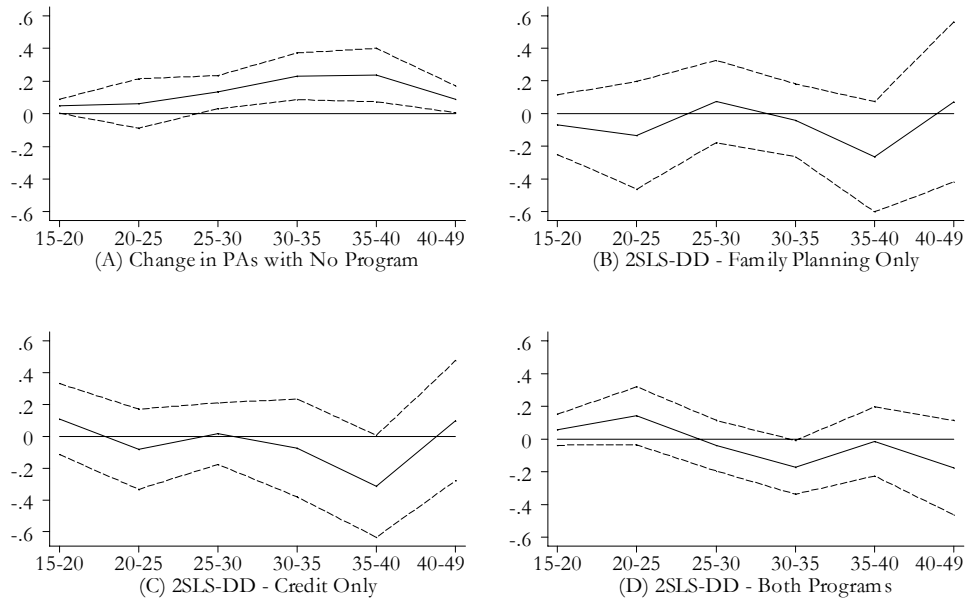

Figure 2

Amhara. Changes in contraception, by intervention and age group. Each continuous line joins comparable coefficients in the model of equation (2) estimated separately at the PA level with 2SLS using samples with women in different cohorts. Cohorts are labeled based on women's ages in 2003. The dashed lines represent the 95% confidence bands, calculated with heteroskedasticity-robust standard errors. The number of PAs (n_{PA}) and the total number of women (n_W) included in the estimation for each cohort-specific regression are as follows: age 15-20 at baseline, n_{PA}= 52, n_W=720; age 20-25 at baseline, n_{PA}= 54, n_W=904; age 25-30 at baseline, n_{PA}= 54, n_W=1228; age 30-35 at baseline, n_{PA}= 54, n_W=888; age 35-40 at baseline, n_{PA}= 52, n_W=609; age 40-49 at baseline, n_{PA}= 49, n_W=854;

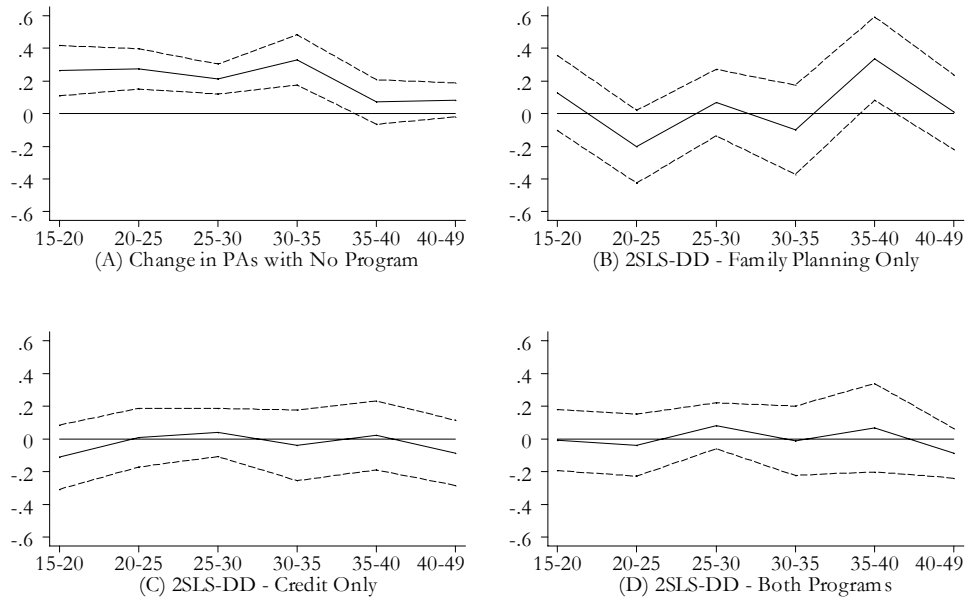

Figure 3

Oromia. Changes in contraception, by intervention and age group. Each continuous line joins comparable coefficients in the model of equation (2) estimated separately at the PA level with 2SLS using samples with women in different cohorts. Cohorts are labeled based on women's ages in 2003. The dashed lines represent the 95% confidence bands, calculated with heteroskedasticity-robust standard errors. The number of PAs (n_{PA}) and the total number of women (n_W) included in the estimation for each cohort-specific regression are as follows: age 15-20 at baseline, n_{PA}= 51, n_W=529; age 20-25 at baseline, n_{PA}= 77, n_W=1089; age 25-30 at baseline, n_{PA}= 78, n_W=1265; age 30-35 at baseline, n_{PA}= 72, n_W=834; age 35-40 at baseline, n_{PA}= 69, n_W=658; age 40-49 at baseline, n_{PA}= 68, n_W=698;

