FAMILY NETWORKS AND SCHOOL ENROLMENT: EVIDENCE FROM A RANDOMIZED SOCIAL EXPERIMENT*

MANUELA ANGELUCCI† GIACOMO DE GIORGI‡

MARCOS A.RANGEL§ IMRAN RASUL¶

Abstract

We present evidence on whether and how a household's behavior is influenced by the presence and characteristics of its extended family. Using household panel data from the *Progresa* social assistance program in rural Mexico, we first exploit information on the paternal and maternal surnames of heads and spouses in conjunction with the patronymic naming convention to identify the inter and intra generational family links of each household to others in the same village. We then exploit the randomized research design of the *Progresa* evaluation data to identify whether the treatment effects of *Progresa* transfers on secondary school enrolment vary according to the presence and characteristics of extended family members. We find that *Progresa* only raises secondary enrolment among households that are embedded in an extended family network. Eligible but isolated households do not respond. The key mechanism through which the extended family influences household schooling choices relates to the redistribution of resources within the family network towards those households on the margin of enrolling children into secondary school.

Keywords: extended family network, *Progresa*, resource sharing, schooling.

JEL Classification: I21, J12, O12.

---

*This research was supported by an IRB approval from the University of Chicago. The paper has been screened to ensure no confidential information is revealed. We thank Orazio Attanasio, Oriana Bandiera, Jere Behrman, Timothy Besley, Martina Bjorkman, Pedro Carneiro, Pierre-Andre Chiappori, Hilary Hoynes, Sylvie Lambert, Steve Machin, Ofer Malamud, Alan Manning, Guy Michaels, Debraj Ray, Jean-Marc Robin, Yoram Weiss, and seminar participants at Chicago, Columbia, IFS, INRA, LSE, Maryland, Toulouse, UCL, BREAD/CESifo Conference (Venice 2006), and the CEPR/EUDN Conference (Paris 2006) for useful comments. All errors remain our own.

†Department of Economics, University of Arizona [angelucm@eller.arizona.edu].

‡Department of Economics, Stanford University [degiorgi@stanford.edu].

§Harris School of Public Policy, University of Chicago [rangelm@uchicago.edu].

¶Department of Economics, University College London [i.rasul@ucl.ac.uk].

1 Introduction

Economists usually focus on the nuclear family as the unit of analysis from which to study household behavior. There are good reasons for this. Theories of household decision making, such as the unitary model [Becker 1981], bargaining models [Manser and Brown 1980, McElroy and Horney 1981], and collective choice models [Chiappori 1988] all emphasize the interplay between the preferences and resources of household members in shaping outcomes. In addition, survey data is typically constrained in the sense that it is not possible to identify the familial ties, or any other ties, between any two surveyed households.

However, each household is actually embedded within a wider network of its extended family members. If extended families shape the objectives and constraints relevant for households within them, then neglecting the role of the extended family when analyzing household behavior may lead to an incomplete understanding of the forces driving household choices. The extended family may be especially relevant for household behavior in economic environments characterized by missing markets, correlated income shocks, the prevalence of informal institutions of contract enforcement, and policy interventions that affect many households in the local economy.

All these features apply to the empirical context of this paper, which uses household panel data from the *Progresa* social assistance program in rural Mexico. This program provides cash transfers to households conditional on their childrens' school attendance. In this paper we first exploit information on the paternal and maternal surnames of heads and spouses in conjunction with the patronymic naming convention in Mexico to identify the inter and intra generational family links of each household to others in the village. We then exploit the randomized design of the *Progresa* evaluation data to identify whether the treatment effects of *Progresa* transfers on school enrolment vary according to the presence and characteristics of extended family members.

Two key intuitions underlie our analysis. The first is that *Progresa* exogenously shifts the resources available to individual households *and* to their family network as a whole. The second is that families are able to enforce implicit contracts of resource sharing. Hence the program can induce differential responses between households embedded within family networks and those that are socially isolated in the sense that none of their extended family live in close proximity and are not therefore subject to the change in resources that *Progresa* provides.[1]

While undoubtedly other households outside the extended family network also influence be-

[1] Our analysis relates to the literature on risk pooling arrangements across households. While such *ex post* risk pooling mechanisms have been documented to exist in a number of developing country settings, formal tests of the Pareto efficient allocation of risk being achieved at the village level are typically rejected [Townsend 1994, Ligon 1998, Dercon and Krishnan 2000]. However there is stronger evidence of risk pooling within ethnic groups [Deaton 1992, Udry 1994], sub-castes [Munshi and Rosenzweig 2005], and family and friends [Rosenzweig 1988, Fafchamps and Lund 2003, La Ferrara 2003, Cox and Fafchamps 2007].

havior, there are good reasons to focus on family networks, rather than friendship networks say, as being the core group within which resources are shared. First, evolutionary biology suggests preferences are defined over the family dynasty, as is commonly modelled within an overlapping generations framework for example. Moreover there are specific inter-generational investments – such as those into children's education, bequests, and the choice of marriage partners – that have no counterpart in relationships between friends, and provide long run incentives for family members to reciprocate in resource sharing arrangements. Finally, the transactions costs of sharing resources with non-family members may be higher because it is both more costly to observe outcomes outside the family network, and fewer mechanisms exist by which to punish non-family members that renege on such arrangements [La Ferrara 2003].

In our descriptive analysis we document the proportions of households that are connected and isolated, and whether husbands and wives differ in the extent to which members of their extended family live in close proximity. In the econometric analysis we identify whether and why responses to *Progresa* in terms of secondary enrolment differ for connected and isolated households, and among connected households, whether and why the characteristics of specific intra-generational and inter-generational family links matter for the response. We use three key features of the data in order to provide evidence on these issues.[2]

First, we combine information on the paternal and maternal surnames of heads of households and their spouses, with the patronymic naming convention in Mexico to build two types of extended family link outside the household but within the village – (i) inter-generational links, such as those from the head and spouse to their parents, and to their adult sons and daughters; (ii) intra-generational links, such as those from the head and spouse to their brothers and sisters. We combine this with information from the household roster to identify extended family members that co-reside in the household. These sources of information together provide an almost complete mapping of extended family structures across 506 villages in rural Mexico, covering around 22,000 households and over 130,000 individuals.

Second, we exploit the fact that *Progresa* has multiple components, each of which provides cash transfers conditional on a different household behavior. One component provides cash transfers to households conditional on the school attendance of children in primary and secondary school on at least 85% of class days. However, as pre-program primary school enrolment rates are above 90%, transfers provided for this purpose operate like a *de facto* unconditional cash transfer and so represent a pure income effect on households with primary school aged children. In contrast pre-

[2] We therefore contribute to the literature on the effects of extended family on household behavior, such as for consumption [Altonji *et al* 1992]; inter-generational transfers [Cox and Jakubson 1995, Altonji *et al* 1997, La Ferrara 2003, Behrman and Rosenzweig 2006]; childrens' education choices [Loury 2006]; and non-resident parental investments into children [Weiss and Willis 1985].

program secondary enrolment rates are closer to 65% so that for many households cash transfers will be obtained only if a change in behavior is induced. This is important because as the value of transfers only corresponds to between one half to two thirds of the full time child wage in the survey villages [Schultz 2004], the value of transfers conditional on secondary enrolment alone do not fully compensate for foregone earnings of children employed full time in the labor market.

Hence an important consequence is that the program effect on secondary enrolment may be a function of the presence of primary school aged children, who receive *de facto* unconditional cash transfers. More specifically, as the number of primary school aged children in the household increases, the household can use these unconditional transfers to supplement their transfers specifically conditioned on secondary school enrolment, thus fully offsetting the opportunity costs of enrolling children into full time secondary school. This channel affects both isolated and connected households. In addition, if families share resources, and in particular they share the unconditional transfers obtained from the primary school component of *Progresa*, the response of connected households will also depend on the demographic composition of eligible households within their family network. This channel drives a wedge between the behavioral responses of connected and isolated households to the program in terms of their secondary school enrolment.

The third key feature of the data is the randomized research design used to evaluate *Progresa*. Of the 506 sampled villages, 320 were randomly assigned to be a treatment group, namely villages where *Progresa* would be implemented, and 186 villages were controls. In each village, household data was collected on a panel of around 22,000 households every six months over the pre and post-program periods. Within each village the baseline survey provides a complete census of households. This covers all households that are either eligible (poor) for any *Progresa* transfers, and all households that are not eligible (not poor) for any component of the program.

This research design allows us to first identify the average treatment effect of the program from a comparison of eligibles in treatment and control villages. Second, we identify whether this treatment effect varies between connected and isolated households by comparing eligibles with a given set of extended family characteristics across treatment and control villages. Third, this research design also enables us to identify an indirect treatment effect of being connected or isolated from a comparison of non-eligibles in treated and control villages. This sheds light on the existence of within village spillovers of the program on secondary school enrolment. Non-eligibles may change behavior in response to the program either because extended family networks span eligible and non-eligible households and resource transfers take place between all households in the network, or because of general equilibrium effects of the program.

Our main descriptive findings are as follows. First, 20% of couple headed households are isolated in the sense that none of their extended family reside within the village, while 80%

are connected and so are embedded within an extended family network. As a benchmark for comparison, we note that 15% of households are single headed. Second, there are however no significant pre-program differences between isolated and connected households in terms of their enrolment rates and poverty levels. Third, among connected households, the extended family of the head is more likely to reside in the same village than the extended family of his spouse. This difference is explained by there being higher levels of female migration at the time of marriage.[3]

We have four main econometric results. First, despite baseline enrolment rates being similar in connected and isolated households, only connected households respond to the program in terms of secondary school enrolment. The average treatment effect of *Progresa* on them is significantly different from zero at around 9%. In contrast, eligible but isolated households do not respond – their treatment effect varies between -.2% and .9% and is never significantly different from zero.[4]

Our second set of results provide direct evidence on the interplay between the design features of *Progresa*, whether a household is connected or isolated, and its response to the program. In particular, we estimate how program responses vary with the intensity of treatment in terms of the monetary value of transfers received by the household and the extended family, and the demographic composition of children in the household and the extended family as a whole. We find that connected households are sensitive to the total value of cash transfers that flow into their own household and the households of eligible members of their extended family. More subtly, the marginal response of connected households depends on the demographic composition of children both in their household and in their family network.

When the minority of family members are both eligible and have primary school aged children, so there are relatively few unconditional resources to redistribute, only connected households with primary school aged children themselves significantly increase secondary enrolment. Connected households that have no primary school aged children do not respond to the program. In contrast, when the majority of family network members are both eligible and have primary school aged children, connected households with and without primary school children are sensitive to the inflow of resources to their own household and their extended family. Taken together, these results point to resources being redistributed within family networks towards households that are on the margin of enrolling children into secondary school, namely those households that themselves receive effectively unconditional *Progresa* transfers for the primary school enrolment of their children.

Third, we find no evidence of any indirect treatment effects – on average, secondary school enrolment rates do not differentially change among non-eligibles between treatment and control villages, for either connected or isolated households. The lack of evidence for within family spillovers

[3] This is similar to the findings of Rosenzweig and Stark [1989] who examine marital arrangements in rural India.

[4] We build on earlier findings on the effects of *Progresa* on enrolment [Schultz 2004, Attanasio *et al* 2005, Todd and Wolpin 2006]. This literature has not emphasized the influence of extended families on schooling choices.

in secondary enrolment implies that even if recipient households redistribute some resources to non-eligibles in the family, then either – (i) the amounts transferred are small; or, (ii) recipient households do not use the resources to increase secondary enrolment.[5]

Finally, we complement our main results by considering which specific family members share resources and why. Among sibling links, household responses are sensitive to the eligibility status of their sibling, and in particular, whether eligible siblings obtain unconditional transfers from having primary school aged children. Households are less sensitive to their sibling having secondary school aged children *per se*, suggesting there are not strong within sibling peer effects specifically related to secondary schooling. Evidence of such peer effects are however found when we analyze how household behavior is influenced by the presence of those that have similar observable characteristics to their true siblings. Among parent-adult child links, parental households are again sensitive to the eligibility status of their adult children and the presence of younger children in their households. Taken together, the results on the influence of specific family links consistently point to resources being redistributed within the network from eligible households that receive largely unconditional transfers for the primary school enrolment of their children, towards eligible households that are on the margin of enrolling children into secondary school.

The paper is organized as follows. Section 2 describes the *Progresa* program and data. Section 3 discusses how we construct family links and provides descriptive evidence on them. Section 4 presents the empirical method and supportive evidence on the underlying identifying assumptions. Section 5 presents our baseline results on whether and why being connected influences program responses. Section 6 presents evidence on how specific family links shape schooling decisions. Section 7 concludes with a discussion of the policy implications of our findings. Additional descriptive evidence and robustness checks are in the Appendix.

2 The Progresa Program and Evaluation Data

2.1 Transfers

Progresa is an ongoing publicly funded social assistance program in Mexico. It is designed to alleviate poverty by fostering human capital accumulation through two channels. First, it provides cash transfers to households conditional on the school attendance of children in primary and secondary school grades. Second, it provides cash transfers and nutritional supplements conditional on attendance at health programs in local facilities. This component of *Progresa* targets benefits

[5] Using a different empirical method to ours, Bobonis and Finan [2006] and Cattaneo and Lalive [2006] present evidence of positive *ITEs* on schooling in treated villages. We return to this issue in the conclusion.

towards households with pregnant or lactating women, and with children aged less than five.[6]

In this paper we focus on how households with secondary school age children respond to the conditional cash transfers provided by *Progresa*. Central to our analysis is the fact that the extended family members of any given household can be eligible for transfers even if they have no secondary school age children themselves. For example, households with primary aged children, very young children and those with no children but a pregnant women in them, can be eligible for the primary school, health, and nutritional components respectively. Hence if families share resources, the response to *Progresa* of households with secondary school age children will in general be determined by the demographic characteristics of *all* their eligible family members, not just the subset of members that also have secondary school aged children.

The school enrolment based cash transfers are paid bimonthly and conditional on the child attending school classes at least 85% of the previous 60 days. There are two important details to note on this program component. First, transfers are larger for higher school grades, and higher for girls within any given grade. The average monthly transfer to beneficiary households is non-negligible, corresponding to 20% of the value of monthly consumption expenditures pre-program [Skoufias 2005].[7] On the other hand, the value of transfers only corresponds to between one half to two thirds of the full time child wage in the survey villages [Schultz 2004]. Hence the most credit constrained households should not be induced to change their schooling choices by such transfers alone. However, given the value of transfers related to primary school are around one third of those for secondary school, a household with primary school aged children has the possibility to use these transfers, in conjunction with transfers specifically conditioned on secondary school enrolment, to fully offset the opportunity cost of sending a children to secondary school.

Moreover, if extended families share resources, this provides a direct mechanism by which households embedded in family networks might respond to the program to a greater extent, in terms of secondary enrolment, than isolated households. In particular, resource sharing within the family network may allow connected households to retain their children in secondary school when they otherwise would have dropped out in the absence of *Progresa*, and to re-enrol their children back into secondary school if they are engaged in labor market activity in the pre-program period.

Second, although both the education and health components are conditioned on household behavior, the extent to which households need to change behavior to obtain the transfers depends

[6] The program was initially offered to 140,544 households in 1997, expanding to more than 2.6 million recipient households throughout rural Mexico by the end of 1999. This constitutes around 40% of all rural families and one ninth of all families in Mexico. The total annual budget of the program in 1999 was equivalent to just under 20% of the federal poverty alleviation budget or .2% of GDP [Skoufias 2005]. The program was expanded to urban areas in 2003 under the name of *Oportunidades*. We do not exploit this expansion into urban areas for our analysis.

[7] By November 1999 the bimonthly transfer ranged from 160 *pesos* for third grade, to 530 (610) *pesos* for boys (girls) in ninth grade. The total amount received bimonthly by a household cannot however exceed 1500 *pesos*.

on their pre-program choices. As documented in Section 4, pre-program primary enrolment rates are above 90%. Hence transfers provided to enrol children into primary school effectively represent a pure income effect on households. In contrast pre-program secondary enrolment rates are closer to 65%. Hence given the value of transfers conditional on secondary enrolment do not compensate for the losses of income earned by children in the labor market, the likelihood the program increases secondary enrolment rates will depend partly on the demographic composition of children in the household. In addition, if families share resources, the response of connected households will in general depend on the demographic composition of eligible households within their entire family network. This channel drives a wedge between the behavioral responses of connected and isolated households to the program in terms of the secondary school enrolment.

Finally, the health and nutrition components of the program require the periodic attendance of mothers at local health clinics. Households are likely to view this component as somewhere between the purely unconditional transfers obtained for primary schooling, and the conditional transfers obtained for secondary schooling. Again it will be the case that the response in terms of secondary school enrolment will in general be influenced by the presence of young children in the household to whom this component is targeted, and for connected households, the presence of young children in eligible households within the family network as a whole.[8]

2.2 Eligibility and The Evaluation Data

In 1997 households were classified as either being eligible (poor) or non-eligible (not poor) for *Progresa* transfers according to a household poverty index. This index is a weighted average of household income (excluding children), household size, durables, land and livestock, education, and other physical characteristics of the dwelling. The index is designed to give relatively greater weight to correlates of permanent income rather than current income.

Around half the households were classified as poor and therefore eligible for *Progresa*. Households were informed that their eligibility status would not change at least until November 1999, irrespective of any variation in their household income. Moreover, the monetary value of transfers that eligibles were entitled to was determined by the age and gender composition of the children resident in the household at baseline. There are therefore no incentives for eligibles to foster children from non-eligibles in the hope of obtaining greater transfers. Finally, a feature of *Progresa* that distinguishes it from many other social assistance programs is that households were

[8] The required frequency of attendance varies depending on the age of the individual. For example, children younger than four months are required to attend three check ups, those aged between 4 and 24 months attend eight check ups, children aged between 2 and 4 attend every four months, children aged between 5 and 16 attend every six months, pregnant women attend five check ups during their prenatal period, lactating women attend two check ups, and adults aged 17 and above are required to attend an annual check up [Skoufias 2005].

clearly informed and made aware of their eligibility status – for example through village assemblies. Hence take-up rates among eligibles are over 90%, and we do not therefore distinguish between intent-to-treat and treatment effects of *Progresa* in our analysis.[9]

To evaluate the effectiveness of *Progresa*, an experimental research design was implemented and household data collected on a panel of around 22,000 households every six months in 506 villages between March 1998 and November 1999.[10] Of the 506 villages, 320 were randomly assigned to the treatment group, namely locations where *Progresa* would be later implemented in May 1998, and 186 villages were assigned to be control villages. The control group therefore contains 36.8% of all villages in the evaluation data, corresponding to 38.2% of all households.

The first two waves of data were collected pre-program (October 1997, March 1998). Transfers were first distributed in May 1998, hence the remaining waves (October 1998, May 1999, November 1999) correspond to the post-program period. To understand whether household behavior is influenced by the presence and characteristics of extended family members of both the head and spouse, we focus attention on the 85% of households that are couple headed throughout.[11]

3 Constructing Extended Family Links

3.1 Surnames and the Matching Algorithm

To identify the extended family links between households in the same village we exploit information on surnames provided in the third wave of data. We combine this information with the patronymic naming convention in Mexico to build two types of family link – (i) intra-generational family links, such as those from the head (spouse) of the household to his (her) brothers and sisters; (ii) inter-generational family links, such as those from the head (spouse) of a household to his (her) parents,

[9] A group of households – referred to as *densificados* – had their eligibility status reclassified from non-eligible to eligible in October 1998. A non-random subset of them began receiving *Progresa* transfers in treatment villages prior to November 1999. As no precise algorithm exists to determine which *densificados* received transfers in treatment villages, no counterfactual set of households exists for them in control villages. As we can define extended family links to and from these households, all the reported descriptive statistics on extended families include links to and from *densificados*. We do not consider changes in enrolment among *densificados* in our analysis.