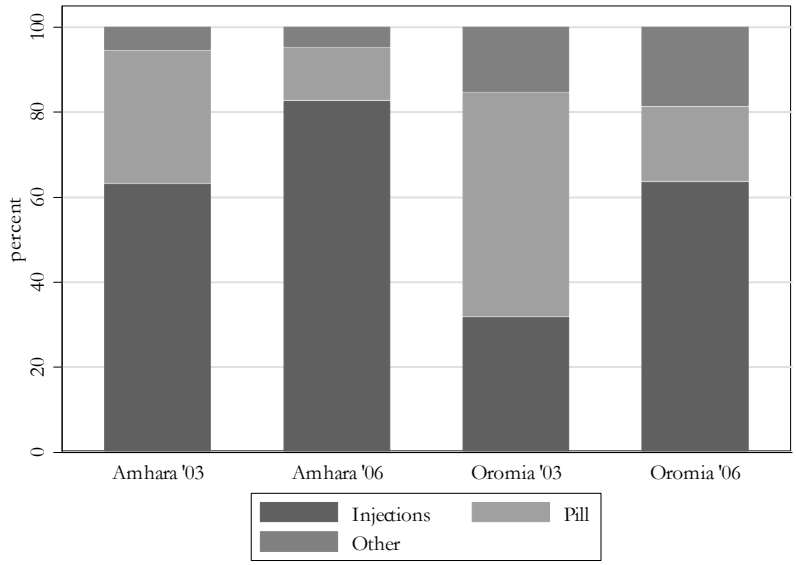

Figure 4
Contraceptives currently in use.

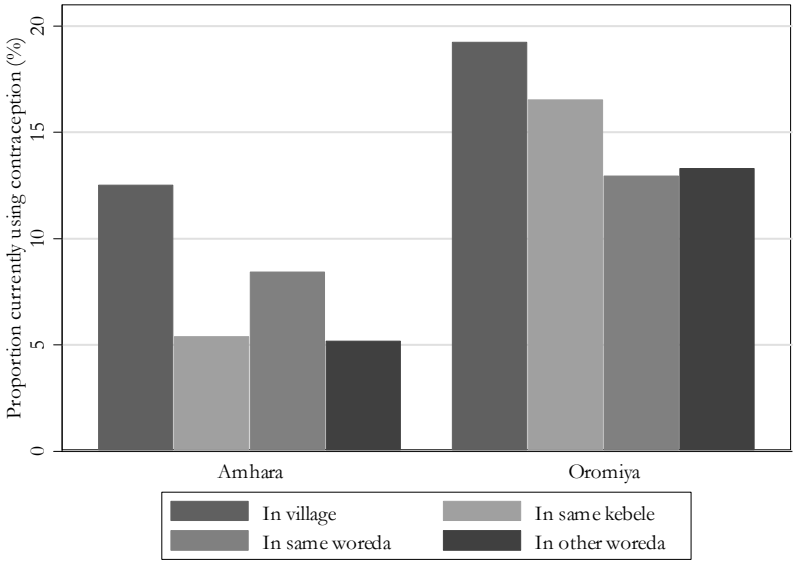

Figure 5
Proportion of women using contraception by location of nearest health center.

Table 1: Assigned *versus* Actual Treatment

	Actual Treatment				
Amhara					
Assigned Treatment	Both	Credit	FP	None	Total
Both	**12**	**1**	**1**	**1**	15
Credit	**2**	**9**	**0**	**3**	14
FP	**5**	**0**	**6**	**2**	13
None	**0**	**1**	**1**	**10**	12
Total	19	11	8	16	54
Oromia					
Assigned Treatment	Both	Credit	FP	None	Total
Both	**18**	**0**	**0**	**2**	20
Credit	**0**	**19**	**0**	**1**	20
FP	**6**	**0**	**12**	**0**	18
None	**0**	**3**	**0**	**17**	20
Total	24	22	12	20	78

Source: Authors' calculations from 2003 and 2006 household surveys and from service statistics data collected by the agencies implementing the study programs.

Table 2: Tests of randomization of PAs to assigned study groups (baseline survey 2003)

	Assigned groups					Assigned groups				
	Both	Credit	FP	None	p-value	Both	Credit	FP	None	p-value
	Amhara					**Oromia**				
All women 15-49 years of age										
Currently using contraception (%)	4.5	2.6	2.8	3.6	0.72	7.5	6.7	7.3	7.3	0.99
Intend to use FP in the future	46.1	48.9	41.0	50.0	0.02	72.8	68.9	70.6	72.0	0.55
Know of any method of FP (%)	86.5	88.3	77.7	82.6	0.36	73.5	79.7	81.2	76.6	0.37
Have heard of pills/injectibles (%)	60.6	60.3	52.6	59.3	0.71	42.0	44.4	44.0	49.8	0.61
No. of births in past 3 years (mean)	0.53	0.51	0.51	0.46	0.45	0.46	0.40	0.53	0.41	0.11
Desired number of children (mean)	4.9	4.8	5.1	4.6	0.12	4.4	4.3	4.6	4.4	0.27
Percent ever attended school	10.9	12.8	7.8	9.5	0.45	46.8	49.7	46.2	47.2	0.87
No. of women	877	937	875	886		1009	1080	1013	1035	
Currently married women										
Currently using contraception (%)	5.1	3.4	2.7	4.5	0.62	10.3	10.4	11.0	10.3	1.00
No. of births in past 3 years (mean)	0.65	0.65	0.60	0.55	0.11	0.65	0.62	0.77	0.59	0.13
Desired number of children (mean)	5.2	5.2	5.4	4.8	0.07	4.7	4.7	4.9	4.7	0.49
No. of women	665	666	682	662		646	638	644	672	
Women less than 25 years old										
Ever married (%)	77.2	68.2	76.9	75.8	0.32	39.1	33.5	34.0	41.2	0.34
Begun childbearing (%)	58.4	48.7	58.8	48.0	0.11	48.7	43.7	38.1	39.0	0.54
Have heard of pills/injectibles (%)	57.0	55.0	45.9	58.1	0.42	35.8	41.2	43.3	46.7	0.44
Desired number of children (mean)	3.99	3.85	4.07	3.90	0.80	3.82	3.57	3.86	3.65	0.15
No. of women	322	370	290	281		456	497	439	472	
Household Income and Economic Activity										
Crop cultivation (% of households)	87.2	90.3	90.7	89.2	0.84	93.7	91.4	92.0	91.3	0.63
Livestock (% of households)	74.6	73.4	80.4	77.7	0.31	49.2	44.6	48.0	47.5	0.90
Manufacturing (% of households)	6.2	7.5	2.0	5.5	0.14	1.9	4.9	3.1	3.3	0.45
Remittances from family/friends (%)	12.7	10.1	5.8	8.8	0.35	1.1	0.7	2.1	1.8	0.31
Crop sales (% of households)	51.3	54.9	59.2	51.5	0.71	30.5	36.7	34.8	38.2	0.54
Coffee cultivation (% of households)	3.5	6.4	0.3	0.5	0.31	37.4	46.5	54.8	47.7	0.16
Cows/Oxen/Bulls (mean no.)	1.9	1.8	2.2	2.0	0.60	1.2	1.1	1.5	1.2	0.56
Calves/Donkeys/Sheep/Goats (mean no.)	2.8	2.8	3.4	4.1	0.17	0.9	0.9	1.2	1.0	0.74
Chicken (mean no.)	1.8	1.7	0.9	1.4	0.10	0.9	0.8	1.3	0.7	0.19
# months of food insecurity last 12 mts	2.2	1.9	2.3	2.3	0.67	2.5	2.5	2.7	2.7	0.68
Household had debt past 12 mts (%)	16.4	25.5	16.0	14.8	0.34	13.9	17.2	14.9	23.6	0.23
Woman in hh had debt past 12 mts (%)	2.5	4.8	2.0	2.0	0.32	1.3	0.9	1.4	4.5	0.16
Households had debt towards revolving credit associations past 12 mts (%)	4.4	1.4	4.6	3.2	0.02	0.6	1.6	1.4	1.8	0.30
Minutes to nearest health facility	126	108	123	97	0.82	81	68	81	61	0.47
Household size	4.9	5.0	5.0	4.7	0.15	5.4	5.6	5.6	5.6	0.71
Religion – Christian Orthodox (%)	95.4	98.0	95.7	96.7	0.63	38.5	31.3	30.3	31.9	0.65
Religion – Christian Protestant (%)	0.0	0.0	0.3	0.0	0.13	41.4	55.6	56.3	52.4	0.35
Religion – Muslim (%)	4.1	1.6	3.7	2.7	0.69	16.9	11.3	11.6	13.0	0.82
Household head ever attended school (%)	12.3	9.7	9.1	11.0	0.63	56.8	54.7	50.7	53.7	0.53
No. of households	800	800	799	799		810	812	810	810	

Notes: Reported p-values refer to tests with a null of equality of means across the four treatment groups. Tests take into account the possible presence of clustering at the PA level.