[10] These villages are located in seven out of thirty one states in Mexico – Guerrero, Hidalgo, Michoacan, Puebla, Queretaro, San Luis Potosi, and Veracruz. Villages were selected on the basis of a marginality index constructed from information on the share of illiterate adults in the village; share of dwellings without water, drainage systems, electricity, and with floors of dirt; average number of occupants per room in village households; share of population working in the primary sector; distance from other villages, and health and school infrastructures in the village.

[11] Control villages began receiving *Progresa* transfers in December 1999. In 1997, eligible households in control villages were informed they would become part of the program at the end of 1999 conditional on them still being eligible and the program continuing.

and from the head and spouse of the household to their adult sons and daughters.[12]

Mexicans use *two* surnames – the first is inherited from the father's paternal lineage and the second from the mother's paternal lineage. For example, former Mexican president Vicente Fox Quesada would be identified by his given name (Vicente), his father's paternal name (Fox) and his mother's paternal name (Quesada). In the evaluation data, respondents were asked to provide the – (i) given name; (ii) paternal surname; and, (iii) maternal surname, for each household member. Hence couple headed households have four associated surnames – the paternal and maternal surnames of the head, and the paternal and maternal surnames of his wife.[13]

Figure 1 provides an illustration of the matching algorithm. To define each family link, we use information on *two* of the four surnames. Consider household **A** at the root of the family tree. The head of the household has paternal and maternal surnames $F1$ and $f1$ respectively. His wife has paternal and maternal surnames $F2$ and $f2$ respectively.[14]

The children of the couple in household **A** will adopt the paternal surnames of their father ($F1$) and mother ($F2$). Hence we define there to be a parent-son relationship between households **A** and **B** if – (i) the paternal surname of the head in household **B** is the same as the paternal surname of the head in household **A** ($F1$); and, (ii) the maternal surname of the head in household **B** is the same as the paternal surname of the spouse in household **A** ($F2$). Parent-daughter relationships can be similarly defined. Moreover, intra-generational family ties between siblings can also be identified. For example, the heads of households **B** and **C** are identified to be brothers if they

[12] Two concerns arise from the surnames data being measured in the first wave of post-program data. First, households may endogenously respond to the program by changing household structures, in particular, by artificially forming new households in order to increase the number of eligibles in the family. This concern is ameliorated by the fact that the register of eligible households was drawn up at baseline, and only households recorded to be eligible at that point were later entitled to receive transfers. Moreover, although there is an increase in the number of households from the baseline to October 1998, this increase is proportionately the same in both treatment and control villages. A second concern is that the program may affect the migration of the household head or of his spouse. However, only .4% (.5%) of households in wave 3 (5) report having a migrant head or spouse. Moreover the share of households with such migrants does not differ across treatment and control villages.

[13] The precise wording of the question in Spanish is, "*Dígame por favor el nombre completo con todo y apellidos de todas las personas que viven en este hogar, empezando por (jefe del hogar) – (i) nombre; (ii) apellido paterno; (iii) apellido materno*". We cleaned the surnames data as follows – (i) we removed non-alphabetical characters, replaced "Sin Apellido" (no surname) with missing values, and corrected some obvious typos based on intra-household surname checks; (ii) we imputed a small number of missing female surnames from wave 2; (iii) we verified surnames using the same information from wave 5, and verified the relationship to the household head using wave 1 data. No information on surnames is available in the first wave of data. The head of household is originally defined to be the main income earner. In a very small number of cases the head of a couple headed household is reported to be a women. To keep clear the exposition, we redefine the head to be male in such cases.

[14] We use the convention that the head's surnames are written in black, and those of his wife are written in red italics. Paternal surnames are indicated in upper case and maternal surnames are indicated in lower case. First names are not shown as they are not relevant for the construction of extended family links. Each household in the family tree is assumed to be couple headed purely to ease the exposition. In Anglo Saxon countries, $F1$ corresponds to the family name and $F2$ corresponds to the spouse's maiden name.

share the same paternal and maternal surnames.

In Figure 1 we assume households are couple headed solely to ease the exposition. To deal with the 15% of households that are single headed we use information on the gender of the head to accurately define each family link. Finally, we impose the following restrictions when defining family links – (i) inter-generational links exist when the relevant individuals have at least 15 years age difference, and no more than 60 years age difference between mother and child; (ii) intra-generational links exist when the individuals have at most 30 years age difference.

However, there are limits to which information on surnames can be used to construct family ties. Consider links from household i to a single headed household j. As Figure 1 shows, the fact that household j is single headed does not affect the construction of links from the head and spouse of household i either to their children or to their siblings. However, links from the head (spouse) of household i to the household of his (her) parents can only be identified if *both* his (her) parents are alive and resident together. This is because this particular family link is identified using information from household j on the paternal surnames of both the head and spouse.[15]

Finally, we only define extended family links within the village because *Progresa* is implemented at the village level. Hence geographically proximate family members are those most relevant to understand household behavior if families share the resources provided by *Progresa*. Isolated households may well have extended family elsewhere and share resources with them. However, their family is unlikely to experience the change in resources specifically due to the *Progresa* intervention because the evaluation data cover a time period when the program was still being rolled out in rural Mexico.

3.2 Measurement Error in Extended Family Links

There are a number of potential forms of measurement error in the surnames data that can be checked for. The first arises from the convention that women change their paternal surname to their husband's paternal surname at the time of marriage. To address this concern, we note that the precise wording of the question specifically asks respondents to name the paternal and maternal surname of each household member. Furthermore, in only 5.8% of households is the spouse's maternal surname recorded to be the same as her husband's paternal surname. This provides an upper bound on the extent to which measurement error of this form is occurring.

Second, if the male head is the respondent, he may not recall his wife's maternal surname and simply replace it with her paternal surname. This may occur because his children only inherit his

[15] However this is unlikely to be a major issue. For example we note that female widows aged above 40 are 37% more likely to live as a dependent within a household, rather than head their own household, relative to a similar married woman. These single parents are then recorded in the household roster.

wife's paternal surname. Reassuringly, this problem occurs in only 4.9% of households. A final circumspect case is households in which the paternal and maternal surnames of both the head and spouse are all reported to be the same. This occurs for 1.6% of households, although the figure drops to .5% if we exclude households with the most common surname in the data.[16]

Some forms of measurement error however cannot be addressed. The first arises from any remaining typos in surnames. Second, there may be two identical families in the village who share the same paternal and maternal surnames of head and spouse but are genuinely unrelated. The matching algorithm then assigns the number of family links to be double what they actually are. A check for the severity of this problem is based on the following intuition. By definition, household i cannot have parental links to more than two other households (the parent's of the head and the parent's of the spouse), conditional on the parents not being present within the household. This is true for 97% of households using our matching algorithm. Third, consider a scenario in which a women's brother marries someone with the same maternal surname as himself. Then the woman's niece will be identified as her sister and although the households are within the same family network, the strength of their tie may be inferred to be stronger than it actually is.

With these caveats and concerns in mind, it is equally important to reiterate that all links are defined across households on the basis of *two* surname matches and the fact that these two surnames map precisely to two of the four surname types. For the econometric analysis, we only exploit information on whether such family links exist for household i within the village, and not the number of such links. We also show the robustness of our main results to dropping households with potential measurement error in their surnames, and to limiting the sample to smaller villages where spurious links between households are less likely to be defined.

3.3 The Number of Extended of Family Links

Table 1 provides details on the number of family links each household has to others in the village. The columns of the upper panel split family links into – (i) inter-generational links to parents and adult children who head their own households; (ii) intra-generational links to siblings who head their own households. We report each type of link from the head and spouse separately. The lower panel reports the number of corresponding family links that co-reside inside the household as measured from the household roster in wave 3. Using information on the relationship of each person to the household head, we can decompose these links into those from the head and spouse.[17]

To ease comparisons across the panels, each type of family link is reported separately for connected and isolated households. By definition, the upper panel reports that isolated households

[16] There are no differences in the incidence of these potential errors between treatment and control villages.

[17] The Appendix provides descriptive evidence on each of the individual surnames used to construct family links.

have no links to family members within the same village. The first important point to note is that 20% of couple headed households are isolated. The incidence of not having extended family members geographically proximate in the same village is therefore at least as high as the incidence of single headedness, which affects 15% of households.

We now consider links to each type of extended family member beginning with links to parents. Table 1 shows that parents are many times more likely to reside outside the household than inside the household of their adult children. The number of parents present is higher for the head than for his spouse, and this is true for parents inside and outside the household. This is consistent with either the spouse migrating to the village, or women moving in with their husband's family within the same village. To shed more light on this we exploit data on spouse's marital history. Wives were asked about where they went to live after marriage – 49.3% stated that they went to live with their in-laws after marriage, and only 6.5% report living with their own parents. The key difference between spouses with and without parents resident in the village is that 85% of spouses that have their parents present in the village report remaining in the same village at the time of marriage. The figure for spouses that have no parental links in the village is only 61%. Along other margins, women in connected households with and without parents in the village are similar. For example, they do not differ significantly in their ages at marriage, nor in the proportions that report their in-laws originally proposing the marriage (56%).[18]

The number of links to adult children are, by construction, identical for head and spouse. Given respondents' ages, there are many more young children inside than adult children outside the household. Sibling links are more likely to be outside the household, which is as expected given the age of respondents. Heads have more siblings links than their spouses, and this is again true for links both inside and outside the household.[19]

Overall, the upper panel shows that the average household has family ties to just over five other households in the same village, and that the majority of these extended family links are those of male heads of household. A key source of this difference arises from women moving households and villages at the time of marriage.

One concern with the constructed extended family data is that isolated households may simply be those in which all family members reside under the same roof. The lower panel of Table 1 shows

[18] Rosenzweig and Stark [1989] examine marital arrangements in rural India and also find women move to distant villages at the time of marriage. They argue that such marriages can be viewed as a consumption smoothing device. We leave for future research a more complete analysis of the relationship between the marriage market and consumption smoothing in rural Mexico.

[19] In October 1998 the average age of heads (spouses) among couple headed households is 45.0 (40.5). The total number of siblings of the head is on average 2.23, implying his parents have 3.23 children that reside within the same village. In contrast, heads have on average 4.70 of their own children somewhere in the village. Again this is as expected given that siblings are older than adult children and so will be more likely to have migrated.

this is not the case. Although isolated households are slightly more likely to have the parents and siblings of the head reside within them, the differences with connected households are not large, and the number of within household parental and sibling links are orders of magnitude smaller than the corresponding links outside the household. Overall, the average household has around seven members and this is not different between connected and isolated households.[20]

Another candidate explanation of how isolated and connected households differ relates to geographic mobility. While isolated and connected households are equally likely to report being resident in the same state of birth, unfortunately, the data does not contain any direct information on how long individuals have been resident in the village. More indirectly, we note that among isolated households, only 52% of spouses report living in same village as at time of marriage, which is lower than that for spouses in connected households as reported above.

Finally, we note that connected and isolated households are as likely to report receiving remittances from family members that have permanently migrated away in the five years prior to the baseline survey. Reassuringly, this suggests isolated households *do* have family located somewhere, and may share resources with them.[21] The key difference for our analysis in terms of program responses is that their families are less likely to receive *Progresa* resources since they live in different villages. This drives a wedge between the program response of isolated and connected households, even if all families share resources.[22]

To provide external validity to this constructed data, in the Appendix we present similar information from an alternative data source that was collected in a comparable economic environment and time period. The *Mexican Family Life Survey* (MxFLS), collected in 2001, provides information on the numbers of each type of family link, by head and spouse, that are alive in *any* location, not just the same village. The MxFLS therefore provides upper bounds on what should be recorded as family links in the *Progresa* data. Reassuringly, we find that in the MxFLS, the number of family links to parents, adult children, and siblings outside the household and located anywhere, are indeed greater than those we construct within the village.

[20] We can also decompose the overall variation in each statistic in Table 1 into the variation between villages, and the variation across households in the same village. In each case the within village variation is around ten times as great as the between village variation. Hence variations in family structure occur within the same village, and it is not that there are systematically different types of family structures in each village.

[21] There is however no information on pre-program transfers from other family members outside the village, including temporary migrants or family members in the place of birth if the household has itself migrated. Hence it is not possible to make inferences about whether the transactions costs of across village transfers are higher than within village transfers, due to asymmetric information say. It is the case that the transactions cost of other forms of exchange – such as labor transfers – are weakly higher across villages than within them, all else equal.

[22] Munshi and Rosenzweig [2005] provide evidence from India that those who migrate away from their sub-caste lose the services of that network, including mutual insurance arrangements. The model they develop and test implies the wealthiest households are those with incentives to withdraw from such arrangements. However, in general, the relationship between wealth and exit is theoretically ambiguous [Banerjee and Newman 1998].

4 Empirical Method

4.1 Preliminaries

The aim of the empirical analysis is to estimate the response of eligible and non-eligibles to *Progresa* in terms of their secondary enrolment rates, and to explore whether these responses vary depending on the presence and characteristics of extended family members. To begin with, we denote the secondary enrolment rate of household h in village v in wave t as Y_{hvt}. This is defined as the fraction of children aged between 11 and 16 resident in the household that are enrolled in school on the survey date in wave t. The behavioral response we focus on is the change in secondary enrolment within the same household over time, between November 1999 (wave 5) and October 1997 (wave 1), denoted ΔY_{hvt}. This corresponds to more than one academic year after the program is initiated in treatment villages.[23]

We define three dummy variables. First, $D_h = 1$ if household h is eligible, and zero otherwise. Second, we define $P_{vt} = 1$ if *Progresa* is in place in village v in wave t, and zero otherwise. Hence $P_{vt} = 0$ ($P_{vt} = 1$) in the pre (post) program waves in treatment villages, and $P_{vt} = 0$ in control villages for all t. Third, we define $L_{jh} = 1$ if household h has family link-j in the village, and zero otherwise. Although in the long run we expect family networks to endogenously adjust to the permanent presence of *Progresa*, in this paper we treat these networks as being fixed over the relatively short period of time considered.[24]

We then run the following first differenced OLS regression separately for non-eligible and eligibles where the differences correspond to those between November 1999 and October 1997,

$$\Delta Y_{hvt} = \alpha + \beta_1 \Delta P_{vt} + \beta_2 \left(\Delta P_{vt} \times L_{jh}\right) + \Delta u_{hvt}, \tag{4.1}$$

and Δu_{hvt} corresponds to unobserved time varying household characteristics. Time invariant household determinants of enrolment – such as household preferences or ability – are differenced out in this specification. Standard errors are clustered by village as that is the level at which *Progresa* is implemented. To ease the exposition, we consider the case in which $L_{jh} = 1$ if the household has *any* extended family links present and $L_{jh} = 0$ if the household is isolated in the

[23] The enrolment rate is constructed from individual child observations that have complete information on the child's age, gender, and residence. Waves 1 and 5 are collected during the school year which runs from September to July. This eases concerns that households mis-report enrolment status either because children are on school vacation or have dropped out part way through the academic year. We later also use primary enrolment rates for some of the analysis – these are analogously defined for 6 to 10 year olds.

[24] In support of this, we reiterate that the share of households reporting a migrant head or spouse by is only .4% (.5%) in wave 3 (5), and that the share of households with such migrants does not differ across treatment and controls. In the Appendix we show our main results to be robust to redefining family networks on the basis of additional information collected in May 1999.

sense that no members of its extended family live in the same village.

When (4.1) is estimated for eligibles, $\widehat{\beta}_1+\widehat{\beta}_2$ identifies the average treatment effect of *Progresa* on eligibles in treatment villages that are embedded within extended family networks ($L_{jh}=1$). We denote this parameter as TTE^1. This difference-in-difference (DD) in secondary enrolment is identified from the comparison of the average changes in enrolment among eligibles in treatment villages with extended family present, relative to eligibles in control villages who also have extended family present. Similarly, $\widehat{\beta}_1$ identifies the average response to the program among eligible but isolated households ($L_{jh}=0$), denoted TTE^0, from the comparison of the DD in enrolment of isolated households in treatment and control villages. If families share resources, there exists a wedge between the behavioral responses of connected and isolated households to the program in terms of their secondary school enrolment, so that $TTE^1 > TTE^0$.

When (4.1) is estimated for non-eligibles, $\widehat{\beta}_1+\widehat{\beta}_2$ identifies the average indirect treatment effect of *Progresa* on non-eligibles in treatment villages that are embedded within extended family networks. We denote this parameter as ITE^1. This DD in school enrolment is identified from the comparison of the average changes in enrolment among non-eligibles in treatment villages with extended family, relative to eligibles in control villages who also have extended family present. Similarly, when (4.1) is estimated for non-eligibles, $\widehat{\beta}_1$ identifies the average indirect treatment effect of the program among non-eligible and isolated households, ITE^0. These $ITEs$ shed light on whether there exist within village spillovers of the program on secondary school enrolment. Spillovers onto non-eligibles may arise either because extended family networks span eligible and non-eligible households and resource transfers take place between all households in the network, or because of general equilibrium effects of the program.

4.2 Unconditional Levels and Changes in Enrolment Rates

The left hand panel in Table 2 provides descriptive evidence on the pre-program levels of secondary school enrolment rates, and how they vary over time in treatment and control villages by eligibility status. While enrolment rates among eligibles at baseline are similar across treatment and control villages, by November 1999 eligibles in treatment villages have 6.9% higher enrolment rates than eligibles in control villages. This DD in enrolment rates, estimated from the unconditional specification in (4.1), is positive and significant for both boys and girls enrolment, and consistent with previous studies, the proportionate increase is greater for girls. In line with the earlier literature, we find pre-program differences in enrolment among non-eligibles across treatment and control villages [Schultz 2004]. However, as shown in the next section, these differences disappear once observables are conditioned on.[25]

The central panel of Table 2 splits the sample between connected and isolated households. This reveals the following information. First, eligible isolated and connected households have similar baseline levels of enrolment. Among non-eligibles, baseline enrolment rates are actually higher among isolated households. This suggests that determinants of the *levels* of enrolment – such as credit constraints, information on the costs and benefits of schooling, or demand for child labor in the home – do not explain differences in the schooling choices between isolated and connected households at baseline. Moreover, this suggests isolated households are at least as well off – in terms of secondary enrolment – as connected households. Isolated households should not therefore be viewed necessarily as the most vulnerable households in society. This notion is further reinforced by the fact that isolated and connected households do not differ in their poverty index at baseline, a measure of permanent income, as shown in Figure A2.[26]

Second, despite their similarities at baseline, there are significant differences in the program response of isolated and connected households. The DD in enrolment rates among eligible and connected households is 8.3% and significantly different from zero. In contrast, the DD for isolated households is close to zero. In short, the previously documented positive effects of *Progresa* on secondary enrolment are a combination of a large and significant effect on the 80% of connected households, and a close to zero effect on the 20% of isolated households. In line with this, we note that take-up rates for transfers conditional on school enrolment – either secondary or primary – are consistently higher among connected households. In the first two months after the program is initiated, connected households have take-up rates that are .8% higher, and this difference rises to 4.0% by November/December 1999.

Third, among non-eligibles, neither connected nor isolated households in treated villages have any significant changes in enrolment vis-à-vis analogous households in control villages.

Taken together the results suggest that although the baseline levels of enrolment between isolated and connected households are similar, there exists an important interplay between the physical presence of the extended family, and the behavioral response of households to the conditional cash transfers provided by *Progresa*.