Table 3: Assigned *versus* Actual Treatment

		(1)	(2)	(3)	(4)
		Amhara		Oromia	
		FP	MC	FP	MC
	PA randomly assigned family planning program	0.763	0.261	0.960	0.087
		[0.113]***	[0.112]**	[0.039]***	[0.095]
	PA randomly assigned micro-credit program	-0.065	0.615	-0.088	0.655
		[0.129]	[0.116]***	[0.054]	[0.104]***
(1)	% of women 15-49 who are pregnant	0.152	-0.539	0.328	-0.623
		[0.980]	[1.564]	[0.285]	[0.916]
(2)	Median # of live births in village (15-49)	0.066	-0.126	-0.015	-0.038
		[0.077]	[0.065]*	[0.030]	[0.050]
(3)	% women 15-49 who did not desire last birth	-2.162	0.219	0.135	1.097
		[1.915]	[1.882]	[0.362]	[0.765]
(4)	% women 15-49 whose husband did not desire last birth	1.51	0.484	0.530	-0.295
		[1.562]	[1.496]	[0.352]	[0.768]
(5)	Median desired family size among 15-49	0.095	0.038	0.028	0.168
		[0.169]	[0.128]	[0.046]	[0.115]
(6)	% women 15-49 who ever used the pill	-0.207	-0.869	-0.085	-0.588
		[1.848]	[1.877]	[0.245]	[0.839]
(7)	% women 15-49 who ever used injectibles	1.53	-3.206	0.512	-0.613
		[2.620]	[1.801]*	[0.788]	[1.409]
(8)	% women 15-49 who obtained contr. ¡ 60min away	-3.997	2.297	-1.159	-1.380
		[3.012]	[2.184]	[0.896]	[1.470]
(9)	% women 15-49 who know where to get contraceptives	0.171	0.235	0.090	0.466
		[0.248]	[0.465]	[0.082]	[0.307]
(10)	% women 15-49 visited by FP worker last 12m	-0.865	-1.012	0.269	0.168
		[0.678]	[0.491]**	[0.263]	[0.599]
(11)	% hhs who borrowed from a Revolving Credit Association	-0.545	-1.414	-0.012	0.221
		[0.789]	[0.880]	[0.186]	[1.124]
(12)	% hhs who borrowed from a Bank or Cooperative	0.483	2.103	0.298	0.413
		[1.149]	[0.976]**	[0.191]	[0.430]
(13)	% of individuals primarily working in agriculture	-0.808	-1.141	0.427	0.197
		[1.089]	[0.953]	[0.330]	[0.556]
(14)	% of individuals primarily working in animal husbandry	-1.306	0.074	1.530	0.811
		[0.774]	[1.070]	[1.060]	[1.637]
(15)	Median distance (in minutes) from nearest health center	0.001	-0.002	0.000	0.000
		[0.001]	[0.001]***	[0.001]	[0.002]
(16)	Median distance (min.) from nearest market	0.000	0.003	-0.001	-0.001
		[0.002]	[0.002]	[0.001]	[0.002]
(17)	Median # months of food insecurity	0.012	-0.089	-0.035	-0.032
		[0.060]	[0.057]	[0.032]	[0.063]
	Constant	0.287	1.497	-0.466	-0.600
		[0.954]	[0.756]*	[0.455]	[0.807]
	Observations	54	54	78	78
	R-squared	0.730	0.720	0.930	0.590
	Tests of joint significance (p-values)				
	All PA characteristics	0.00	0.00	1.000	0.710
	Fertility variables (1, 2)	0.69	0.14	0.440	0.660
	FP Demand (3, 4, 5, 6, 7)	0.83	0.37	0.750	0.300
	FP Supply (8, 9, 10, 15)	0.46	0.02	0.660	0.430
	Banking (11, 12)	0.78	0.03	0.300	0.600
	Economy (13, 14, 16, 17)	0.57	0.13	0.580	0.960

Notes: Authors' calculations from baseline survey (2003). Heteroskedasticity-robust standard errors in brackets. Statistical significance is indicated as * (10 percent), ** (5 percent) or *** (1 percent). The unit of observation is the peasant association (PA) and each statistic has been calculated using observations from sample households residing in a given PA. The figures in the last six rows are p-values for joint tests of significance (the figures in parenthesis indicate the variables whose coefficients are equal to zero under the null hypothesis).

Table 4: Demographic changes between baseline (2003) and follow-up survey (2006)

	Amhara			Oromia		
	2003	2006	Diff. in means p-value	2003	2006	Diff. in means p-value
Age-specific fertility rates (past 3 years)						
15-19	0.063	0.054	0.489	0.020	0.022	0.484
20-24	0.224	0.232	0.177	0.195	0.177	0.094
25-29	0.233	0.252	0.059	0.245	0.251	0.429
30-34	0.208	0.248	0.002	0.232	0.225	0.540
35-39	0.185	0.193	0.751	0.175	0.171	0.877
40-44	0.116	0.143	0.175	0.085	0.069	0.255
45-49	0.065	0.075	0.391	0.068	0.044	0.252
Total fertility rate	5.47	5.99	0.037	5.10	4.80	0.184
All women 15-49 years of age						
Currently using contraception (%)	3.5	12.2	0.000	7.4	21.1	0.000
Intend to use FP in the future	46.4	64.8	0.000	71.2	73.5	0.042
Know of any method of FP (%)	84.5	97.0	0.000	78.3	97.5	0.000
Have heard of pills/injectibles (%)	58.2	80.0	0.000	45.2	79.7	0.000
No. of births in past 3 years (mean)	0.51	0.55	0.003	0.45	0.42	0.094
Desired number of children (mean)	4.83	4.84	0.841	4.38	3.94	0.000
Percent ever attended school	10.5	18.4	0.000	47.3	57.6	0.000
No. of women	3575	3573		4137	4376	
Currently married women						
Currently using contraception (%)	4.0	14.8	0.000	10.9	32.8	0.000
No. of births in past 3 years (mean)	0.62	0.68	0.001	0.65	0.64	0.601
Desired number of children (mean)	5.11	5.27	0.003	4.70	4.35	0.000
No. of women	2675	2677		2600	2665	
Women less than 25 years old						
Ever married (%)	74.2	62.9	0.000	36.7	30.0	0.000
Begun childbearing (%)	53.4	47.9	0.006	41.5	27.9	0.000
Have heard of pills/injectibles (%)	54.4	75.8	0.000	41.2	77.3	0.000
Desired number of children (mean)	3.91	3.48	0.000	3.70	3.21	0.000
No. of women	1263	1215		1864	1945	

Notes: Reported p-values refer to tests with a null of equality of means between the two periods. Tests take into account the possible presence of clustering at the PA level. The total fertility rate is estimated as the total number of children that a woman would have between 15 and 49 years of age, given the fertility rates observed in the sample for the different age groups indicated in the table.

Table 5: Socio-economic changes between baseline survey (2003) and follow-up survey (2006)

	Amhara			Oromia		
	2003	2006	Diff. in means p-value	2003	2006	Diff. in means p-value
Crop cultivation (% of households)	89.2	93.2	0.000	92.8	96.9	0.000
Livestock (% of households)	76.2	86.0	0.000	48.1	72.9	0.000
Manufacturing (% of households)	5.7	5.7	0.959	2.7	8.7	0.000
Remittances from family/friends (%)	9.6	5.8	0.000	1.4	2.2	0.143
Crop sales (% of households)	54.3	65.4	0.000	36.5	70.4	0.000
Coffee cultivation (% of households)	2.6	4.7	0.020	48.3	72.7	0.000
Cows/Oxen/Bulls (mean no.)	2.0	2.7	0.000	1.3	2.1	0.000
Calves/Donkeys/Sheep/Goats (mean no.)	3.3	5.3	0.000	1.0	2.1	0.000
Chickens (mean no.)	1.5	2.0	0.000	0.9	2.2	0.000
Total value of livestock (2003 Birr)	1351	2723	0.000	651	1943	0.000
# months of food insecurity last 12 months	2.2	1.7	0.000	2.6	0.9	0.000
Household had debt past 12 mts (%)	18.0	44.2	0.000	17.5	36.9	0.000
Woman in household had debt past 12 mts	2.8	10.1	0.000	2.0	13.7	0.000
Households had debt towards revolving credit associations past 12 mts	3.4	19.8	0.000	1.5	18.5	0.000
Minutes to nearest health facility	113	94	0.000	74	68	0.001
Age-specific school attendance rate (6-10 years)	17.3	41.7	0.000	36.2	45.0	0.000
Age-specific school attendance rate (11-14 years)	39.5	66.9	0.000	66.7	79.2	0.000
Age-specific school attendance rate (15-18 years)	28.4	51.5	0.000	48.1	62.9	0.000

Notes: Data from 2003 and 2006 household surveys. Reported p-values refer to tests with a null of equality of means between the two periods and are robust to clustering at the PA level.

Table 6: Impact of interventions on current use of contraception among currently married women

	(1)	(2)	(3)	(4)	(5)	(6)
	Amhara			Oromia		
	OLS-DD	2SLS-DD	2SLS-DD	OLS-DD	2SLS-DD	2SLS-DD
α_0 Intercept	0.128	0.124	-0.209	0.231	0.241	0.994
	[0.022]***	[0.031]***	[0.173]	[0.023]***	[0.024]***	[0.254]***
α_1 Both	-0.007	-0.007	0.01	0.001	-0.023	-0.066
	[0.028]	[0.042]	[0.031]	[0.036]	[0.045]	[0.043]
α_2 Credit	-0.066	-0.036	-0.044	-0.016	-0.027	-0.076
	[0.027]**	[0.060]	[0.051]	[0.041]	[0.050]	[0.042]*
α_3 FP	-0.026	-0.039	-0.067	-0.036	-0.029	-0.03
	[0.036]	[0.066]	[0.053]	[0.037]	[0.047]	[0.049]
Additional controls	No	No	Yes	No	No	Yes
Obs.	54	54	54	78	78	78
R^2	0.1			0.01		
Tests (p-values)						
(1) H_0: $\alpha_1=\alpha_2=\alpha_3=0$	0.0500	0.9332	0.5330	0.7500	0.8662	0.3031
(2) H_0: $\alpha_1=\alpha_2$	0.0193	0.6178	0.2652	0.6887	0.9357	0.7821
(3) H_0: $\alpha_1=\alpha_3$	0.5607	0.6596	0.1406	0.3470	0.9316	0.5590
(4) Exogeneity		0.6160	0.1802		0.5130	0.0671
(5) Underidentification		0.0057	0.0051		0.0001	0.0000