The last panel in Table 2 shows the baseline levels and changes over time in primary enrolment

[25] Recall that the program design is such that households have no incentives to foster children in order to obtain greater transfers. Indeed, we note that these changes in enrolment rates are driven by changes in the number of children enrolled in school, and not by the number of children resident in the household. The pre-program differences in enrolment rates of non-eligibles are reduced if the *densificados* – namely those households that had their eligibility status changed in October 1998 – are also used to calculate baseline enrolment rates.

[26] Such factors may explain heterogeneous responses within connected and isolated households, but not the differences between them. Consistent with this we note that the variance in baseline enrolment rates, or the variance in the household poverty index, are not different between connected and isolated households.

rates for connected and isolated households. The baseline levels of primary enrolment are very high to begin with, and neither types of eligible household significantly increase their primary enrolment rates relative to analogous households in control villages. This confirms that transfers from this component of the program are obtained by households without inducing any changes in behavior, and therefore essentially act as a pure income effect.

This opens up the possibility that if a household receives a sufficiently large amount of such unconditional transfers, they can use these to supplement transfers conditioned specifically on secondary enrolment to overcome the opportunity costs of enrolling children in secondary school. Moreover, if resources are shared within family networks, connected households have access to these unconditional transfers from family members. If these resources are channelled towards members with secondary school aged children, there should be a wedge between the response of connected and isolated households in terms of secondary school enrolment, as suggested by the descriptive evidence in Table 2.

4.3 Identification

Equation (4.1) makes precise the assumptions required for the four parameters of interest, TTE^1, TTE^0, ITE^1, ITE^0, to be identified. First, we require there to be no spillover effects from treatment to control villages, so that household behavior is driven by whether they reside in a treatment or control village and not on the status of other villages. Second, we require there to be random assignment of villages into treatment and control groups. This implies whether a household is in a treatment or control village, is orthogonal to unobservables that drive its response to *Progresa* ($Cov(\Delta P_{vt}, \Delta u_h) = 0$), and to the presence of its extended family ($Cov(\Delta P_{vt}, L_{jh}) = 0$).

These twin assumptions of no cross village spillovers and random assignment are standard requirements for the identification of TTE and ITE effects. Together they imply the change in enrolment for eligibles and non-eligibles in the absence of treatment in treatment villages would have been, on average, identical to the change in enrolment in the absence of the program in control villages. In other words, eligibles (non-eligibles) in control villages provide a valid counterfactual for eligibles (non-eligibles) in treatment villages.[27]

Third, there may be other time varying factors that would differentially drive enrolment rates in treatment and control villages, even in the absence of *Progresa*. Given random assignment, these omitted variables will be orthogonal to ΔP_{vt} but not necessarily to L_{jh}. To address this we control for a series of characteristics of the head, spouse, household, and village, in all specifications.

[27] In support of the first identifying assumption, we note that villages were, in part, included in the evaluation data because they were geographically remote. In terms of the second identifying assumption, it is has been previously documented that household and village characteristics do not significantly differ in treatment and control villages suggesting random assignment [Schultz 2004, Behrman *et al* 2005].

Moreover we allow there to be a direct impact of the family link-j on secondary enrolment rates to allow for the effect of having family present in the village to change over time, say because of changes in other public assistance programs. We therefore estimate the following specification for eligibles and non-eligibles,

$$\Delta Y_{hvt} = \alpha + \beta_1 \Delta P_{vt} + \beta_2 (\Delta P_{vt} \times L_{jh}) + \beta_3 L_{jh} + \boldsymbol{\lambda}' \mathbf{X}_{hv} + \Delta u_{hvt}, \tag{4.2}$$

where $\mathbf{X}_{hv}$ includes characteristics of the head, spouse, household, and village. Hence we estimate program responses conditional on any potential differences between isolated and connected households such as the characteristics of their heads, spouses, or the villages in which they reside.[28]

Fourth, to identify whether the $TTEs$ and $ITEs$ vary with regards the presence of extended family links of type-j, namely to consistently estimate β_2, requires an additional assumption that the presence of link type-j is uncorrelated with unobservables that drive the response to *Progresa* ($Cov(L_{jh}, \Delta u_h) = 0$). However a central concern is that β_2 captures two effects – (i) that the response to the *Progresa* program truly differs according to whether a household is isolated or embedded within a family network; (ii) connected and isolated households differ in characteristics that drive responses to *Progresa* and are correlated to the presence of the extended family.

To isolate the impact of the extended family *per se*, we need to purge the estimates of factors that both drive the presence of extended family and responses to *Progresa*. To do this we proceed as follows. First, we estimate what are the correlates of extended family link of type-j being present in the village or not – this is detailed in the Appendix. These results, reported in Table A3, highlight the following characteristics are robust predictors of whether extended family members reside in the same village or not – whether the head's (spouse's) age is above or below the median among couple headed households, whether the head (spouse) is literate, whether the household owns land, and whether household size at baseline is above or below the median among couple headed households. Using this analysis to guide our approach we then allow for a household's response to *Progresa* to be heterogeneous along each of these dimensions. We then estimate the following specification for eligibles and non-eligibles,

$$\Delta Y_{hvt} = \alpha + \beta_1 \Delta P_{vt} + \beta_2 (\Delta P_{vt} \times L_{jh}) + \beta_3 L_{jh} + \boldsymbol{\lambda}' \mathbf{X}_{hv} + \sum_i \gamma_{2i} Z_i + \sum_i \gamma_{3i} (\Delta P_{vt} \times Z_i) + \Delta u_{hvt}, \tag{4.3}$$

[28] These controls are the head and spouse's ages, literacy status, whether (s)he speaks an indigenous language, the household poverty index, whether the household owns any land, and household size at baseline. At the village level we control for the number of households in the village to capture any scale effects, the share of households that are eligible to capture any aggregate income effects, the marginality index for the village, and the village level enrolment rate at baseline among eligible and non-eligibles. This may correlate to distances to school facilities for example. Hence we do not exploit data from villages that have all eligible or all non-eligible households so the $TTEs$ and $ITEs$ are identified from the same set of villages. As randomization into treatment and control groups takes place within region, we control for regional fixed effects throughout.

where Z_i refers to each dimension along which we allow the response to *Progresa* to be heterogenous, in addition to allowing for heterogeneous responses with the presence of the extended family link type-j. We additionally allow Z_i to include the average village level enrolment rates among eligible and non-eligibles, as measured at baseline, to pick up any differential trends in enrolment by eligibility, across treatment and control villages. Comparing the estimates from specifications (4.2) and (4.3) then indicates whether there are likely to exist unobserved characteristics, not captured in Z_i, that drive the differential responses of connected and isolated households.[29]

We have so far emphasized the identification and interpretation of four parameters of interest. However, the differences between these $TTEs$ and $ITEs$ are also of economic interest. Estimating (4.3) among eligibles for example, the difference between TTE^1 and TTE^0, which corresponds to a triple difference, captures the differential effect of *Progresa* between eligible connected and isolated households within treatment villages, conditional on characteristics that are correlated to the existence of extended family, and potentially other sources of heterogeneous responses to *Progresa*. Similarly, the difference between ITE^1 and ITE^0 captures the differential effect of *Progresa* between non-eligible connected and isolated households within treatment villages.

The remaining econometric concerns relate to unobservables that – (i) are correlated to the existence of extended family; (ii) cause heterogeneous responses to *Progresa*; and, (iii) do not drive the baseline levels of enrolment to be different across connected and isolated households (Table 2). In the Appendix we present a series of robustness checks that allow program responses to vary with more village level characteristics, we explore the robustness of the results in alternative subsamples of villages, and address concerns over unobserved household level characteristics such as economic shocks driving enrolment changes.[30]

4.4 Evidence in Support of the Identifying Assumptions

Table 3 shows how the probability of having an extended family link of type-j varies by eligibility across treatment and control villages. This sheds light on whether one of the identifying assumptions, $Cov(\Delta P_{vt}, L_{jh}) = 0$, is likely to hold in the data. Consistent with the earlier descriptive evidence we see that – (i) 80% of households are connected, 20% are isolated; (ii) links from the head of the household are significantly more likely to exist than links from his spouse; (iii) intra-generational links are significantly more common than inter-generational links. This pattern of

[29] As the effects of *Progresa* are allowed to be heterogeneous along other dimensions Z_i, the $TTEs$ and $ITEs$ are evaluated at the mean value of each variable in Z_i.

[30] A concern over the interpretation of the results stems from *Progresa* having a direct effect on fertility. If so the opportunity cost of female time will change and this may in turn drive schooling responses. Todd and Wolpin [2006] develop and test a dynamic structural model of schooling and fertility choices, and find that fertility rates are insensitive to the value of *Progresa* transfers, a result confirmed using a reduced form approach [Schultz 2004].

links holds for both eligible and non-eligibles in treatment and control villages. Reassuringly, in nearly all cases there are no significant differences in the extended family links of both eligible and non-eligible households between treatment and control villages.[31]

To provide further support for the identifying assumption, Table 4 provides descriptive evidence on the characteristics of family networks as a whole by treatment and control villages.[32] There are 1379 (817) family networks in treatment (control) villages covering 10559 (6471) households. In treatment villages, the first column shows that on average there are around 7.6 households in each family. The average village has around seven family dynasties so each dynasty encompasses 16% of all village households. The third column provides information on the diameter of the network – the largest distance between any two households in the network. This is around 2.4, implying family networks are unlikely to span more than three generations as expected.

The remaining columns of Table 4 highlight how the characteristics of family networks vary. We consider characteristics that relate specifically to the design of the program as discussed in Section 2.1, and therefore provide sources of variation from which to understand how households embedded within family networks may respond to changes in the resources available to the network as a whole. To begin with, we see that family networks span eligibility status – there are an almost equal number of eligibles and non-eligibles in the average family. Hence there is considerable scope for *Progresa* transfers to be redistributed within the family.

Family networks also vary in the share of households within them that have primary and secondary school aged children. Hence, as shown in the final column, the potential value of transfers each household is eligible for and hence the amount of resources to be redistributed within the family, varies considerably between networks.[33]

Finally, Table 4 also decomposes the variation in each statistic into that arising between family networks across different villages, and across family networks within the same village. In nearly all cases, there is more variation within networks in the same village than across villages. This is important given that identification arises from an across village comparison of eligibles (non-eligibles) with a given set of family characteristics.

[31] The one exception is the proportion of non-eligibles with links from the spouse to her brother being significantly higher in controls. Our later analysis however focuses on the influence of same gender siblings.

[32] To be precise, suppose the heads of households i and j are linked because they are brothers. Household j may itself be linked to household k, say, because the parents of the spouse of household j reside in household k. Households i and k then lie within the same family network, even though they do not have a *direct* family link between them. Households i and j are then said to be of distance one from each other and households i and k are of distance two from each other. Two households i and j are then defined to be within the same family network if the distance between them, d_{ij}, is finite. Isolated households do not contribute to these network statistics.

[33] Note that these figures relate to the potential value of transfers to all eligible households in the family network. The values for the subset of eligible households with secondary school aged children are higher. We later exploit this information in the empirical analysis.

5 Does The Presence of Extended Family Matter?

5.1 Baseline Results

To begin with, Column 1a of Table 5 benchmarks our estimates against the existing literature by reporting standard TTE and ITE estimates of *Progresa* from specification (4.2), by pooling eligibles and non-eligibles and interacting ΔP_{vt}, L_{jh}, and $(\Delta P_{vt} \times L_{jh})$ with the dummy for eligibility status, D_h. These are averaged across households irrespective of their extended family structure. The TTE estimate implies eligibles have a 7.8% increase in their secondary enrolment rate compared to eligibles in control villages, which is similar in magnitude to the previously documented program effects [Schultz 2004].

The ITE estimate implies non-eligibles in treatment and control villages have similar changes in their secondary enrolment. This suggests that, on average, there are no significant spillover effects of *Progresa* on the secondary enrolment of non-eligibles.

Columns 1b and 1c split the sample by connected and isolated households. Consistent with the earlier evidence, we find that only connected households respond to *Progresa*. Conditional on observables, eligible connected households in treatment villages increase enrolment rates by 9.8% more than analogous households in control villages. In contrast, eligible isolated households do not respond to the program on this margin.

Column 2a estimates all four parameters of interest from a single regression. The result confirms that the TTE of *Progresa* is actually comprised of a large and significant effect on eligibles that have extended family members present in the village ($TTE^1 > 0$), and a negligible and non significant effect among isolated households ($TTE^0 = 0$).

There is no evidence of any $ITEs$ on average for either households in family networks or isolated households. The fact that the ITE^1 is close is to zero itself implies there are no within family spillovers in secondary enrolment rates. If recipient households redistribute some fraction of their transfers to non-eligibles within the same family network, then either – (i) the amount transferred is too small for the average ITE^1 to be positive and significant; or, (ii) recipient households do not themselves use such resources to increase secondary enrolment, but rather use them to change behavior along other margins such as consumption, say.[34]

At the foot of the table we report the triple differences, $\Delta TTE = TTE^1 - TTE^0$, and $\Delta ITE =$

[34] Two further points are of note. First, the estimated parameters of interest are very similar whether they are estimated from separate regressions for connected and isolated households, or from a pooled regression. This implies the marginal effects of the other controls do not differ between connected and isolated households. Second, the coefficient on L_{jh}, β_3, which recall is not used to estimate any of our parameters of interest, is not significantly different from zero. This suggests the effect of having an extended family on the level of enrolment in the absence of *Progresa*, is not changing over this time period.

$ITE^1 - ITE^0$. Column 2a shows that connected eligibles experience a 8.4% increase in their secondary enrolment rate relative to eligible but isolated households within treatment villages. Moreover, the ΔITE estimate implies non-eligibles with and without extended family present have almost identical changes in enrolment from the baseline period.[35]

Column 2b estimates (4.3) and therefore allows responses to *Progresa* to vary along a number of additional dimensions that have previously been documented in Table A3 to be correlated with the presence of extended family members in the same village. In addition, we control for potentially differential trends in enrolment among eligibles and non-eligibles by allowing the response to *Progresa* to vary with the baseline village enrolment rates among eligibles and non-eligibles. The result shows that the previous pattern of coefficients is robust to the inclusion of these interactions. In particular, connected eligibles have a – (i) 9.2% increase in their enrolment rate relative to connected eligibles in control villages; (ii) 8.6% increase in their enrolment rate relative to isolated eligibles in treatment villages. Reassuringly, the previous estimates of the response of households to the program were not merely picking up heterogenous responses along observable dimensions that are correlated with the existence of extended family links.[36]

While we can never rule out with certainty that there exists some unobserved characteristic that drives the results, the stability of the estimates across specifications (4.2) and (4.3) is reassuring, and the requirements for such a variable to explain the data are stringent. In particular the unobserved characteristic should – (i) be correlated to the presence of extended family in the village; (ii) drive responses to *Progresa*; (iii) not drive the baseline levels of enrolment to be different across connected and isolated households.

Estimating (4.3) allows us to benchmark the magnitude of the effect of being embedded in a family network vis-à-vis other observable characteristics. We find that – (i) the magnitude of TTE^1 is larger than the TTE of most other observable characteristics; (ii) there are few observable dimensions along which household behavior is as heterogenous as with regards to whether members of the extended family are present or not. For example, the TTE of the household owning land is

[35] As ΔITE corresponds to coefficient on the interaction term, $L_{jh}.D_h$, the results suggest there is no differential effect of having extended family by eligibility status. Reassuringly, this is in line with the presence of extended family not proxying for some unobservable that varies with eligibility status and drives enrolment over time.

[36] Three simple explanations of this result would be that – (i) isolated households have higher enrolment rates to begin with and so have less scope to respond to the program; (ii) trends in enrolment rates among isolated households are different to others households over this time period; (iii) isolated households are not aware of the program, while connected households may simply share information rather than resources within their family network. None of these explanations is supported by the data. Although isolated households do have slightly higher enrolment rates than households embedded in family networks at baseline, these enrolment rates are overtaken by November 1999 by the other households. Second, the pattern of coefficients on the $ITEs$ rule out the explanation that isolated households are naturally changing their enrolment rates over this time period. Third, villages in the evaluation sample are small – the average number of households in each is 45. It is hard to conceive of a significant proportion of households remaining unaware of such a large scale policy intervention.

.082, the TTE of the household not owning land is .065. While both these $TTEs$ are significant at the 1% level, the difference between them is not significantly different from zero.[37]

We next repeat the analysis for boy's and girl's secondary enrolment separately. Columns 3a and 4a estimate the basic specification (4.2) and show that, in line with the existing literature, household responses are greater for girls enrolment than for boys, with the magnitude of the point estimate of the TTE for girls being double that for boys. Columns 3b and 4b estimate specification (4.3) and show that, first, connected eligibles significantly increase the enrolment of boys and girls more than analogous households in control villages. Second, eligible but isolated households do not respond to the program in terms of boys enrolment, although the TTE^0 estimate is imprecisely estimated. As a result, the ΔTTE shows that the null hypothesis that within treated villages connected and isolated households have the same response cannot be rejected. Third, eligible but isolated households do not respond to the program in terms of girls enrolment, despite the monetary value of conditional cash transfers being higher for girls' enrolment, and the baseline level of girls's enrolment being lower.

Having established the importance of being embedded within a family network for a households response to *Progresa*, we now begin to unpack why this is the case. To begin with, a necessary condition behind why connected and isolated households respond differentially to the program, is that the level of transfers conditional on secondary enrolment is not sufficient to compensate households for the opportunity costs of child labor.

To provide direct evidence on this we use cross sectional variation in adult wage rates in October 1998 to identify villages with above and below the median level of adult wages. The results show that in villages where wages are sufficiently high – (i) in terms of boys secondary enrolment, neither connected nor isolated households respond to *Progresa* (Column 5a); (ii) in terms of girls secondary enrolment, it remains the case that only connected households respond to the program (Column 5b). If children are substitutes for adults in the local labor market so that their wages are positively correlated with adult wages, and if boys are more likely to be engaged in labor market activities relative to girls, then the results suggest the magnitude of transfers is not large enough to induce many households to change their behavior and increase the school enrolment of boys. Moreover, for connected households, there exist limits to the resources family networks can redistribute towards members with secondary school aged children.[38]

[37] The $TTEs$ of the head being above (below) the median age among all couple headed households is .090 (.062), the spouse being above (below) the median age is .070 (.082), the head being literate (illiterate) is .061 (.118), the spouse being literate (illiterate) is .061 (.099), and the household size at baseline being above (below) the median of all couple headed households is .041 (.096). These $TTEs$ are all significant at the 1% level. The difference in $TTEs$ are not significantly different from zero except along two margins – the head being literate or not ($\Delta TTE = .057$ and is significant at the 10% level), and household size at baseline being above or below the median ($\Delta TTE = -.054$ and is significant at the 5% level).

[38] Using wages as measured in October 1998 may be problematic if in villages where children withdraw from the

A second important factor underlying why connected and isolated households may respond differentially to the program is that extended families share resources. To explore this possibility we check whether family wealth is correlated to secondary enrolment rates at baseline for connected households. In particular we regress the share of children of secondary school aged children that are enrolled in school at baseline against the set of controls described above, and additionally control for the average poverty index – a measure of permanent income – among households in the extended family. The result in Column 6 shows that wealthier households have significantly higher enrolment rates at baseline. The magnitude of the coefficient implies that if there were to be a one standard deviation increase in the poverty index of the average family member, the households secondary enrolment rate would rise by 2%, relative to a baseline enrolment of 65%.

In the Appendix we present a series of robustness checks on our main result that only connected households respond to *Progresa*. The first series of checks relate to concerns over the surnames information and matching algorithm. In particular we show the robustness of the baseline results to – (i) potential measurement error in surnames; (ii) limiting the sample to smaller villages where there is less likelihood of spurious family links being defined; (iii) the fact that our matching algorithm may measure the intrinsic value of surnames rather than having anything inherently to do with extended family links. The second series of checks address – (i) the concern that there are unobserved village level characteristics that drive both the presence of isolated households and their differential response to *Progresa*; (ii) the concern that there may be unobserved time varying household characteristics that drive their program responses such as whether they are subject to economic shocks post-program; (iii) the underlying identifying assumption that there are no spillover effects from treatment to control villages; (iv) the underlying assumption that extended family networks are not changing over time.[39]

To summarize, the evidence suggests the behavioral response of households to *Progresa* differs according to whether members of their extended family are resident in the village or not. In particular – (i) only households embedded within family networks respond to the program in terms of their schooling outcomes; (ii) eligible but isolated households do not respond; (iii) connected households respond significantly more than isolated households in treatment villages; (iv) neither

labor market as a result of the program, are those that experience the greatest adult wage increases. However, October 1998 corresponds to only five months after the introduction of the program. Hence wages are unlikely to have adjusted to such an extent so as to change the cross sectional ranking of villages by adult wages. Moreover we note that the correlation over time in adult wages in the 212 villages in which wages are available both at baseline and in October 1998, is .749.

[39] An additional concern is that if children can flexibly supply any amount of hours to the labor market, then children may enrol in school 85% of their time – just sufficient to obtain *Progresa* transfers – and then devote the remaining 15% of their time to continue earning in the labor market. This hypothesis is not supported by the data. Attendance rates in November 1999 are over 95% for both boys and girls, suggesting, as is intuitive, there are indivisibilities or fixed costs in supplying labor to the market.

type of household responds in terms of boys enrolment in villages where adult wages are sufficiently high; (v) there are no indirect treatment effects of the program on secondary enrolment rates among non-eligibles. Given the final point, in the remainder of the paper we focus on the TTE parameter estimates and return to discuss the lack of $ITEs$ in the final section.

5.2 Understanding Why Families Matter

To begin to understand why connected and isolated households respond differently to *Progresa* we now exploit the design feature of the program that transfers provided conditional on primary enrolment act as a pure income effect on households, as documented in Table 2. This has two important consequences for changes in household behavior. The first is that households with and without primary school aged children should respond differentially to the program in terms of secondary enrolment, all else equal. The second is that if families share resources, then the wedge between the behavioral response of connected and isolated households depends on the presence of primary school aged children in eligible households within the extended family.

To provide direct evidence on the first mechanism we estimate how the average treatment effect varies with the value of transfers received by the household itself, and the presence of primary school aged children in the household. To provide evidence on the second mechanism we estimate how the average treatment effect varies with the value of transfers received by eligibles in the extended family of connected households, and the presence of primary school aged children in the family network. As transfers are provided conditional on household behavior, the actual transfers received are endogenously determined. We therefore proxy actual transfer receipts with the value of potential transfers the household could have received, as measured in October 1998. This is determined by the demographic characteristics of children in the household at baseline. Figure 2a plots the histogram of the potential transfers available to isolated and connected households. These distributions are similar because the demographic characteristics of children in isolated and connected households do not differ at baseline.

Figure 2b plots the potential transfers each eligible connected household with secondary school age children can receive against the average potential transfer eligible households in its family network are entitled to. The correlation between these two potential sources of income for connected households is only -.01. This is because, as documented in Table 4, there exists considerable variation in the demographic characteristics of children across households within the same family network. Hence the potential transfers available to a household and those in its extended family provide independent sources of variation from which to identify the effects of resources that flow into household h and those that flow into the households of its extended family. We use the following specification to estimate how the treatment effects of *Progresa* on secondary enrolment rates vary with the potential transfers received by household h itself (T_h^O), and with the average value of transfers received by eligibles among the extended family (T_h^F), for eligibles,

$$\Delta Y_{hvt} = \begin{array}{l} \alpha + \beta_1^O \left(\Delta P_{vt} \times T_h^O\right) + \beta_1^F \left(\Delta P_{vt} \times T_h^F\right) + \beta_2^O T_h^O + \beta_2^F T_h^F \\ +\boldsymbol{\lambda}'\mathbf{X}_{hv} + \sum_i \gamma_{2i} Z_i + \sum_i \gamma_{3i} \left(\Delta P_{vt} \times Z_i\right) + \Delta u_{hvt}, \end{array} \tag{5.1}$$

where all other controls are as previously defined, and standard errors are clustered by village. The parameter of interest is how the program response (TTE) varies with the intensity of treatment,

$$\frac{\partial \Delta Y_{hvt}}{\partial \Delta P_{vt}} = \beta_1^O T_h^O + \beta_1^F T_h^F + \sum_i \gamma_{3i} Z_i. \tag{5.2}$$

As potential transfers are defined for eligibles in both treatment and control villages, this is identified from a comparison of eligibles in treatment villages with a given value of potential transfers, relative to eligibles in control villages that have an identical value of potential transfers. Both sets of households also have the same family characteristics (connected or isolated). By controlling for the level of potential transfers the household and average family member are eligible for in (5.1), we capture any direct effects the demographic composition of the household and its extended family have on changes in secondary enrolment.[40]

Table 6 presents the results. The upper (lower) panel shows the results for households with (without) primary school aged children in them at baseline. This allows us to assess any interplay between the own and family transfers and whether the household is on the margin of enrolling children into secondary school because it receives unconditional transfers for primary school aged children. Across columns, we first compare the TTE arising from the potential transfer to isolated and connected households, and then among connected households, the later columns present estimates of the TTE when potential transfers to the household and eligible members of its extended family are both taken into account. We evaluate (5.2) at the mean values of T_h^O, T_h^F, and each Z_i.

Column 1 estimates the program response among isolated households. We see that the TTE is not significantly different from zero for isolated households, irrespective of whether they have any primary school aged children in them or not. Repeating the exercise for connected households in Column 2 shows that, on average, there is a positive and significant TTE for households both with and without primary school aged children.

[40] For example, families with a larger share of primary school aged children may be learning more quickly about the net benefits of secondary education for their older children, other things equal. The mean value of potential transfers eligible households with secondary school aged children receive in October 1998 is 1225 pesos, and the mean value of potential transfers received by households in their extended family – whatever the demographic composition of their children at baseline – is 541 pesos. This difference reflects the fact that the value of transfer increases for older children.

The remaining Columns shed light on how these $TTEs$ vary with the demographic composition of the family network that is eligible for *Progresa*. Column 3 focuses on families in which the minority of members are both eligible and have primary school aged children. In these networks, *Progresa* leads to a relatively small increase in the resources available per household. The upper panel shows that households with primary school aged children in such families significantly respond to the program. In contrast, the lower panel shows the behavior of connected households that receive no unconditional transfers themselves is not significantly influenced on the margin by the flow of resources into their household, or into the households of eligible family members.

Therefore, a subset of connected households who do not respond to the program are those that receive no unconditional transfers themselves, and are part of family networks that potentially receive only a small cash injection from *Progresa*. One explanation that reconciles this with resource sharing within families is that in such families, resources are first redistributed towards those with primary school children who are on the margin of being able to enrol their children into secondary school.

A useful thought experiment is to ask what would have been a household's response to the program if it were the *only* household in its family network to obtain transfers? To answer this, Column 3 reports (5.2) evaluated at $T_h^F = 0$. The estimate implies if a household is the only member of its extended family to obtain transfers, it does not significantly respond to the program in terms of secondary school enrolment. As shown in the upper panel, this is true even for those households who have primary school aged children and are closer to being able to fully offset the opportunity costs of secondary school enrolment. This pattern of behavior – that connected households do not respond to the program if others in their network are non-eligible – is also found throughout the next Section, where we consider how program responses vary with the presence and eligibility status of particular extended family members.

Column 4 focuses on the case in which the majority of family members are both eligible and have primary school aged children. In these networks, *Progresa* leads to a relatively large increase in resources available per household. In such families, connected households with and without primary school aged children are sensitive on the margin to the total inflow of resources to the family network. To see how responses to the program vary with the extent of the inflow of resources to the family network, Column 4 also reports the TTE where we evaluate (5.2) at the mean of T_h^O and the 75th percentile of T_h^F. The estimates imply that the response of households to *Progresa* in family networks that experience a large injection of transfers, is increasing in the magnitude of these transfers. Comparing the $TTEs$ in Column 4 we note that in such families, households without primary school aged children are more sensitive on the margin to the flow of resources into the family network than are households with primary school aged children.

Comparing the pattern of $TTEs$ in Columns 3 and 4 reveals a non-linear relationship between the resources received by the average family member and household secondary school responses, and that this non-linear relationship differs for households with and without primary school aged children. For both types, if they are the sole recipient of transfers they have a response close to zero (Column 3). For those with primary school aged children, their response is more sensitive to the inflow of resources into the extended family when these amounts are small (Column 3, upper panel), and they are less sensitive to them when the magnitude of inflows increases (Column 4, upper panel). In contrast, those without primary school aged children only become sensitive to the inflow of resources to their family when the magnitude of this inflow is sufficiently large (Column 4, lower panel). This pattern of results is consistent with families first redistributing resources towards extended family members with primary school aged children, namely those that receive unconditional transfers and are on the margin of enrolling children into secondary school.

6 Which Family Members Matter and Why?

We now investigate whether and why particular members of the extended family influence the behavior of households within the network. Exploiting the full richness of the constructed data on extended family links helps provide further credibility to the earlier results suggesting the behavioral response of connected households to *Progresa* differs from that of isolated households because families share resources among their members. Furthermore, distinguishing the influence of particular family members is of intrinsic importance for the following reasons.

First, evolutionary biology predicts social concerns may differ both between men and women towards their family, and between older and younger generations of the family towards each other [Hamilton 1964, Trivers 1972]. Second, among siblings, there may be peer pressure, learning, conformity, cooperation, or complementarities in the schooling choices across households. Third, the pre-program demand and supply of private transfers across different pairs of households within the family network will vary. Whatever are these precise *ex ante* patterns of monetary and non-monetary transfers, each may be crowded out or crowded in by *Progresa* transfers. Finally, if some family members are more pivotal for the behavior of others, then this may shed light on which households in the network should optimally be targeted in other policy interventions.[41]

We focus on connected households throughout and for any given family link type-j, we first

[41] Strauss and Thomas [1995] review the empirical evidence on whether mothers behave more altruistically towards their children than fathers. Relatedly, Duflo [2003] presents evidence consistent with maternal and paternal grandmothers having different degrees of altruism towards their grandchildren. Albarran and Attanasio [2003] present evidence on the crowding out of private transfers by *Progresa* transfers, and how this varies with the variance and persistence of income. Cox and Fafchamps [2007] review the literature on the crowding out of within family transfers by publicly provided transfers.

estimate specification (4.2) only for the subset of households with that link in the village, and shed light on whether the treatment effect of *Progresa* varies with specific characteristics of the link. Two characteristics in particular are of relevance for our analysis. First, if families share resources, the response of any household should depend on the eligibility status of their family links. Second, given the design of *Progresa*, the demographic composition of children in linked households determines the extent to which they receive largely unconditional transfers for the primary school enrolment of their children. This in turn should influence responses if these unconditional transfers are redistributed to others in the family.[42]

6.1 The Eligibility Status of Siblings

Table 7 presents estimates of how the treatment effects of *Progresa* vary with the presence and eligibility status of the same gender siblings of the head and spouse of household h – namely the brother of the head of household, and the sister of his spouse.[43] Column 1a shows that eligible households in treatment villages that have an eligible brother of the head present, namely the uncle of the children in household h, increase secondary enrolment rates significantly more than analogous eligible households with eligible brothers in control villages ($TTE^1 > 0$). In contrast, there is no program response among eligible households with non-eligible brothers ($TTE^0 = 0$). Column 1b shows a similar pattern of responses among households by the eligibility status of the sister of the spouse, namely the aunt of the children in household h. The fact that households respond only if both they and their sibling are eligible for *Progresa* transfers is consistent with there being few resources to redistribute within the family when other members of the extended family are non-eligible for the program.

A concern with this result is that it merely captures unobservables that both drive the response to the program and that cause the eligibility status of siblings to be similar. We address this in two ways. We first estimate (4.3) and so allow household responses to be heterogeneous along a number of dimensions in addition to the eligibility status of their siblings. The results, reported in Columns 2a and 2b, show that it remains the case that households do not respond to *Progresa* if they are eligible for transfers but the siblings of the head and spouse are not. A second approach

[42] For example, to understand whether the program response of household h varies with the eligibility status of the brother of the head of household, we define our link variable, L_{jh}, to be equal to one if the head of household h has brothers present in the village and at least one brother is himself eligible for *Progresa* transfers, and zero if brothers are present in the village but all are non-eligible. The average couple headed household with secondary school aged children has 1.24 brothers of the head and .657 sisters of his spouse present in the village.

[43] We focus on same gender siblings because, first, these household pairs are likely to be at similar stages of the life cycle and so more similar on observables than cross-gender sibling pairs still resident in the village. In particular secondary and primary school aged children are likely to be observed in both households. Second, as shown in Table 3, the proportion of households with same gender siblings is the same across treatment and control villages by eligibility – the same is not true for cross-gender sibling pairs.

is to allow program responses to vary with characteristics that reflect heterogeneity within the extended family. A natural candidate to capture this is the standard deviation of the household poverty index among extended family members. The results, reported in Columns 3a and 3b, continue to show that, allowing for household responses to vary with the degree of inequality among the family network, households only significantly increase secondary enrolment when both they and their siblings are eligible for the receipt of transfers.

Another concern is that the results do not distinguish whether household h's behavior is driven by the eligibility status of their true siblings, or whether it is driven by the eligibility status of those households that are similar on observables to the true siblings of household h, such as members of their network of friends. To shed light on this we randomly reassign each household the family links of another household in the village. This method leaves the likelihood of any household being connected to a family link type-j in a village to be the same as in the original data.[44]

The results, in Columns 4a and 4b, show that households are insensitive to the eligibility status of such randomly assigned siblings. Households in treatment villages significantly increase their secondary enrolment vis-à-vis those in control villages, in the presence of both eligible and non-eligible randomly assigned siblings ($TTE^1 > 0$, $TTE^0 > 0$).

6.2 The Demographic Composition of Siblings' Children

The next table explores whether for a given eligibility status of the sibling, the demographic composition of children in the sibling's household affects the response of household h to the program. This sheds light both on whether there is a redistribution of unconditionally received transfers from households with primary school aged children, and whether there exist complementarities in the secondary schooling choices across siblings. Such complementarities may arise for example from the fixed costs of school attendance such as transportation and books.

We therefore define our link variable, L_{jh}, to be equal to one if the head of household h has an eligible brother that has secondary school aged children, and equal to zero if the head of household h has eligible brothers that only have primary school aged children. $TTE^{L_j 1}$ is then the average treatment effect on households with eligible siblings where the siblings have secondary school age children, and $TTE^{L_j 0}$ is the average treatment effect on households with eligible siblings where the siblings have only primary school age children.

[44] As a result, households and their randomly assigned siblings are likely to be similar to each other on observables. This may approximate the friendship ties of the household because of the homophyly principle, namely that similarity between individuals on their socioeconomic and behavioral characteristics leads them to be more likely to form social ties with each other. Homophyly has been well documented to be a major driving force in the formation of social ties in a wide range of contexts including friendship, marriage, work advice, information transfer, exchange, and co-membership of organizations [McPherson *et al* 2001].

Column 1a of Table 8 shows households with eligible brothers of the head present significantly increase secondary enrolment relative to analogous households in control villages. The magnitude of the response is more than twice as large if siblings receive largely unconditional transfers through only having primary school aged children. Column 1b shows a similar pattern of results when we consider how the response of households varies with the demographic composition of children in the eligible household of the sister of the spouse.

One concern is that households whose siblings have younger children may differ on unobservables to those whose siblings have older children, and such unobservables drive program responses. This concerned is ruled out by the results in Columns 2a and 2b. These show that conditional on the sibling being non-eligible, the demographic composition of the children of siblings does not influence the response of household h. It is not therefore the case that households that have siblings with young children respond differently to the program *per se.* Rather it is the interaction of those siblings being eligible for *Progresa*, and the siblings receiving unconditional transfers for the primary school enrolment of their children, that allows the household to itself respond to the program. This is consistent with the transfer of resources from eligible households with primary school aged children to the households of their eligible siblings with secondary school aged children. It is precisely such households that are likely to be on the margin of enrolling their children into secondary school.

The pattern of coefficients in the first four columns also show that the presence of siblings with secondary school aged children only influences program responses if those siblings are themselves eligible for *Progresa*. This rules out that idea that households share the direct costs of secondary school with all of their siblings. Moreover this pattern of results is inconsistent with behavior being driven by conformity or peer effects within siblings based on secondary school outcomes, because if so, the presence of secondary school aged children in siblings' households should influence behavior in household h independent of the eligibility status of siblings.

The remaining columns repeat the analysis with each household h being randomly assigned a same gender sibling from within the village. The results show that – (i) households significantly respond to the program in the presence of randomly assigned siblings that have secondary school aged children; (ii) the presence of primary school aged children in the household of the randomly assigned sibling has no effect on the response to the program of household h; (iii) household responses are independent of the eligibility status of the randomly assigned sibling.

These results suggest that household behavior is influence both by the presence and characteristics of their actual siblings, and the presence and characteristics of households that are similar on observables to their true siblings. However the mechanism through which household behavior is influenced by these two groups is different. Among actual siblings, the results imply resources are transferred from eligibles with primary school aged children to eligibles with secondary school aged children. In contrast, with regards to randomly assigned siblings, household behavior is shaped by the mere presence of secondary school aged children – but not primary school aged children, and this effect is independent of the eligibility status of such assigned siblings. This suggests such pairs of households either share the direct costs of secondary schooling with each other, or that there are peer influences or conformity among them with regards to secondary school outcomes.

6.3 Adult Sons and Daughters

We now consider how household behavior is shaped by the characteristics of those in the family network that are inter-generationally linked. We first consider the links between the head and spouse of household h to their adult sons or daughters that head their own household j in the village. These individuals correspond to the adult siblings of the children actually resident in household h. Given the age structure of these links, we consider how the program response of household h in terms of secondary school enrolment rates varies with the eligibility status and the presence of young children in household j. The presence of young children matters if the resources available to be shared are those transfers received on a largely unconditional basis, such as those for primary school enrolment, or the attendance of children aged under five to local health clinics.

Table 9 presents the results. As with the effects of sibling links, Column 1a shows that households respond to *Progresa* if their adult son is also eligible, but do not if the household is in receipt of transfers but their adult son is not ($TTE^1 > 0$, $TTE^0 = 0$). Column 1b confirms this finding applies equally to adult daughters resident in the village.[45]

The magnitude of TTE^1 is larger than the baseline estimate and the corresponding treatment effect found in relation to the eligibility status of siblings. For example, eligible households with eligible adult sons in a treated village increase secondary school enrolment rates by 20.6% more than corresponding households with such links in control villages. This may reflect the fact that eligible adult children give part of the transfers they receive to their parents, thus enabling their younger siblings to enrol into secondary school. The adult son is himself unlikely to have secondary school aged children and so the transfers he receives relate either to his primary school aged children, or the attendance of his young children or pregnant wife to health clinics. As such, these transfers are likely to be viewed as a pure income effect to his household.