Notes: All coefficients are 2SLS estimates of the model in DD form, estimated at the PA level (see equation (2) in the text). The regressions in columns 1, 2, 4 and 5 include only the intercept and dummies for actual treatment status as regressors, while the regressions in columns 3 and 6 also include controls for the pre-intervention levels of the same 17 PA characteristics used to evaluate systematic deviation from protocol compliance in Table 3. The full set of results which include the estimated slopes for the controls are available upon request from the authors. The standard errors in brackets are heteroskedasticity-robust. Asterisks indicate significance at the 10 (*), 5 (**) or 1 (***) percent level. In rows denoted (4), the null hypothesis is that the three dummies indicating actual treatment are exogenous. In row (5) we report the results of Kleibergen-Paap tests for the null hypothesis that the equation is underidentified. Both tests are heteroskedasticity-robust.

Table 7: Impact of interventions, other fertility outcomes

	Intend to use FP		No. of methods heard of		Births past 3 years		Desired family size	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Amhara								
α_0 Intercept	0.163	0.574	0.477	-1.571	0.036	0.708	-0.213	-2.132
	[0.046]***	[0.542]	[0.350]	[3.404]	[0.041]	[0.390]*	[0.195]	[1.497]
α_1 Both	0.004	-0.037	0.606	0.304	-0.005	0.005	0.165	0.396
	[0.061]	[0.077]	[0.471]	[0.448]	[0.064]	[0.061]	[0.267]	[0.277]
α_2 Credit	-0.044	-0.146	-0.32	-0.65	0.004	-0.112	0.976	0.891
	[0.103]	[0.129]	[1.121]	[0.815]	[0.124]	[0.097]	[0.535]*	[0.551]
α_3 FP	0.208	0.19	0.817	0.643	0.159	0.132	-0.239	-0.21
	[0.131]	[0.126]	[0.724]	[0.626]	[0.157]	[0.095]	[0.621]	[0.485]
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	54	54	54	54	54	54	54	54
Tests (p-values)								
(1) H_0: $\alpha_1=\alpha_2=\alpha_3=0$	0.2155	0.0240	0.2326	0.3502	0.7538	0.0644	0.1778	0.1267
(2) H_0: $\alpha_1=\alpha_2$	0.6285	0.4078	0.3982	0.1834	0.9466	0.2930	0.1246	0.3494
(3) H_0: $\alpha_1=\alpha_3$	0.1409	0.0665	0.7934	0.6009	0.3414	0.2266	0.5392	0.2329
(4) Exogeneity	0.0374	0.0213	0.2261	0.1902	0.4111	0.1240	0.0822	0.1356
(5) Underidentification	0.0068	0.0049	0.0057	0.0051	0.0059	0.0051	0.0051	0.0049
Oromia								
α_0 Intercept	0.008	0.406	1.712	4.479	0.088	0.024	-0.626	0.39
	[0.038]	[0.196]**	[0.289]***	[1.312]***	[0.043]**	[0.531]	[0.124]***	[0.637]
α_1 Both	-0.013	-0.067	0.416	0.342	-0.151	-0.154	0.123	0.286
	[0.051]	[0.042]	[0.361]	[0.291]	[0.061]**	[0.068]**	[0.178]	[0.166]*
α_2 Credit	0.027	-0.087	-0.104	-0.301	-0.106	-0.108	0.285	0.378
	[0.056]	[0.051]*	[0.384]	[0.346]	[0.056]*	[0.075]	[0.149]*	[0.173]**
α_3 FP	0.089	0.05	-0.418	-0.507	-0.165	-0.166	0.374	0.417
	[0.061]	[0.051]	[0.458]	[0.391]	[0.126]	[0.089]*	[0.191]*	[0.166]**
Additional controls	No	Yes	No	Yes	No	Yes	No	Yes
Obs.	78	78	78	78	78	78	78	78
Tests (p-values)								
(1) H_0: $\alpha_1=\alpha_2=\alpha_3=0$	0.4320	0.0415	0.1811	0.0713	0.0637	0.0562	0.1586	0.0542
(2) H_0: $\alpha_1=\alpha_2$	0.3752	0.6331	0.0633	0.0232	0.3583	0.4322	0.2268	0.5937
(3) H_0: $\alpha_1=\alpha_3$	0.1252	0.0277	0.0690	0.0271	0.9179	0.9086	0.2764	0.5169
(4) Exogeneity	0.3493	0.2401	0.3277	0.2329	0.2449	0.1559	0.7949	0.5090
(5) Underidentification	0.0002	0.0000	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000