To shed light on this we next estimate how the treatment effects vary in household h with the demographic composition of children in the eligible households of their adult sons and daughters. Given the age profile of such households, we consider the influence of eligible adult child households

[45] If households are randomly assigned the adult sons and daughters of another household in the village, the average treatment effects are insensitive to the eligibility status of the randomly assigned family link.

that have primary school aged children versus those that have no primary school aged children and are thus eligible either because of the presence of very young children or a pregnant wife. The results in Columns 2a and 2b show the response of households, in terms of secondary school enrolment, increases conditional on their adult sons and daughters being eligible, and irrespective of the demographic composition of children in their adult son and daughter's households.

This suggests adult sons and daughters view both components – transfers conditional on the primary school enrolment of their children and those conditional on their attendance to health clinics – as being pure income effects for the household. The pattern of coefficients is consistent with there being a transfer of resources from eligible households with primary school aged or very young children, to the households of their eligible parents. This allows receiving households to fully offset the opportunity costs of secondary school attendance, and so enables the younger siblings of adult sons and daughters to enrol into secondary school.

6.4 Paternal and Maternal Grandparents

Finally, we consider inter-generational links from the head and spouse to their parents – namely, the paternal and maternal grandparents of the children in household h. Such elderly households can be eligible for a health component of the program where those aged over 60 are required to attend a clinic once per year. The results in Columns 3a and 3b show the response of households to the program are generally weaker in the presence of the paternal or maternal grandparents of the children. Unlike the other family links considered, this is the case irrespective of the eligibility status of the household in which the paternal or maternal grandparents reside. This may reflect the low monetary value of transfers such elderly households are eligible for [Skoufias 2005].[46]

7 Discussion

We have presented evidence from the *Progresa* social assistance program on whether and how household behavior is influenced by the presence and characteristics of its extended family. Our central finding is that *Progresa* raises secondary enrolment only among households that are embedded in extended family networks. Eligible but isolated households do not respond. Our results suggest a key channel through which the extended family influences household schooling choices relates to the redistribution of resources among family members to enable eligible family members

[46] The evaluation data contains limited information on pre-program transfers. We note that those households defined to be of older generations are more likely to report receiving transfers at baseline. This is consistent with evidence from the Mexican Family Life Survey that transfers tend to flow from younger generations to older generations in the family network.

to fully overcome the opportunity costs of enrolling their children into secondary school.

There are a number of common themes running throughout our analysis that we now bring to the fore. The first stems from the design feature of *Progresa* that transfers are provided directly to women. If this increases females' bargaining power, household behavior may be differentially influenced by the characteristics of the extended family of the wife relative to that of her husband. However, throughout the paper we find no evidence that the family links of the spouse are more influential than are the family links of her husband. The mechanism through which extended family affects household schooling behavior – resource sharing – is qualitatively the same for both the head and spouse's family links.[47]

A second common theme is that there is little evidence of any indirect treatment effect of *Progresa*, namely that secondary enrolment rates among non-eligibles in the family network do not increase on average. This suggests either – (i) resources are not transferred from eligibles to non-eligibles in the network; (ii) the amounts transferred are small; (iii) non-eligibles do not themselves use these resources to increase secondary enrolment.[48]

To shed light on this, we consider program responses along another margin, consumption. In particular we estimate the average treatment and indirect treatment effects of *Progresa* on household consumption, and explore how these vary with whether the household is connected or isolated. Our sample covers all couple headed households, not just those with secondary school aged children, and we use two measures of consumption – (i) the change in weekly consumption of meat, a luxury good; (ii) the proportionate change in household non-food expenditures.[49]

---

[47] This result does not however imply that female bargaining power or the distribution of resources within the household is unaltered by *Progresa* [Attanasio and Lechene 2002].

[48] Note that we do not find robust evidence of an effect of the program on the school enrolment of the *average* non-eligible. Spillover effects in schooling have been reported by Bobonis and Finan [2006] and Cattaneo and Lalive [2006]. These alternative findings may stem from methodological differences. For example, Bobonis and Finan [2006] and Cattaneo and Lalive [2006] match children over the waves of the data and consider changes in each individual's enrolment status. Cattaneo and Lalive [2006] focus on the effects of the program in its first year up to October 1998 (wave 3), and consider the enrolment among primary school aged children and those transiting from primary to secondary school. Bobonis and Finan [2006] consider changes in enrolment of those children that have completed primary school at baseline, and find such spillover effects predominantly among those in non-eligible households on the margin of being eligible. Taken together, these results suggest there may be heterogeneous spillover effects on non-eligibles, and heterogeneous effects on the children within non-eligibles. Whether extended families influence the nature of such spillovers remains a question for future research.

[49] Consumption data is not available at baseline so we consider changes in these measures from the second pre-program wave of data (March 1998) until November 1999. For meat consumption, households were asked the number of times they consumed meat in the week prior to survey day. For non-food expenditures we normalize changes over time by their initial values in March 1998. We remove the top and bottom 2.5% of values in the proportionate change in non-food expenditures so that the results are not driven by outliers. Although there may be seasonal changes in consumption, these should affect isolated and connected households equally. Moreover, the qualitative results are similar when changes between March 1998 and March 1999 are considered instead. Non-food expenditures include those on clothing, adult goods such as alcohol and tobacco, hygiene products, medical products, kitchen utensils, and children's toys. We do not include school related expenditures in this measure.

The results reported in Table 10 show that eligibles – whether connected or isolated – significantly increase consumption as defined by both measures. Reassuringly, this confirms that isolated households are not dysfunctional – they are both aware of the program and take-up *Progresa* transfers. Isolated households respond on this margin because the mere receipt of transfers allows consumption to increase. In contrast, the value of transfers given to any one household for the secondary school enrolment of their children, are not by themselves sufficient to offset the opportunity costs of sending children to school.

Furthermore, there are indirect treatment effects on meat consumption for both connected and isolated households. The fact that the ITE's are both positive suggests there might be – (i) redistribution of resources within families from eligibles to non-eligibles; (ii) redistribution of resources to isolated households; or, (iii) village-wide effects of the program that alter the consumption bundles of all households in treated villages. The fact that there are no $ITEs$ for non-food consumption may reflect that on the margin, households prefer to use additional resources for the consumption of luxury food rather than non-food items.[50]

In ongoing work, we develop and test a formal model of risk sharing in villages, that reflects the fact that while all households can potentially insure each other, the transactions costs of transferring resources within and outside family networks may differ. This has important implications for how consumption patterns in eligible and non-eligible connected and isolated households change in response to an exogenous income shock such as the introduction of *Progresa*.

The final theme running throughout our analysis has been the interplay between the design of conditional cash transfer programs, the presence of extended family members, and household responses to the program. There are three policy implications of our findings for the design of such programs. First, if families share resources, how connected households respond to such policies on any given margin, will generally depend on the eligibility status of others in their network. The particular design features of *Progresa* also lead to the behavior of connected households to depend on the demographic composition of others in the family network. More generally, our findings highlight that ignoring the presence and characteristics of the extended family can lead to an incomplete understanding of the forces driving the behavioral responses of households to large scale policy interventions in developing country settings.

All non-food expenditures are valued in pesos. Angelucci and De Giorgi [2006] and Gertler *et al* [2006] provide evidence that local food prices do not change significantly over time between treatment and control villages.

[50] These results are in line with those of Angelucci and De Giorgi [2006] who find there to be indirect effects on the consumption of non-eligibles in treatment villages. They present evidence that this occurs through the insurance and credit markets – households indirectly benefit from their neighbors higher income by receiving more transfers, by borrowing more when hit by a negative idiosyncratic shock, and by reducing their precautionary savings. They find a limited impact of the program on food prices directly, a result confirmed in Gertler *et al* [2006], suggesting the $ITE^0 > 0$ result in Table 10 reflects the redistribution of resources through insurance networks in the village rather than general equilibrium effects of *Progresa*.

Second, in this setting the value of transfers conditional on secondary school enrolment is not sufficiently high to offset any potential loss in children's labor market earnings if they were to enrol into full time secondary education. Our results suggest that the resources spent on *Progresa* could therefore be more efficiently targeted if the policy aim is to increase secondary school enrolment. In particular, if the program's entire budget were to be channelled into transfers conditional on secondary school enrolment – with no component conditioned on primary enrolment, and such that the value of transfers offset children's labor market earnings – then we would expect *both* isolated and connected households to increase their secondary enrolment rates.[51]

Finally, our analysis has general implications for understanding the role of family networks in developing economies. Using data from over 22,000 households in 506 villages in rural Mexico, we have documented that 20% of couple headed households are isolated in that none of their extended family members are geographically proximate in the same village. The incidence of this type of isolatedness is therefore at least as high as the incidence of single headedness, which affects 15% of households in our data. While there exists a large literature on the effects of single headedness on household welfare, our data and results suggest the importance of designing future surveys to identify isolated households in other settings, and more generally, to establish the social ties between households in survey data. Such information can then be used to understand the effects of isolation and being part of geographically proximate kinship networks on household behavior and welfare. This forms the basis of a broad and challenging research agenda.[52]

8 Appendix

8.1 Descriptive Evidence: Surnames

Table A1 provides descriptive evidence on each surname type – the paternal and maternal surnames of the head ($F1$, $f1$) and spouse ($F2$, $f2$). For both head and spouse, there are fewer paternal than maternal surnames reported. As Figure 1 shows, this reflects the fact that under a patronymic naming convention, paternal surnames have a greater survival rate across generations. There

[51] Todd and Wolpin [2006] develop and test a structural dynamic model of the schooling and fertility effects of *Progresa*. They find the policy as currently designed increases years of schooling on average by .5 years. They then estimate the effect of a revenue neutral reform that eliminates transfers to children in grades 3 to 5 and increasing transfers to those in grades 6 to 9 by around 50%. They find such a reform would increase average completed schooling by an additional .1 years. Attanasio *et al* [2005] conduct a similar experiment and also find larger effects on secondary enrolment of such revenue neutral policy alternatives.

[52] Several recent contributions have collected detailed information on the structure of family and other networks in village economy settings. These studies have, for example, provided important insights on the role of social networks in learning about agriculture in Ghana [Conley and Udry 2005], child fostering in Burkina Faso [Akresh 2005] and implicit insurance arrangements in Ethiopia [Dercon *et al* 2005].

are 1696 different paternal surnames reported by heads ($F1$), lower than for the other types of surname including those reported as the spouse's paternal surname ($F2$). This is both because the patronymic naming convention implies spouse's paternal surnames have lower survival rates across generations than those of male heads of household, and also be partly due to spouses moving into the 506 villages in the data from villages outside the evaluation sample.

The second row shows that the majority of surnames are mentioned at least twice. For each surname type, the most frequent surname covers around 9% of households, and the half the households have one of the 50 most frequent surnames for each surname type. The third row shows the probability of two randomly matched households having the same surname type is close to zero, and the expected number of households with the same head's paternal surname is 13.3. This is higher than the expected number of households with the same spouse's paternal surname, again suggestive of women moving into *Progresa* villages from other locations.[53]

The next two rows report the same information but at the village level. The probability (without replacement) of two randomly chosen households in the village having the same surname is orders of magnitude larger than in the population. Hence households are not randomly allocated by surname type into villages. On the other hand, the fact that the expected number of households in the village with the same surname is smaller than in the population implies households do not perfectly sort into villages by surname either.[54]

The final row sheds light on the degree of sorting of households into villages by surname type. This is measured by an odds ratio, defined as the ratio of the probability that two randomly chosen households from the same village have the same surname, divided by the probability that two randomly chosen households from the population of *Progresa* villages have the same surname. This odds ratio suggests that households are, for example, 356 times more likely to match within a village on their head's paternal surname than if they were randomly allocated by this surname across villages.[55]

[53] These population values are calculated as follows for any given surname type. Let n_i denote the number of households with surname i and let N denote the number of households that report some surname of the given type. The probability, without replacement, that two randomly chosen households have surname i is then $P_i = \left(\frac{n_i}{N}\right) . \left(\frac{n_i-1}{N-1}\right)$, and the expected number of households in the population with name i is $E_i = n_i . \left(\frac{N-1}{N}\right)$. The values reported in Table A1 are the averages of P_i and E_i over all surnames i.

[54] These village values are calculated as follows for any given surname type. Let n_{iv} denote the number of households with surname i in village v and n_v denote the number of households that report some surname of the type in village v. The probability, without replacement, that two randomly chosen households in the village have surname i is then $p_{iv} = \left(\frac{n_{iv}}{n_v}\right) . \left(\frac{n_{iv}-1}{n_v-1}\right)$, and the expected number of households in the village with name i is $e_{iv} = n_{iv} . \left(\frac{n_v-1}{n_v}\right)$. The values reported in Table A1 are the weighted averages of p_{iv} and e_{iv} over all villages v, where the weights are $\frac{n_{iv}}{n_v}$. These weights account for the same name being reported to different extents across villages. The expected number of matches in the village is based on only one surname, and so provides an upper bound on the total number of extended family links our matching algorithm actually defines.

[55] This odds ratio is calculated as follows for any given surname type. We first take the weighted average of

8.2 External Validity of the Extended Family Links: MxFLS Data

To provide external validity to the constructed data on extended family links in the *Progresa* data, we present similar information from an alternative data set that was collected in a comparable economic environment and time period. The *Mexican Family Life Survey* (MxFLS), collected in 2001, provides information on the number of each type of link, by head and spouse, that are still alive in *any* location, not just the same village. This data set therefore provides an upper bound on what should be recorded as family links in the *Progresa* data, in which we only construct links in the same village. In addition, we exploit information from the household roster in the MxFLS to also construct the number of family links inside the household, by each type of family link, and for the head of household and his spouse separately. To make the MxFLS data comparable, we restrict the sample to couple headed households that reside in locations with less than 2500 inhabitants in states that are also covered in the *Progresa* data. There are 580 such households.[56]

Table A2 reports the findings from the MxFLS. The number of family links to parents, children and siblings outside the household and located anywhere, are greater than those we construct using surnames data within the village from the *Progresa* data. The fact that more parents of the spouse are alive is likely to be driven by spouses being younger than their husbands. Moreover, the differences between husbands and spouses in the number of parents and siblings are less dramatic in the MxFLS data, presumably because these statistics refer to family links in any location and so are unaffected by the geographic mobility of women at the time of marriage.

The comparison of family links within the household is also informative. Here the number of each type of family is similar to that found in *Progresa*, although the number of children is slightly lower. This may be driven by differences in the age of respondents in the two data sets – the age of spouses is 40.5 (43.5) in the *Progresa* (MxFLS) data. Heads and spouses are also more educated in the MxFLS data – the mean years of schooling for heads (spouses) in MxFLS is 3.91 (3.46) in comparison to 2.77 (2.27) in *Progresa*. These differences would explain the lower numbers of children in the household in the MxFLS data if more educated couples have lower fertility rates.[57] Moreover, it remains the case that in the MxFLS data as in *Progresa*, the number of family links of the head inside the household are greater than those of the spouse.

p_{iv} over all names i where the weights are $\frac{n_v}{N}$. These weights take account of the fact that if two households are drawn from the population at random, they are more likely to come from a larger village. Denote this weighted probability as $\widetilde{p}_i$. The reported odds ratio is then given by $\frac{\widetilde{p}_i}{P_i}$.

[56] As discussed in Section 3, one restriction on the matching algorithm used in the *Progresa* data is that we are unable to identify links to parental households if only one of the parents is alive. To ensure the MxFLS data is therefore comparable, we do not include information from couple headed households that report only having a single parent alive in another household. There are no such concerns for parental links defined inside the household.

[57] This underestimates the true difference in average years of education of couples between the two data sets because in the MxFLS, years of schooling are top coded at 12.

8.3 Correlates of Extended Family Links

Before exploiting information of extended family networks to explain household behavior, a useful stepping stone is to first establish the correlates of the extended family being present in the same village. This focuses attention on the econometric concerns stemming from the endogenous formation of family networks. Moreover, this analysis also provides further supportive evidence on the accuracy of the constructed extended family links.

A number of mechanisms drive the presence of extended family members, such as the need for insurance and the choice of marriage partners for children [Rosenzweig and Stark 1989], the value of services provided by social networks [Munshi and Rosenzweig 2005], inheritance of land and other household assets [Foster 1993], and the nature of household production [Foster and Rosenzweig 2002]. Our aim here is not to replicate such analyses, but to identify correlations between the presence of extended family ties and three classes of observable characteristics.

First, the age of the head and spouse should be negatively correlated with the likelihood their parents are in close proximity, and positively correlated with the probability of having adult children in the village. These correlations are somewhat mechanical as they depend primarily on the life cycle rather than on economic mechanisms. Second, there can be a positive correlation between wealth or land ownership and the presence of an extended family because – (i) wealthier family dynasties may have higher fertility and lower mortality rates; (ii) landed households both have more need for and can support larger family sizes; (iii) wealthier families may also be more likely to own land – as rural land markets are typically missing, the ability to inherit land, or to acquire land specific human capital, may lead adult children to be more likely to remain within the village than otherwise. A third mechanism driving extended family structures is the need to insure against risk. This leads to the formation of networks of related families with negatively correlated shocks, the strategic marriage of daughters into families with less correlated shocks, and migration of some family members to other locations.

To shed light on these channels we estimate a conditional logit regression where the dependent variable, L_{jh}, is a dummy equal to one if extended family link type-j exists for household h in the village, and zero otherwise. We consider the correlates of each of the specific family links that we identify, as well as on whether household h has *any* family connections in the village. For each link type, L_{jh}, we control for the ages of the head and spouse, and dummy variables for whether they are literate and speak an indigenous language. At the household level, we control for whether the household owns its home, whether any land is owned, whether any member of the household temporarily migrated in the last year, the eligibility status of the household, the household poverty index, and household size at baseline.[58]

[58] We experimented with other specifications before settling on this set of controls. For example, we do not control

We group the conditional logit regression by village to take account of differences across villages that drive the formation of extended family networks. For example, villages may vary in the riskiness of their economic environments, altering the need for households to insure each other and therefore potentially causing alternative patterns of extended family networks to form. Standard errors are clustered by village, we report log odds ratios so that tests of significance relate to the log odds being significantly different from one. All continuous variables are divided by their standard deviation so the corresponding coefficients can be interpreted as the effect of a one standard deviation change in the continuous variable.

The results, reported in Table A3, highlight the following. First, the mechanical correlations with age are as expected with older heads and spouses being significantly less likely to have their parents outside the household and resident in the village, and significantly more likely to have their adult children in other households in the village. Older heads and spouses are more likely to have brothers present and less likely to have sisters present, presumably because, as highlighted in Section 3, women move village at the time of marriage. Second, literate heads and spouses are more likely to have their parents present. If such correlations persist across generations, then parents that educate their children increase the likelihood their children remain geographically proximate, other things equal.

Third, home and land ownership are positively correlated with the likelihood that children and siblings reside in the same village, other things equal. The coefficients are of similar magnitude for brothers and sisters as well as for the adult children of the head and spouse (not shown). This pattern of coefficients is consistent both with inheritance norms in rural Mexico that do not favor men over women, and with a dynastic wealth effect such that wealthier families accumulate greater assets and have higher rates of fertility. In contrast, households in which at least one member has temporarily migrated in the last year – 18.5% of all households – are not more or less likely to have extended family links present.

Fourth, although there is a slight positive correlation between the household poverty index and the presence of adult children, there is no discontinuous effect of eligibility on the presence of any extended family ties. This aids the interpretation of the econometric evidence we provide. Whether the head and spouse speak an indigenous language also does not predict the presence of extended family ties. This is again reassuring because the unconditional number of extended family ties, for each type of tie, are no different between indigenous and non indigenous households.

Fifth, households that have a greater number of individuals within them are also significantly more likely to have a greater number of extended family members residing within the same village.

for years of education because it is highly correlated with literacy – 89% (90%) of heads (spouses) have no formal schooling if they are illiterate. We focus on temporary rather than permanent migration because the proportion of households that report any members permanently migrating in the five years prior to 1997 is only 3.3%.

This may be due to persistent differences in fertility levels within the same family dynasty across generations. Alternatively, the presence of extended family members may reduce the costs of having and raising children because extended family members are able to supply of time, labor, and other resources to the household.

A comparison across the columns is also informative. For example, the controls have a differential effect on the likelihood that the parents of the head or spouse are present. This is in line with the earlier evidence suggesting the process that drives the presence of parents are very different for the head and his spouse. In contrast, most of the controls have similar effects on the likelihood of brothers or sisters of the head and spouse being present.

In summary, the final column shows that connected and isolated households differ on a range of observable characteristics that drive the presence of extended family. In the empirical analysis it will therefore be important to both condition on these observables, and to allow household responses to *Progresa* to also vary with them. Hence the analysis sheds light on whether there exists a differential effect of being embedded within a family network or not, over and above potentially heterogeneous effects of characteristics that predict the existence of family links.[59]

8.4 Robustness Checks on the Baseline Estimates

We present a series of robustness checks on the main finding in Table 5 that only households embedded within extended family networks respond to *Progresa*. The first series of checks, presented in Table A4, relate to concerns over the information on surnames and the matching algorithm.

We first address concerns over potential sources of measurement error in the recorded surnames data that can lead to erroneous inference on the presence of extended family members. Column 1 shows our baseline result to be robust to dropping households with any of three potential types of measurement error in their surnames – (i) the spouse's maternal surname is the same as their husband's paternal surname; (ii) the spouse's paternal and maternal surnames are the same; (iii) the paternal and maternal surnames of both the head and spouse are all reported to be identical. The point estimate of the TTE^1 is larger than the baseline estimate, suggesting the previous estimates may have indeed been subject to attenuation bias.[60]

[59] As a further check on whether connected and isolated households differ in their overall distribution of observables, we estimated a propensity score for each type of household by eligibility status. The propensity score is based on the same set of covariates as in Table A3 including village fixed effects, but we control for a richer set of household demographic composition variables rather than just household size at baseline. Figure A1 shows the distribution of the propensity score by extended family links and eligibility status. The distribution of the score for connected (isolated) households is shown on the positive (negative) section of the y-axis in red (blue). This analysis reveals that connected and isolated households appear overall to be balanced in household specific observables.

[60] The results are also robust to dropping households with the most common paternal surname of either the head or spouse in the village.

The second check addresses the concern that in larger villages, the matching algorithm is more likely to spuriously link two households that happen to have the same paternal and maternal surnames of head and spouse but are genuinely unrelated. Column 2 shows the baseline results to be robust to dropping villages in the top quartile of the village size distribution, as measured by the number of households in the village. Moreover the point estimate on the TTE^1 is larger than the baseline estimate. This may again reflect a downward bias in TTE^1 because isolated households we previously being ascribed to be connected.

A third concern with the matching algorithm is that it actually measures the intrinsic value of surnames rather than having anything inherently to do with extended family links. For example, individuals with the most frequent surnames are both most likely to be found to have extended family members present, and may belong to family dynasties that have different unobservables that cause them to respond differently to *Progresa* than households with less common surnames. We address this concern in two ways.

First, we randomly reassign households in our baseline sample to another village within the same municipality and then rerun our matching algorithm based on the surname matches in these neighboring villages. We then explore whether our main results capture the effects of true extended family links that are present in the same village, or merely capture the effects of having more or less frequent surnames *per se*. The result reported in Column 3 shows there are no heterogeneous responses to *Progresa* on the basis of these surname based links in neighboring villages – the difference in the $TTEs$ is not significantly different from zero.[61]

Second, we estimate whether the frequency of paternal surnames predicts fertility levels at baseline. If for example some surnames are more frequent because those dynasties have lower mortality rates or are better able to insure against income shocks, then we expect them to have higher fertility, other things equal. We estimate an OLS specification analogous to (4.1) where the left hand side variable is the number of children aged 0 to 16 in the household at baseline, and we control for the share of households with the same paternal surname of the head and spouse, as well as the basic set of controls previously discussed. The result in Column 4a shows that the relative frequency of paternal surnames is uncorrelated to fertility levels at baseline. This result is robust to focusing attention to only the most common paternal surnames, namely those that

[61] Three points are of note. First, we only reassign the subset of 6227 households in the baseline sample. If we were to reassign all 22,553 households then the newly constructed family ties would be more likely to capture the actual family ties originally used. Second, we reassign households to other villages within the same municipality because as discussed in relation to Table A1, households sort across geographic locations by surnames. Doing so, 59% of households are found to be connected. Randomly reassigning households to any other village in the data would however dramatically reduce the likelihood any household is constructed to have family links. Finally, there are 115 municipalities in the data, the median municipality contains six villages, and we drop municipalities that only contain one village.

are shared by at least .5% of the population (Column 4b).

The second series of checks, presented in Table A5, relate to a number of remaining concerns. First, there may be unobserved village level characteristics that drive both the presence of isolated households and their differential response to *Progresa*. We address this concern in two ways. First, we estimate (4.3) where we add in an additional series of village level characteristics in Z_i. These village level controls are the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, and the share of households that report being affected by any natural shocks from October 1998 to November 1999. These shocks include being affected by droughts, floods, frosts, fires, pests, earthquakes, or hurricanes. Column 2 shows the baseline estimates to be robust to allowing the household responses to *Progresa* to also vary by these village characteristics.

An alternative way to address this concern is to randomly reassign each household the family links of another household in the same village. This method leaves the likelihood of any given household being connected in a village to be the same as in the original data, as well as leaving unchanged the overall share of connected households in each village. If our previous measure of being connected or isolated were merely capturing some village level phenomenon, we would not expect the results to differ according to whose family ties any given household were assigned. Reassuring the result in Column 2 shows this not to be the case – when households are assigned to be connected or isolated at random, then both TTE^1 and TTE^0 are significantly greater than zero, and the difference between them is not different from zero.

The third check relates to time varying household characteristics that drive enrolment. Of particular is the fact that households are subject to economic shocks that cause them to take their children out of school [Jacoby and Skoufias 1997]. To address this we use information on whether the household reports being affected been any type of natural shock from October 1998 to November 1999. We control for this directly in (4.3) and allow the response to *Progresa* to vary depending on whether the household has been affected by such shocks. The result in Column 3, shows the results to be robust to the inclusion of such time varying household shocks.

The fourth check relates to the underlying identifying assumption that there are no spillover effects from treatment to control villages. To check for this we restrict our sample to households in villages that are below the median distance (5km) from any health facility, as recorded in May 1999. The result in Column 4 shows that within such villages, where concerns over spillover effects are perhaps greatest, the signs, significance, and magnitude of the baseline estimates continue to hold. An alternative subset of villages in which concerns over spillover effects may be particularly acute are those villages in which there are no secondary or middle schools present. Given that children resident in such locations attend secondary schools outside their own village, these children may be particularly likely to be in schools with children from both treatment and control villages present. Reassuringly, the result in Column 5 shows the previous parameter estimates to be robust to restricting the analysis to this subset of villages.

The final check directly addresses the assumption that extended family networks are not endogenously changing over time in response to the program. To address this we use data from the marital history module collected in May 1999 that explicitly asked spouses about whether their parents were present in the village or not. We use this information to reconstruct extended family ties to parents and hence to reconstruct whether households are connected or isolated. The result in Column 6 shows the previous estimates to be robust this redefinition of extended family links.

References

[1] AKRESH.R (2005) Risk, Network Quality, and Family Structure: Child Fostering Decisions in Burkina Faso, mimeo Yale University.

[2] ALBARRAN.P AND O.ATTANASIO (2003) "Limited Commitment and Crowding Out of Private Transfers: Evidence from a Randomised Experiment", *Economic Journal* 113: C77-85.

[3] ALTONJI.J.G, F.HAYASHI, AND L.J.KOTLIKOFF (1992) "Is the Extended Family Altruistically Linked? Direct Tests Using Micro Data", *American Economic Review* 85: 1177-98.

[4] ALTONJI.J.G, F.HAYASHI, AND L.J.KOTLIKOFF (1997) "Parental Altruism and Inter Vivos Transfers: Theory and Evidence", *Journal of Political Economy* 105: 1121-66.

[5] ANGELUCCI.M AND G.DE GIORGI (2006) Indirect Effects of an Aid Program: How Do Liquidity Injections Affect Non-Eligibles' Consumption?, mimeo Stanford University.

[6] ATTANASIO.O AND V.LECHENE (2002) "Tests of Income Pooling in Household Decisions", *Review of Economic Dynamics* 5: 720-48.

[7] ATTANASIO.O, C.MEGHIR, AND A.SANTIAGO (2005), Education Choices in Mexico: Using a Structural Model and a Randomised Experiment to Evaluate *Progresa*, IFS Working Paper.

[8] BANERJEE.A AND A.NEWMAN (1998) "Information, the Dual Economy, and Development", *Review of Economic Studies* 65: 631-53.

[9] BECKER.G.S (1981) *A Treatise on the Family*, Cambridge: Harvard University Press.

[10] BEHRMAN.J.R AND M.ROSENZWEIG (2006) "Parental Wealth and Adult Children's Welfare in Marriage", *Review of Economics and Statistics* 88: 496-509.

[11] BEHRMAN.J.R, P.SENGUPTA, AND P.TODD (2005), "Progressing Through *Progresa*: An Impact Assessment of a School Subsidy Experiment in Rural Mexico", *Economic Development and Cultural Change* 54: 237-76.

[12] BOBONIS.G AND F.FINAN (2006) Endogenous Peer Effects in School Participation, mimeo, University of Toronto.

[13] CATTANEO.A AND R.LALIVE (2006) Social Interactions in Schooling Decisions, CEPR Working Paper 5816.

[14] CHIAPPORI.P.A (1988) "Rational Household Labor Supply", *Econometrica* 56: 63-90.

[15] CONLEY.T AND C.UDRY (2005) Learning About a New Technology: Pineapple in Ghana, mimeo Yale University.

[16] COX.D AND M.FAFCHAMPS (2007) "Extended Family and Kinship Networks: Economic Insights and Evolutionary Directions", forthcoming *Handbook of Development Economics* Volume 4.

[17] COX.D AND G.JAKUBSON (1995) "The Connection Between Public Transfers and Private Interfamily Transfers", *Journal of Public Economics* 57: 129-67.

[18] DEATON.A (1992) "Saving and Income Smoothing in Côte D'Ivoire", *Journal of African Economies* 1: 1-24.

[19] DERCON.S AND P.KRISHNAN (2000) "In Sickness and in Health: Risk Sharing Within Households in Rural Ethiopia", *Journal of Political Economy* 108: 688-727.

[20] DERCON.S, J.HODDINOTT, AND P.KRISHNAN (2005) Networks and Informal Mutual Support in 15 Villages in Ethiopia, mimeo Oxford University.

[21] DUFLO.E (2003) "Grandmothers and Granddaughters: Old Age Pension and Intra-household Allocation in South Africa", *World Bank Economic Review* 17: 1-25.

[22] FAFCHAMPS.M AND S.LUND (2003) "Risk-sharing Networks in Rural Philippines", *Journal of Development Economics* 71: 261-87.

[23] FOSTER.A (1993) "Household Partition in Rural Bangladesh", *Population Studies* 47: 97-114.

[24] FOSTER.A AND M.ROSENZWEIG (2002) "Household Division and Rural Economic Growth", *Review of Economic Studies* 69: 839-69.

[25] GERTLER.P, S.MARTINEZ, AND M.RUBIO-CODINA (2006) Investing Cash Transfers to Raise Long-Term Living Standards, mimeo University of California Berkeley.

[26] HAMILTON.W.D (1964) "The Genetical Evolution of Social Behavior I and II", *Journal of Theoretical Biology* 7: 1-52.

[27] JACOBY.H.G AND E.SKOUFIAS (1997) "Risk, Financial Markets, and Human Capital in a Developing Country," *Review of Economic Studies* 64: 311-35.

[28] LA FERRARA.E (2003) "Kin groups and Reciprocity: A Model of Credit Transactions in Ghana", *American Economic Review* 93: 1730-51.

[29] LIGON.E (1998) "Risk-Sharing and Information in Village Economies", *Review of Economic Studies* 65: 847-64.

[30] LOURY.L.D (2006) "All in the Extended Family: Effects of Grandparents, Aunts, and Uncles on Educational Attainment", *American Economic Review Papers and Proceedings* 96: 275-8.

[31] MANSER.M AND M.BROWN (1980) "Marriage and Household Decision Theory - A Bargaining Analysis", *International Economic Review* 21: 21-34.

[32] MCELROY.M AND M.HORNEY (1981) "Nash-bargained Decisions: Towards a Generalization of the Theory of Demand", *International Economic Review* 22: 333-49.

[33] MCPHERSON.M, L.SMITH-LOVIN, AND J.M.COOK (2001) "Birds of a Feather: Homophily in Social Networks", *Annual Review of Sociology* 27: 415-44.

[34] MUNSHI.K AND ROSENZWEIG.M (2005) Why is Mobility in India so Low? Social Insurance, Inequality, and Growth, mimeo Brown University.

[35] ROSENZWEIG.M.R (1988) "Risk, Implicit Contracts and the Family in Rural Areas of Low Income Countries", *Economic Journal* 98: 1148-70.

[36] ROSENZWEIG.M.R AND O.STARK (1989) "Consumption Smoothing, Migration, and Marriage: Evidence From Rural India", *Journal of Political Economy* 97: 905-26.

[37] SCHULTZ.T.P (2004) "School Subsidies for the Poor: Evaluating the Mexican *Progresa* Poverty Program", *Journal of Development Economics* 74: 199-250.

[38] SKOUFIAS.E (2005) *Progresa and Its Impacts on the Human Capital and Welfare of Households in Rural Mexico: A Synthesis of the Results of an Evaluation by IFPRI*, Research Report 139, Washington D.C.: International Food Policy Research Institute.

[39] STRAUSS.J AND D.THOMAS (1995) "Human Resources: Empirical Modelling of Household and Family Decisions", in J.R.Behrman and T.N.Srinivasan (editors) *Handbook of Development Economics* Volume 3, Amsterdam: Elsevier Science.

[40] TODD.P AND K.I.WOLPIN (2006) "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility", forthcoming *American Economic Review.*

[41] TOWNSEND.R (1994) "Risk and Insurance in Village India", *Econometrica* 62: 539-91.

[42] TRIVERS.R.L (1972) "Parental Investment and Sexual Selection", in C.Campbell (editor) *Sexual Selection and the Descent of Man,* Chicago: Aldine.

[43] UDRY.C (1994) 'Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria', *Review of Economic Studies* 61: 495-526.

[44] WEISS.Y AND R.J.WILLIS (1985) "Children as Collective Goods and Divorce Settlements", *Journal of Labor Economics* 3: 268-92.

Table 1: The Number of Extended Family Links, by Type of Link

Couple Headed Households

Mean, standard error in parentheses clustered by village

	Outside the Household and in the Village									
	Parent		Children Aged 0-16		Adult Children		Siblings		All	
	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated
From head of household to:	.461	-	-	-	.652	-	2.23	-	3.34	-
	(.010)				(.066)		(.111)		(.164)	
From spouse of household to:	.250	-	-	-	.652	-	1.63	-	2.54	-
	(.007)				(.066)		(.103)		(.160)	

	Inside the Household									
	Parent		Children Aged 0-16		Adult Children		Siblings		All	
	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated
From head of household to:	.062	.079	3.23	3.10	.821	.787	.033	.052	4.15	4.02
	(.003)	(.006)	(.027)	(.043)	(.015)	(.021)	(.002)	(.007)	(.035)	(.056)
From spouse of household to:	.018	.021	3.23	3.10	.821	.787	.013	.017	4.09	3.93
	(.001)	(.003)	(.027)	(.043)	(.015)	(.021)	(.001)	(.003)	(.034)	(.054)

Notes: The sample is restricted to couple headed households in the baseline survey. Standard errors are clustered by village. Of the 22553 households that can be tracked in the first and third waves of *Progresa*, 84.2% report to be couple headed in October 1997 (wave 1). We define the head of the household to be the male among the couple. By construction, the number of family links to parental households is always two conditional on such a family link existing. By construction, the number of children of the couple inside and outside the household are identical for the head and the spouse. Adult children are defined to be at least 17 years of age.

Table 2: Descriptive Evidence on Enrolment Rates

Couple Headed Households

Mean, standard errors in parentheses clustered by village

	Secondary School Enrolment Rates (children aged 11 to 16)							Primary School Enrolment Rates (children aged 6 to 10)			
	Eligibles, by Village Type			**Eligibles, by Family Link Type**				**Eligibles, by Family Link Type**			
	Control	Treatment	All Households	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated
	October 1997	October 1997	Difference in Difference	October 1997	October 1997	Difference in Difference	Difference in Difference	October 1997	October 1997	Difference in Difference	Difference in Difference
All children	.651	.654	.069***	.653	.654	.083***	.001	.927	.890	.013	-.011
	(.016)	(.012)	(.016)	(.011)	(.015)	(.017)	(.031)	(.005)	(.013)	(.012)	(.021)
Boys	.681	.685	.043**	.681	.698	.044*	.040	.932	.887	.020	-.001
	(.017)	(.013)	(.022)	(.012)	(.016)	(.023)	(.047)	(.006)	(.014)	(.015)	(.031)
Girls	.603	.612	.102***	.608	.610	.131***	-.030	.925	.901	.005	-.015
	(.018)	(.014)	(.022)	(.012)	(.019)	(.024)	(.044)	(.007)	(.012)	(.015)	(.026)
	Non-eligibles, by Village Type			**Non-eligibles, by Family Link Type**				**Non-eligibles, by Family Link Type**			
	Control	Treatment	All Households	Connected	Isolated	Connected	Isolated	Connected	Isolated	Connected	Isolated
	October 1997	October 1997	Difference in Difference	October 1997	October 1997	Difference in Difference	Difference in Difference	October 1997	October 1997	Difference in Difference	Difference in Difference
All children	.594	.676	-.019	.638	.666	-.020	-.011	.960	.956	.009	-.003
	(.020)	(.018)	(.026)	(.015)	(.025)	(.027)	(.061)	(.006)	(.012)	(.024)	(.035)
Boys	.616	.693	.009	.662	.668	.004	.018	.966	.938	.005	-.040
	(.026)	(.021)	(.036)	(.018)	(.032)	(.039)	(.081)	(.007)	(.018)	(.033)	(.066)
Girls	.589	.646	-.016	.616	.652	-.021	.009	.952	.972	.038	.060
	(.023)	(.022)	(.037)	(.018)	(.032)	(.041)	(.074)	(.009)	(.015)	(.034)	(.041)

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in secondary school at the time of the survey. A household's primary school enrolment rate is defined to be the fraction of children aged 6 to 10 resident in the household that are full-time enrolled in primary school at the time of the survey. In the left hand panel, the difference in difference is defined to be the difference in enrolment rates between households in treatment and control villages in November 1999, minus the corresponding difference at baseline in October 1997. In the centre and right hand panels, this difference in difference is reported for connected and isolated households separately. Standard errors on the differences are derived from an OLS regression, estimated on eligible and non-eligibles separately, of school enrolment rates on a dummy equal to one for *Progresa* villages and zero otherwise. Standard errors are clustered by village.