Notes: All coefficients are 2SLS estimates of the model in DD form, estimated at the PA level. The regressions in columns 1, 3, 5 and 7 include only the intercept and dummies for actual treatment status as regressors, while the regressions in columns 2, 4, 6 and 8 also include controls for the pre-intervention levels of the same 17 PA characteristics used to evaluate systematic deviation from protocol compliance in Table 3. The full set of results which include the estimated slopes for the controls are available upon request from the authors. The standard errors in brackets are heteroskedasticity-robust. Asterisks indicate significance at the 10 (*), 5 (**) or 1 (***) percent level. In rows denoted (4), the null hypothesis is that the three dummies indicating actual treatment are exogenous. In row (5) we report the results of Kleibergen-Paap tests for the null hypothesis that the equation is underidentified. Both tests are heteroskedasticity-robust.

Table 8: Demand for Contraception

	n	Mean	s.d.	Median	min	max
Amhara	55	0.19	0.10	0.19	0.00	0.38
Oromia	78	0.36	0.15	0.37	0.02	0.81

Notes: Authors' calculations from 2003 survey. Demand for contraception is constructed as the fraction of ever-pregnant women of child-bearing age in the PA who (or whose partner) would have preferred to avoid or delay the current pregnancy (if any) or the previous one (if any).

Table 9: Program Impacts and Demand for Contraception, Amhara

	(1) Current FP use	(2) Intend to use FP	(3) No. of methods heard of	(4) Births past 3 years	(5) Desired family size
Both (α_1)	0.239	0.351	1.169	-0.234	-0.615
	[0.496]	[0.563]	[2.900]	[0.902]	[1.543]
Credit (α_2)	0.001	0.847	-0.7	0.1	-0.905
	[0.303]	[0.758]	[2.300]	[0.517]	[1.787]
FP (α_3)	-0.074	0.71	0.255	0.54	-0.94
	[0.888]	[0.677]	[4.606]	[1.587]	[1.949]
Demand (α_4)	-1.653	-1.138	-5.688	2.249	3.748
	[3.884]	[4.143]	[21.584]	[7.027]	[11.152]
Both × Demand (α_5)	0.568	-5.36	5.21	-1.886	10.363
	[3.197]	[4.540]	[18.311]	[5.557]	[10.262]
Credit × Demand (α_6)	2.488	-4.231	12.465	-6.649	3.9
	[12.054]	[8.917]	[61.657]	[21.492]	[24.127]
FP × Demand (α_7)	0.396	2.166	-1.672	0.374	-3.425
	[0.833]	[1.857]	[6.867]	[1.533]	[4.645]
Intercept (α_0)	-0.007	-0.255	0.684	0.03	0.463
	[0.157]	[0.327]	[1.363]	[0.296]	[0.879]
Obs.	54	54	54	54	54
Tests (p-values)					
(1) H_0: $\alpha_1=\alpha_2=\alpha_3=0$	0.7508	0.4482	0.4155	0.9895	0.8755
(2) H_0: $\alpha_5=\alpha_7=0$	0.7491	0.5617	0.9636	0.9499	0.7473
(3) Kleibergen-Paap	0.7554	0.7308	0.7554	0.7542	0.7754

Notes: Each column reports the results of a DD regression where the dependent variable is the fertility-related outcome indicated on top of each column. All regressions are estimated with 2SLS, using as instruments the randomly determined treatment assignment as well as treatment interacted with demand. Demand for contraception is measured as the fraction of ever-pregnant women of child-bearing age in the PA who (or whose partner) would have preferred to avoid or delay the current pregnancy (if any) or the previous one (if any). Standard errors, in brackets, are adjusted for heteroskedasticity. Statistical significance is indicated by * (10 percent), ** (5 percent) and *** (1 percent). The unit of observation in the regressions is the PA and all regressions are weighted using the number of observations in each PA as weight. The Kleibergen-Paap test in the last row of the table tests the null hypothesis of underidentification.

Table 10: Program Impacts and Demand for Contraception, Amhara

	(1) Current FP use	(2) Intend to use FP	(3) No. of methods heard of	(4) Births past 3 years	(5) Desired family size
Both (α_1)	-0.065	0.159	0.11	-0.309	0.203
	[0.151]	[0.290]	[0.893]	[0.311]	[0.659]
Credit (α_2)	-0.236	0.353	0.299	-0.094	0.229
	[0.160]	[0.229]	[1.143]	[0.282]	[0.575]
FP (α_3)	0.067	0.164	0.096	0.141	0.729
	[0.094]	[0.173]	[0.719]	[0.276]	[0.407]*
Demand (α_4)	0.112	-0.469	0.84	0.422	-0.223
	[0.384]	[0.730]	[2.451]	[0.746]	[1.661]
Both × Demand (α_5)	0.623	-0.966	-1.212	-0.036	0.171
	[0.486]	[0.639]	[3.371]	[0.766]	[1.647]
Credit × Demand (α_6)	-0.272	-0.269	-1.316	-0.856	-1.043
	[0.241]	[0.441]	[2.291]	[0.937]	[1.138]
FP × Demand (α_7)	-0.008	0.527	-1.11	-0.103	0.535
	[0.185]	[0.460]	[1.648]	[0.392]	[1.006]
Intercept (α_0)	0.244	-0.173	2.092	0.123	-0.81
	[0.076]***	[0.171]	[0.557]***	[0.155]	[0.365]**
Obs.	78	78	78	78	78
Tests (p-values)					
(1) H_0: $\alpha_1=\alpha_2=\alpha_3=0$	0.1757	0.2677	0.9943	0.6627	0.2440
(2) H_0: $\alpha_5=\alpha_7=0$	0.4653	0.7961	0.6472	0.4990	0.5999
(3) Kleibergen-Paap	0.0275	0.0235	0.0275	0.0278	0.0249

Notes: Each column reports the results of a DD regression where the dependent variable is the fertility-related outcome indicated on top of each column. All regressions are estimated with 2SLS, using as instruments the randomly determined treatment assignment as well as treatment interacted with demand. Demand for contraception is measured as the fraction of ever-pregnant women of child-bearing age in the PA who (or whose partner) would have preferred to avoid or delay the current pregnancy (if any) or the previous one (if any). Standard errors, in brackets, are adjusted for heteroskedasticity. Statistical significance is indicated by * (10 percent), ** (5 percent) and *** (1 percent). The unit of observation in the regressions is the PA and all regressions are weighted using the number of observations in each PA as weight. The Kleibergen-Paap test in the last row of the table tests the null hypothesis of underidentification.

Table 11: Presence of non-ADA/ODA family planning services in study villages (%)

	Both	FP	Credit	None	Total
Amhara					
(1) Health center, Health Post, Health Extension workers	33.3	33.3	18.2	20.8	26.5
(2) Other sources	7.0	16.7	3.0	37.5	16.7
(3) No non-ADA sources	59.7	50.0	78.8	41.7	56.8
No. of villages	57	24	33	48	162
Chi-square test of association (p-value)			0.000		
Oromia					
(4) Health center, Health Post, Health Extension workers	50.0	38.5	38.2	42.9	43.2
(5) Other sources	5.0	11.5	12.7	4.1	7.9
(6) No non-ODA sources	45.0	50.0	49.1	53.1	49.0
No. of villages	60	26	55	49	190
Chi-square test of association (p-value)			0.535		

Source: authors' calculations from post-intervention (2006) community questionnaire data. Figures in rows 1-3 and 4-6 indicate the fraction of villages, reported by actual treatment group, with the presence of the family planning services indicated. In the chi-square tests, the null hypothesis is that the fractions are identical among the four actual treatment groups.

Table 12: Number of family planning methods women are aware of and current contraceptive use

	Both		Credit		FP		None		Total	
	Awareness of FP methods									
	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N
Amhara										
Did not borrow	3.6	603	3.4	282	4.0	334	3.8	661	3.7	1,880
Borrowed from ACSI/OCSSCO	4.4	399	3.5	176					4.1	575
Borrowed from other source	3.8	260	3.1	136	3.8	175	4.0	425	3.8	996
Total	3.9	1,262	3.3	594	3.9	509	3.9	1,086	3.8	3,451
Oromia										
Did not borrow	4.5	779	4.2	589	4.5	446	4.3	699	4.4	2,513
Borrowed from ACSI/OCSSCO	4.8	331	4.7	367					4.7	698
Borrowed from other source	4.7	235	4.0	154	4.6	147	4.6	301	4.5	837
Total	4.6	1,345	4.3	1,110	4.6	593	4.4	1,000	4.5	4,048
	Current contraceptive use									
Amhara										
Did not borrow	9.8	603	4.8	282	8.4	334	12.4	661	9.6	1,880
Borrowed from ACSI/OCSSCO	15.0	399	6.6	176					12.4	575
Borrowed from other source	14.2	260	6.7	136	15.7	175	23.9	425	17.2	996
Total	12.3	1,262	5.8	594	10.8	509	16.9	1,086	12.2	3,451
Oromia										
Did not borrow	18.5	779	18.1	589	22.1	446	22.4	699	20.2	2,513
Borrowed from ACSI/OCSSCO	24.7	331	26.0	367					25.4	698
Borrowed from other source	18.1	235	25.2	154	21.2	147	19.9	301	20.6	837
Total	20.0	1,345	21.7	1,110	21.9	593	21.6	1,000	21.2	4,048