Table 3: Probability of an Extended Family Link

Couple Headed Households

Mean, standard errors in parentheses clustered by village

	Any Family Link (Connected)	Any Family Link of the Head	Any Family Link of the Spouse	Intra-generational Family Links: Head to Head (Brothers)	Intra-generational Family Links: Head to Spouse	Intra-generational Family Links: Spouse to Head	Intra-generational Family Links: Spouse to Spouse (Sisters)	Inter-generational Family Links: Parents to Son	Inter-generational Family Links: Parents to Daughter	Inter-generational Family Links: Son to Parent	Inter-generational Family Links: Daughter to Parent
Eligible Households											
Treatment	.817	.693	.550	.506	.351	.338	.306	.150	.077	.169	.108
	(.011)	(.012)	(.014)	(.013)	(.015)	(.014)	(.013)	(.009)	(.006)	(.008)	(.007)
Control	.800	.682	.541	.503	.364	.348	.314	.149	.079	.163	.097
	(.017)	(.020)	(.023)	(.023)	(.026)	(.026)	(.026)	(.014)	(.009)	(.011)	(.008)
Difference	.017	.012	.009	.003	-.013	-.010	-.008	.002	-.002	.006	.011
	(.020)	(.023)	(.028)	(.026)	(.030)	(.030)	(.030)	(.016)	(.010)	(.013)	(.011)
Non-eligible Households											
Treatment	.808	.712	.523	.532	.332	.258	.248	.198	.112	.142	.074
	(.016)	(.018)	(.021)	(.021)	(.018)	(.016)	(.017)	(.015)	(.010)	(.014)	(.008)
Control	.802	.694	.562	.481	.353	.313	.272	.226	.112	.126	.089
	(.019)	(.022)	(.024)	(.026)	(.028)	(.022)	(.022)	(.021)	(.016)	(.014)	(.010)
Difference	.006	.019	-.039	.051	-.021	-.055**	-.024	-.028	-.000	.016	-.015
	(.025)	(.028)	(.032)	(.034)	(.034)	(.027)	(.028)	(.026)	(.019)	(.020)	(.013)

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered by village. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Means and differences are reported for those households that have secondary school age children (aged 11 to 16) in the baseline survey of October 1997. The standard errors on the differences are calculated from running a corresponding OLS regression, which allows for the error terms to be clustered by village.

Table 4: Family Network Descriptives

Means, standard deviation between villages in parentheses, standard deviation within villages in brackets

Treatment Villages	Size of Global Family Network	Network Size/Number of Households in Village	Diameter	Share that are Eligible	Share With Primary School Aged Children at Baseline	Share With Secondary School Aged Children at Baseline	Average Value of Potential Transfers Households in Network are Eligible for (March 1998, pesos)
Mean	7.66	.169	2.42	.518	.480	.480	746
Standard deviation between villages	(.249)	(.153)	(1.20)	(.233)	(.150)	(.139)	(181)
Standard deviation within villages	[.153]	[.153]	[2.14]	[.259]	[.262]	[.263]	[281]

Control Villages	Size of Global Family Network	Network Size/Number of Households in Village	Diameter	Share that are Eligible	Share With Primary School Aged Children at Baseline	Share With Secondary School Aged Children at Baseline	Average Value of Potential Transfers Households in Network are Eligible for (March 1998, pesos)
Mean	7.92	.163	2.51	.525	.484	.491	756
Standard deviation between villages	(.230)	(.141)	(1.14)	(.249)	(.145)	(.139)	(188)
Standard deviation within villages	[.155]	[.155]	[2.07]	[.229]	[.252]	[.258]	[297]

Notes: The sample is restricted to households that can be tracked over the first and third *Progresa* waves. There is one observation per family network so that each network has the same weight irrespective of the number of households within it. There are 1379 family networks in treatment villages covering 10559 households. There are 817 family networks in control villages covering 6471 households. The size of the network is the number of households in the network. The diameter of the networks is the longest distance between two households that exists in a network. We define two households that are directly connected to be of distance one to each other. Primary school aged children are defined to be those aged 6 to 10 and resident in the household. Secondary school aged children are defined to be those aged 11 to 16 and resident in the household. The average value of potential transfers households in the network are eligible for, are calculated among eligible households only. The standard deviations between and within villages take account of the fact that there are an unequal number of family networks in each village.

Table 5: Extended Families and the Response to *Progresa*, Baseline Specification

Dependent Variable (Columns 1a-5b): Change in Household's Secondary School Enrolment Rate (November 1999 - October 1997)
Dependent Variable (Columns 6): Household's Secondary School Enrolment Rate at Baseline (October 1997)
OLS regression estimates, standard errors are clustered by village

	All Children					Boys		Girls		Boys	Girls	All Children
	Standard	Connected	Isolated	Family Links	Interactions	Standard	Interactions	Standard	Interactions	High Adult Wage Villages		Secondary Enrolment Rate at Baseline (October 1997)
	(1a)	(1b)	(1c)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(6)
TTE	.078***	.098***	-.002			.054**		.132***				
	(.016)	(.017)	(.032)			(.024)		(.026)				
ITE	-.015	-.012	-.008			.008		-.024				
	(.026)	(.028)	(.063)			(.039)		(.042)				
TTE [connected]				.093***	.092***		.055**		.126***	.040	.095***	
				(.017)	(.018)		(.025)		(.026)	(.028)	(.029)	
TTE [isolated]				.009	.006		.041		-.035	.038	-.013	
				(.032)	(.032)		(.048)		(.046)	(.056)	(.050)	
ITE [connected]				-.016	-.027		.008		-.035	.005	-.020	
				(.028)	(.030)		(.042)		(.046)	(.049)	(.050)	
ITE [isolated]				-.007	-.012		.020		-.006	.056	.019	
				(.061)	(.061)		(.083)		(.077)	(.089)	(.080)	
Average poverty index of households in the extended family / 100												.018***
												(.006)
ΔTTE				.084**	.086**		.013		.161***	.002	.107*	
				(.035)	(.036)		(.052)		(.052)	(.061)	(.057)	
ΔITE				-.009	-.015		-.012		-.029	-.052	-.039	
				(.065)	(.065)		(.093)		(.085)	(.102)	(.091)	
Observations	6227	5115	1112	6227	6227	3947	3947	3760	3760	2844	2765	5919

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in school at the time of the survey. In Columns 2a onwards the link variable is defined to be equal to one if household h has any family links in the village, and zero otherwise. All specifications also control for the following - the husband's age, literacy, whether he speaks an indigenous language, the spouse's age, literacy, whether she speaks an indigenous language, the household poverty index, whether the household owns any land, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, regional fixed effects, and the village level enrolment rate at baseline among eligible and non-eligible households. In Columns 2b, 3b, 3c, 4b, 5a, and 5b the effects of the following controls are also allowed to vary with eligibility status, *Progresa*, and the interaction of the two - whether the head's (spouse's) age is above or below the median among couple headed households, whether the head (spouse) is literate, whether the household owns land, whether the household size at baseline is above or below the median among couple headed households, and the village level enrolment rates at the baseline among eligible and non-eligible households. The samples in Columns 5a and 5b are restricted to villages that have above the median level of adult wages as recorded in October 1998. In Column 6 the sample is restricted to connected households that have at least one secondary school aged child at baseline (October 1997), the dependent variable is the household's secondary enrolment rate at baseline, and the same set of controls as in Column 1a is included. The household poverty index increases as the household has higher permanent income.

Table 6: Extended Families, Transfers, and the Response to *Progresa*

Dependent Variable: Change in Household's Secondary School Enrolment Rate (November 1999 - October 1997)
OLS regression estimates, standard errors are clustered by village

	Isolated	Connected	Connected Households, Varying Characteristics of the Local Family Network	
	Own Transfer	Own and Family Transfer	Minority of Family is Eligible and Has Primary School Aged Children	Majority of Family is Eligible and Has Primary School Aged Children
Households With Primary School Aged Children	(1)	(2)	(3)	(4)
TTE [evaluated at mean of own transfer]	-.010			
	(.043)			
TTE [evaluated at mean of own + family transfer]		.093***	.133***	.076**
		(.021)	(.035)	(.030)
Evaluated at mean of own transfer, family transfer = 0			.045	
			(.030)	
Evaluated at mean of own transfer, family transfer at 75th percentile				.083***
				(.031)
Observations	659	3180	1622	1558
Households With No Primary School Aged Children				
TTE [evaluated at mean of own transfer]	.077			
	(.086)			
TTE [evaluated at mean of own + family transfer]		.119***	.053	.116*
		(.044)	(.079)	(.067)
Evaluated at mean of own transfer, family transfer = 0			.096	
			(.077)	
Evaluated at mean of own transfer, family transfer at 75th percentile				.142**
				(.064)
Observations	194	814	457	357

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to eligible couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in school at the time of the survey. For the own transfer effects, the value of potential transfers to the households is evaluated at the mean of the distribution of such transfers in October 1998 among all eligible couple headed households that have secondary school aged children. In Column 2 onwards, the family refers to the local family network at degree one to the household. The value of potential transfers to the local family network that is eligible is evaluated at the mean of the distribution of such transfers in October 1998 among the local family network (excluding the household itself) of all connected eligible couple headed households that have secondary school aged children. All specifications also control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, regional fixed effects, and the village level enrolment rate at baseline among eligible and non-eligible households. In all columns the effects of the following controls are also allowed to vary with *Progresa* - whether the head's (spouse's) age is above or below the median among couple headed households, whether the head (spouse) is literate, whether the household owns land, whether the household size at baseline is above or below the median among couple headed households, and the village level enrolment rates at the baseline among eligible and non-eligible households.

Table 7: Eligibility Status of Intra-generational Family Links

Dependent Variable: Change in Household's Secondary School Enrolment Rate (November 1999 - October 1997)
OLS regression estimates, standard errors are clustered by village

Intra-generational Family Link Type:	Brothers	Sisters	Brothers	Sisters	Brothers	Sisters	Brothers	Sisters
	Baseline		Interactions		Family Heterogeneity		Randomly Assigned	
Relationship to secondary school age children in household h:	Father to Uncle	Mother to Aunt	Father to Uncle	Mother to Aunt	Father to Uncle	Mother to Aunt	Father to Uncle	Mother to Aunt
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
TTE [linked household is eligible]	.134***	.106***	.125***	.131***	.125***	.096**	.064**	.083**
	(.038)	(.040)	(.041)	(.044)	(.037)	(.041)	(.029)	(.038)
TTE [linked household is non-eligible]	-.010	.061	-.000	.068	.015	.094	.102***	.120***
	(.046)	(.058)	(.046)	(.059)	(.047)	(.061)	(.028)	(.046)
Observations	1366	1064	1366	1064	1366	1064	2990	1671

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in school at the time of the survey. All specifications also control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, regional fixed effects, and the village level enrolment rate at baseline among eligible and non-eligible households. In Columns 2a and 2b the effects of the following controls are also allowed to vary with eligibility status, *Progresa*, and the interaction of the two - whether the head's (spouse's) age is above or below the median among couple headed households, whether the head (spouse) is literate, whether the household owns land, whether the household size at baseline is above or below the median among couple headed households, and the village level enrolment rates at the baseline among eligible and non-eligible households. In Columns 3a and 3b, the standard deviation of the poverty index within the extended family members of households h (including household h itself) is allowed to vary with eligibility status, *Progresa*, and the interaction of the two.

Table 8: Schooling and Eligibility Status of Intra-generational Family Links

Dependent Variable: Change in Household's Secondary School Enrolment Rate (November 1999 - October 1997)
OLS regression estimates, standard errors are clustered by village

Inter-generational Family Link Type:	True Siblings				Randomly Assigned Siblings			
	Brothers	Sisters	Brothers	Sisters	Brothers	Sisters	Brothers	Sisters
Relationship to secondary school age children in household h:	Father to Uncle	Mother to Aunt	Father to Uncle	Mother to Aunt	Father to Uncle	Mother to Aunt	Father to Uncle	Mother to Aunt
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
TTE [linked household is eligible and has secondary school age children]	.096*	.139***			.080**	.103*		
	(.052)	(.053)			(.036)	(.055)		
TTE [linked household is eligible and has only primary school age children]	.215***	.183**			.031	-.041		
	(.075)	(.085)			(.067)	(.085)		
TTE [linked household is non-eligible and has secondary school age children]			.008	.040			.102**	.174**
			(.056)	(.078)			(.044)	(.072)
TTE [linked household is non-eligible and has only primary school age children]			.106	-.025			.088	-.040
			(.176)	(.152)			(.103)	(.133)
Observations	631	524	369	302	1153	743	686	297

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in school at the time of the survey. All specifications also control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, regional fixed effects, and the village level enrolment rate at baseline among eligible and non-eligible households.

Table 9: Eligibility Status of Inter-generational Family Links

Dependent Variable: Change in Household's Secondary School Enrolment Rate (November 1999 - October 1997)
OLS regression estimates, standard errors are clustered by village

Inter-generational Link Type:	Parents to Son	Parents to Daughter	Parents to Son	Parents to Daughter	Son to Parents	Daughter to Parents
Relationship to secondary school age children in household h:	Father to Adult Brother	Mother to Adult Sister	Father to Adult Brother	Mother to Adult Sister	Father to Paternal Grandparents	Mother to Maternal Grandparents
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
TTE [linked household is eligible]	.206***	.173***			.085	.102*
	(.045)	(.054)			(.053)	(.060)
TTE [linked household is non-eligible]	.078	.088			.001	.077*
	(.077)	(.120)			(.037)	(.046)
TTE [linked household is eligible and has primary school aged children]			.148**	.171**		
			(.063)	(.078)		
TTE [linked household is eligible and has no primary school aged children]			.269***	.196**		
			(.060)	(.083)		
Observations	965	499	690	384	1028	649

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in school at the time of the survey. All specifications also control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, regional fixed effects, and the village level enrolment rate at baseline among eligible and non-eligible households.

Table 10: Extended Families and the Response to *Progresa* in Consumption

OLS regression estimates, standard errors are clustered by village

Dependent Variable:	Change in Weekly Consumption of Meat (Nov 1999 - Mar 1998)	Proportionate Change in Non Food Expenditures (Nov 1999 - Mar 1998)
	(1)	(2)
TTE [connected]	.409***	.159***
	(.055)	(.048)
TTE [isolated]	.312***	.221***
	(.087)	(.071)
ITE [connected]	.228***	.027
	(.076)	(.072)
ITE [isolated]	.298***	.114
	(.113)	(.111)
ΔTTE	.097	-.062
	(.084)	(.076)
ΔITE	-.070	-.087
	(.110)	(.116)
Mean (sd) of consumption variable [March 1998]		
Connected households	.867	362
	(.867)	(336)
Isolated households	.922	367
	(.901)	(335)
Observations	11681	10844

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village. The dependent variable in Column 1 is the change in the number of days in a week the household reports consuming meat between November 1999 and March 1998. The dependent variable in Column 2 is the proportionate change in monthly non food expenditures, measured in pesos, between November 1999 and March 1998. This is defined as the change in non food expenditures over this time period, divided by the level of non food expenditures in March 1998. This does not include any schooling related expenditures. All specifications also control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, regional fixed effects, and the village level enrolment rate at baseline among eligible and non-eligible households. In all columns the effects of the following controls are also allowed to vary with eligibility status, *Progresa*, and the interaction of the two - whether the head's (spouse's) age is above or below the median among couple headed households, whether the head (spouse) is literate, whether the household owns land, and whether the household size at baseline is above or below the median among couple headed households.

Table A1: Descriptive Statistics on Surnames, by Surname Type

Mean, standard errors in parentheses, percentages in brackets

	Head's Paternal Surname (F1)	Head's Maternal Surname (f1)	Spouse's Paternal Surname (F2)	Spouse's Maternal Surname (f2)
Number of surnames	1696	1996	1912	2025
Number [percentage] of surnames mentioned more than once	1064 [62.7]	1188 [59.5]	1088 [56.9]	1100 [54.3]
Probability of same surname in population	9.50×10^{-6}	7.54×10^{-6}	8.60×10^{-6}	8.33×10^{-6}
	(5.48×10^{-6})	(4.16×10^{-6})	(4.95×10^{-6})	(4.95×10^{-6})
Expected number of same surname matches in population	13.3	11.2	9.92	9.26
	(1.66)	(1.36)	(1.25)	(1.19)
Probability of same surname in the village	.042	.021	.022	.020
	(.0005)	(.0004)	(.0004)	(.0004)
Expected number of same surname matches in the village	7.55	5.31	5.42	4.98
	(.039)	(.036)	(.036)	(.040)
Odds ratio	355.7	344.8	345.4	353.0
	(8.26)	(7.47)	(7.55)	(8.18)

Notes: For the matching probabilities and expected number of same surname matches in the population, the standard errors are clustered by surname for each surname type. The sample is restricted to those households that can be tracked for the first and third waves of the *Progresa* data, namely in the baseline survey in October 1997 (wave 1) and the first post program survey in October 1998 (wave 3). There are 22553 such households.

Table A2: The Number of Family Links, by Type of, as Reported in the Mexican Family Life Survey

Couple Headed Households

Mean, standard error in parentheses clustered by village

	Outside of the Household (ANY location)				
	Parent	**Children Aged 0-16**	**Adult Children**	**Siblings**	**All**
From head of household to:	.476	-	1.23	3.27	4.97
	(.035)		(.089)	(.116)	(.014)
From spouse of household to:	.669	-	1.23	3.50	5.39
	(.039)		(.089)	(.113)	(.148)

	Inside of the Household				
	Parent	**Children Aged 0-16**	**Adult Children**	**Siblings**	**All**
From head of household to:	.047	2.02	.571	.019	2.66
	(.009)	(.079)	(.039)	(.007)	(.084)
From spouse of household to:	.002	2.02	.571	.009	2.60
	(.002)	(.079)	(.039)	(.005)	(.082)

Notes: The sample is taken from the first wave of the Mexican Family Life Survey, 2001. Standard errors are clustered by village. We restrict this sample to the seven Mexican states that are also covered in the *Progresa* evaluation data, and to couple headed households, in locations with less than 2500 inhabitants. There are 580 such households. By construction, the number of family links to parental households is always conditional on two such family links existing. We do not therefore use information on households that have single parents in any location. By construction, the number of children of the couple inside and outside of the household are identical for the head and the spouse. The number of children outside of the household is restricted to be 17 and older (based on spouses' reports).

Table A3: Correlates of Extended Family Links

Couple Headed Households

Conditional logit estimates, grouped on village, standard errors clustered by village, log odds ratios reported

	Inter-generational Family Links			Intra-generational Family Links				
Type of Family Link:	**Parents of Head**	**Parents of Spouse**	**Adult Child**	**Brothers of Head**	**Sisters of Head**	**Brothers of Spouse**	**Sisters of Spouse**	**Any Link [Connected]**
Head age [years]	.373***	.307***	1.48***	.953	.805***	.792***	.841***	.768***
	(.023)	(.024)	(.079)	(.041)	(.037)	(.034)	(.041)	(.041)
Spouse age [years]	.755***	1.02	2.97***	.872***	.941	1.20***	.969	1.18***
	(.043)	(.072)	(.166)	(.037)	(.041)	(.052)	(.047)	(.063)
Head literate [yes=1]	1.45***	1.13	.911*	1.07	1.13***	.966	.993	1.05
	(.099)	(.095)	(.049)	(.046)	(.056)	(.041)	(.046)	(.052)
Spouse literate [yes=1]	1.14**	1.30***	.785***	1.11**	1.03	1.08	.956	1.13**
	(.074)	(.100)	(.039)	(.047)	(.046)	(.048)	(.049)	(.062)
Head speaks indigenous language [yes=1]	1.01	.843	1.16	.963	1.02	.967	1.02	.911
	(.186)	(.138)	(.167)	(.107)	(.125)	(.102)	(.105)	(.115)
Spouse speaks indigenous language [yes=1]	.854	1.05	1.02	.936	.998	1.33	1.17	.998
	(.144)	(.148)	(.145)	(.094)	(.133)	(.175)	(.151)	(.154)
House is owned [yes=1]	1.03	1.08	1.34**	1.54***	1.38***	1.26***	1.33***	1.43***
	(.094)	(.122)	(.200)	(.112)	(.112)	(.107)	(.133)	(.126)
Any land is owned [yes=1]	.846***	1.00	1.26***	1.14***	1.19***	1.11***	1.14***	1.14**
	(.046)	(.069)	(.077)	(.050)	(.059)	(.047)	(.056)	(.062)
Any member temporarily migrated in last year [yes=1]	1.09	.993	1.15**	.937	.967	1.05	.993	1.01
	(.066)	(.071)	(.071)	(.044)	(.044)	(.051)	(.050)	(.064)
Eligible [yes=1]	.963	1.07	1.02	1.01	1.03	1.03	1.05	1.02
	(.069)	(.091)	(.070)	(.051)	(.056)	(.055)	(.059)	(.064)
Poverty index	1.02	1.03	1.19***	1.08**	1.05	.975	1.02	1.00
	(.042)	(.052)	(.048)	(.035)	(.036)	(.031)	(.036)	(.042)
Household size	.967	1.18***	1.11***	1.10***	1.11***	1.07***	1.11***	.992
	(.030)	(.040)	(.028)	(.021)	(.022)	(.021)	(.024)	(.023)
Mean of Dependent Variable	.187	.101	.217	.485	.332	.306	.274	.807
Number of Observations	18309	17046	18634	18907	18686	18740	17648	18611

Notes: *** denotes that the odds ratio is significantly different from one at 1%, ** at 5%, and * at 10%. In each column a conditional logit specification is estimated, grouped on village, where the standard errors are clustered by village, and the log odds ratios are reported. All continuous variables are divided by their standard deviation so that the corresponding coefficients can be interpreted as the effect of a one standard deviation change in the continuous variable. The underlying sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. The sample varies across the columns because villages in which all or no households have the given type of family link are dropped when the conditional logit regression is estimated. All characteristics are measured in the third wave (October 1998) except household size which is measured at baseline. A higher household poverty index implies the household has a higher level of permanent income and so is less poor.

Table A4: Matching Algorithm and Surnames Based Robustness Checks

Dependent Variable (Columns 1 to 3): Change in Household's Secondary School Enrolment Rate (November 1999 - October 1997)
OLS regression estimates, standard errors are clustered by village

	Omit Households With Potential Measurement Error in Surnames	Drop Villages in Top Quartile of Village Size	Random Reassignment of Households To Another Village in the Same Municipality	Number of Children Aged 0-16 in Household At Baseline	
	(1)	(2)	(3)	(4a) OLS	(4b) Most Frequent Names
TTE [connected]	.096***	.131***	.069***		
	(.019)	(.023)	(.026)		
TTE [isolated]	.005	.016	.072***		
	(.033)	(.040)	(.020)		
ITE [connected]	-.015	.004	.016		
	(.031)	(.040)	(.043)		
ITE [isolated]	-.010	-.082	-.047		
	(.062)	(.070)	(.040)		
Share of households with same head's paternal surname				-.293	-.054
				(.444)	(.483)
Share of households with same spouse's paternal surname				.043	.329
				(.473)	(.489)
ΔTTE	.091***	.115**	-.003		
	(.037)	(.047)	(.029)		
ΔITE	-.005	.086	.063		
	(.065)	(.074)	(.058)		
Observations	5490	3121	5447	6227	3954

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village in each column. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in school at the time of the survey. The link variable is defined to be equal to one if household h has any family links of type j in the village, and zero otherwise. The specifications in all columns except 4a and 4b also control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, and the village level enrolment rate at baseline among eligible and non eligible households. In all columns except 4a and 4b the effects of the following controls are also allowed to vary with eligibility status, *Progresa*, and the interaction of the two - whether the head's (spouse's) age is above or below the median among couple headed households, whether the head (spouse) is literate, whether the household owns land, whether the household size at baseline is above or below the median among couple headed households, and the village level enrolment rates at the baseline among eligible and non-eligible households. In Column 1 we drop from the sample households in which the - (i) spouse's maternal surname is the same as her husband's; (ii) wife's paternal and maternal surnames are the same; (iii) paternal and maternal surnames of both the head and spouse are the same. In Column 2 the sample is restricted to villages with less than 57 households in them. In Column 3 we randomly reassign each household in our baseline sample to another village within the same municipality and recalculate their extended family links if they actually lived in that alternative village. In Columns 4a and 4b the dependent variable is the number of children aged 0-16 in the household at baseline and control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, and village fixed effects. In Column 4b the sample is restricted to households in which either the head's or spouse's paternal surname is shared by .5% of the population.

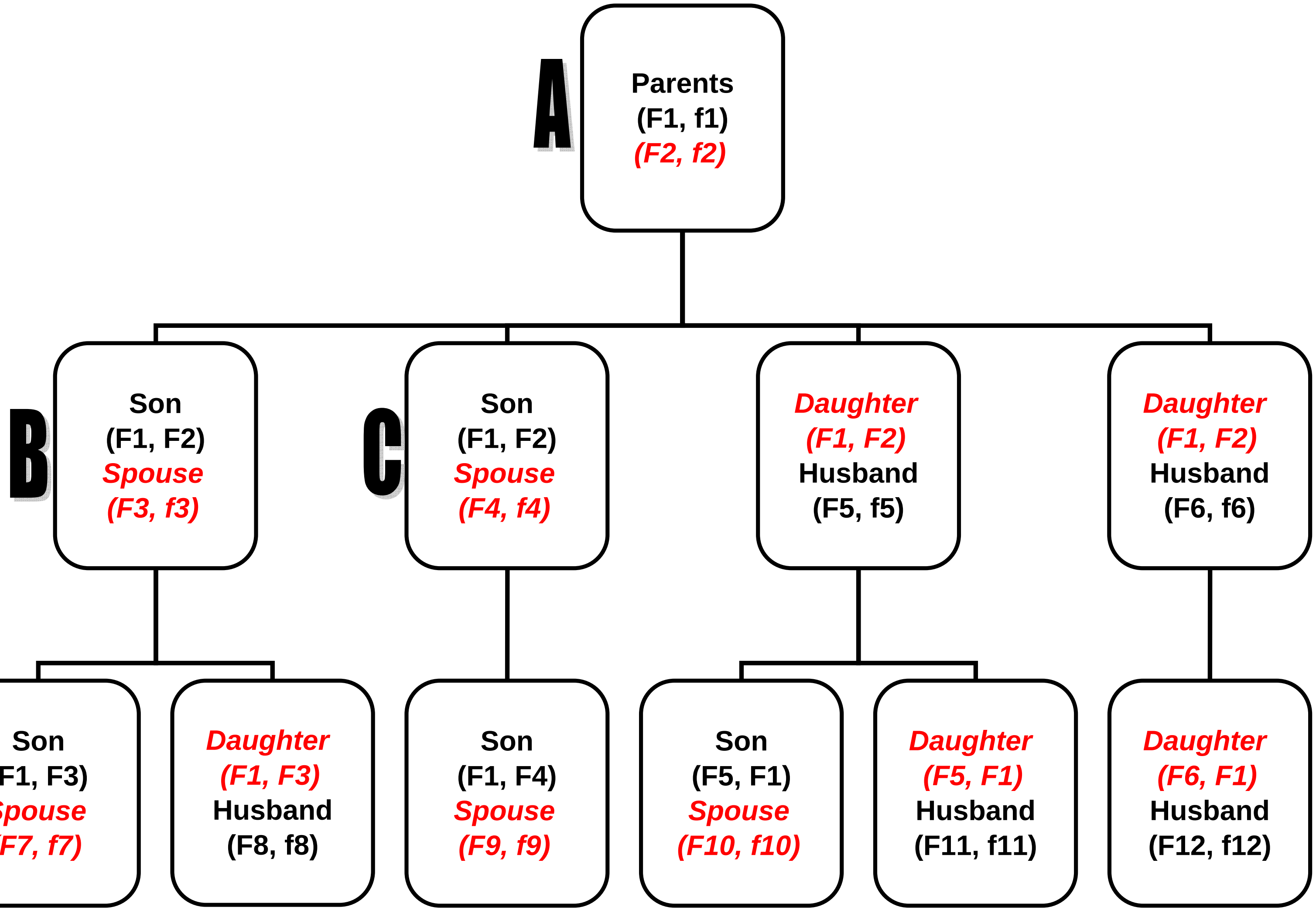
Table A5: Village and Household Level Characteristics Based Robustness Checks

Dependent Variable: Change in Household's Secondary School Enrolment Rate (November 1999 - October 1997)
OLS regression estimates, standard errors are clustered by village

	Village Interactions	Random Reassignment To Another Household in the Same Village	Household Shocks	Close to Any Health Facility (Less Than 5km)	No Secondary or Middle School in the Village	Marital History Module Defined Links (May 1999)
	(1)	(2)	(3)	(4)	(5)	(6)
TTE [connected]	.095***	.074***	.095***	.111***	.114***	.095***
	(.017)	(.017)	(.018)	(.031)	(.021)	(.018)
TTE [isolated]	.013	.098***	.010	.011	.020	-.023
	(.033)	(.029)	(.032)	(.048)	(.038)	(.033)
ITE [connected]	-.004	-.031	-.001	-.077	-.011	-.020
	(.039)	(.032)	(.029)	(.061)	(.040)	(.030)
ITE [isolated]	.010	.011	.001	-.069	-.012	-.055
	(.062)	(.052)	(.060)	(.090)	(.077)	(.064)
ΔTTE	.083**	-.024	.086**	.100*	.094**	.118***
	(.036)	(.030)	(.035)	(.053)	(.042)	(.037)
ΔITE	-.014	-.041	-.002	-.008	-.001	.035
	(.065)	(.059)	(.064)	(.107)	(.083)	(.066)
Observations	6227	6227	6227	2118	4347	6227

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. The sample is restricted to couple headed households that can be tracked over the first and third *Progresa* waves. Standard errors are clustered by village in each column. A household's secondary school enrolment rate is defined to be the fraction of children aged 11 to 16 resident in the household that are full-time enrolled in school at the time of the survey. The link variable is defined to be equal to one if household h has any family links of type j in the village, and zero otherwise. The specifications in all columns also control for the following - the husband's age, years of schooling, literacy, whether he speaks an indigenous language, the spouse's age, years of schooling, literacy, whether she speaks an indigenous language, the household poverty index, the number of individuals in the household at baseline, the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, regional fixed effects, and the village level enrolment rate at baseline among eligible and non eligible households. In all columns the effects of the following controls are also allowed to vary with eligibility status, *Progresa* , and the interaction of the two - whether the head's (spouse's) age is above or below the median among couple headed households, whether the head (spouse) is literate, whether the household owns land, whether the household size at baseline is above or below the median among couple headed households, and the village level enrolment rates at the baseline among eligible and non-eligible households. In Column 1 the effects of the following village characteristics are also allowed to vary with eligibility status, *Progresa* , and the interaction of the two – the number of households in the village, the share of households in the village that are eligible, the marginality index for the village, and the share of households that report being affected by any natural shocks from October 1998 to November 1999. These shocks include being affected by droughts, floods, frosts, fires, pests, earthquakes, or hurricanes. In Column 2 we randomly reassign each household in our baseline sample the family links of another household in the same village. In Column 3 we allow household responses to *Progresa* to vary by whether they themselves have been affected by any shock from October 1998 to November 1999. In Column 4 the sample is restricted to villages that are less than the median distance (5km) from any health facility as measured in November 1999 (wave 5). In Column 5 the sample is restricted to the 410 villages in which there is no secondary or middle school present. In Column 6 we redefine the family links based on information on the presence of parental links in the village collected in the marital history module in May 1999 (wave 4).

Figure 1: Family Tree



Notes: We use the convention that the head's surnames are written in standard (black) font, and those of his wife are written in (red) italics. Paternal surnames are indicated in upper case (F1, *F2*) and maternal surnames are indicated in lower case (f1, *f2*). First names are not shown as they are not relevant for the construction of extended family ties. Each household in the family tree is assumed to be couple headed purely to ease the exposition.

Figure 2a: Potential Transfers in Eligible Isolated and Connected Households

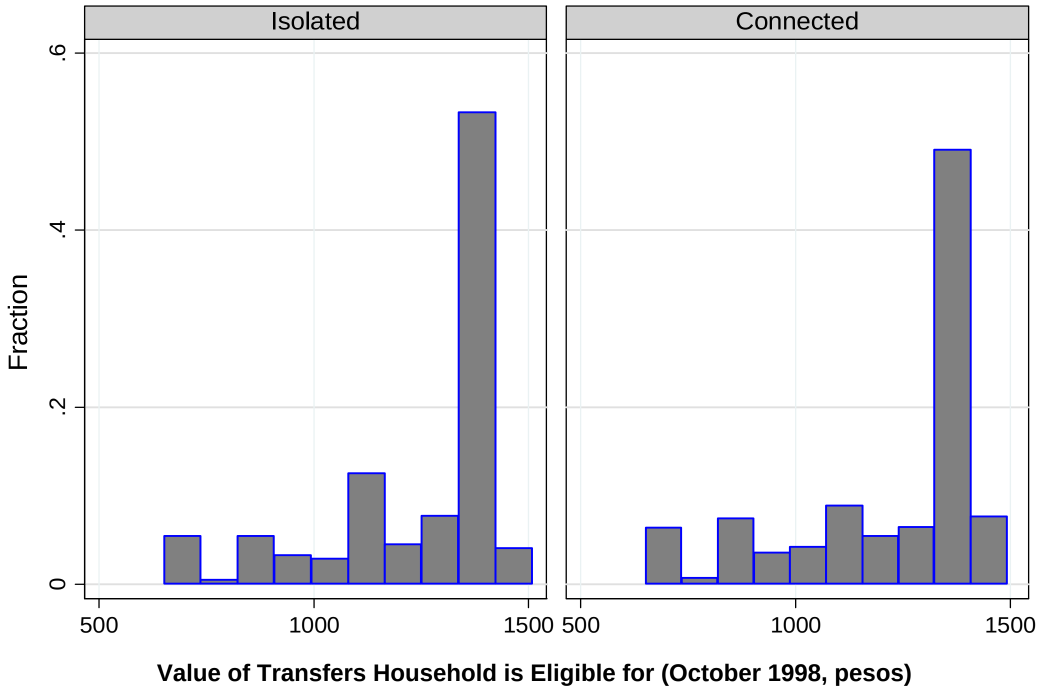
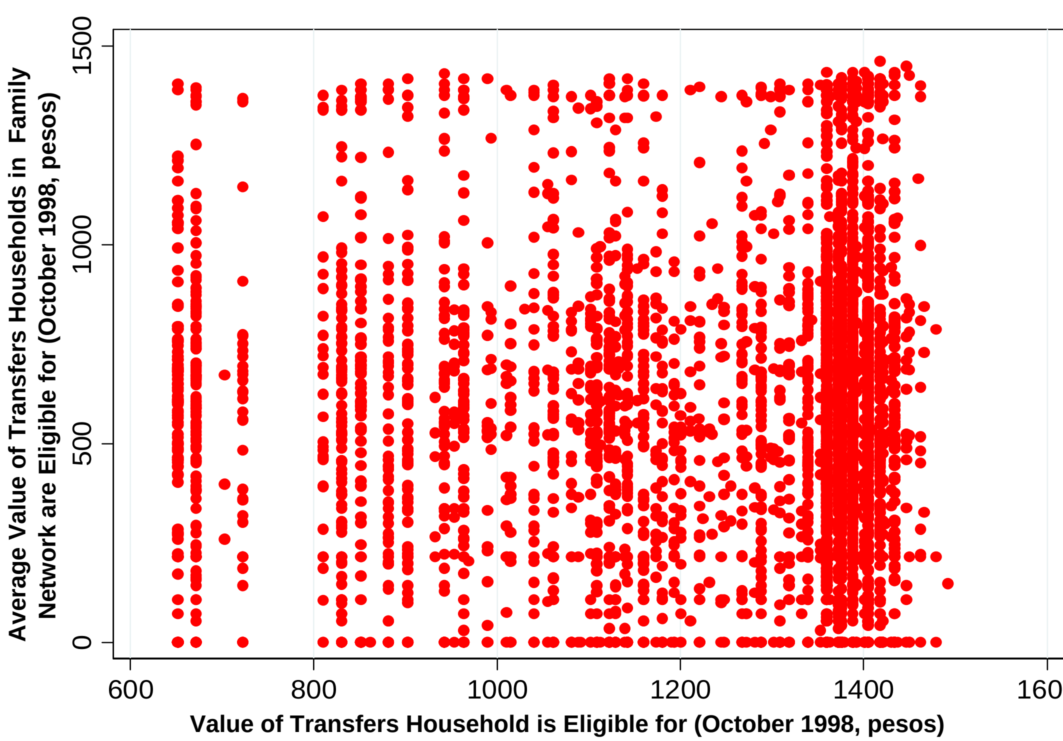


Figure 2b: Potential Transfers in Eligible Households and Their Family Network



Notes: The samples for both figures are based on eligible couple headed households that can be tracked over the first and third *Progresa* waves. Figure 2a plots the potential transfers each eligible connected household with secondary school aged children may receive, as recorded in October 1998, split for isolated and connected households. Figure 2b plots the potential transfers each eligible connected household with secondary school aged children may receive against the average potential transfer eligible households in its family network of degree 1, are entitled to, as recorded in October 1998. The correlation between these two potential sources of income for connected households is -.01.

Figure A1: Propensity Scores For Connected and Isolated Households, by Eligibility Status

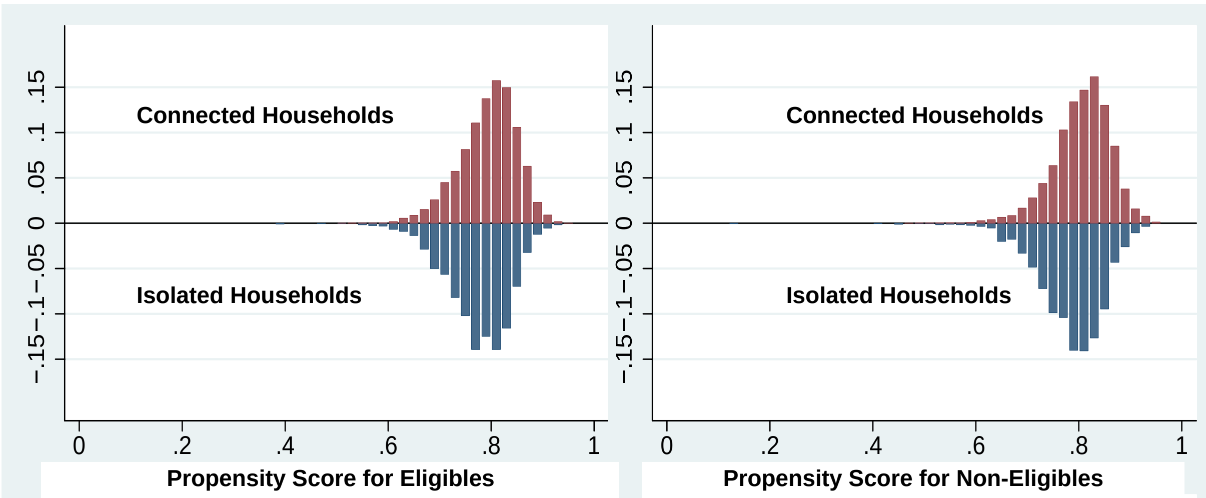
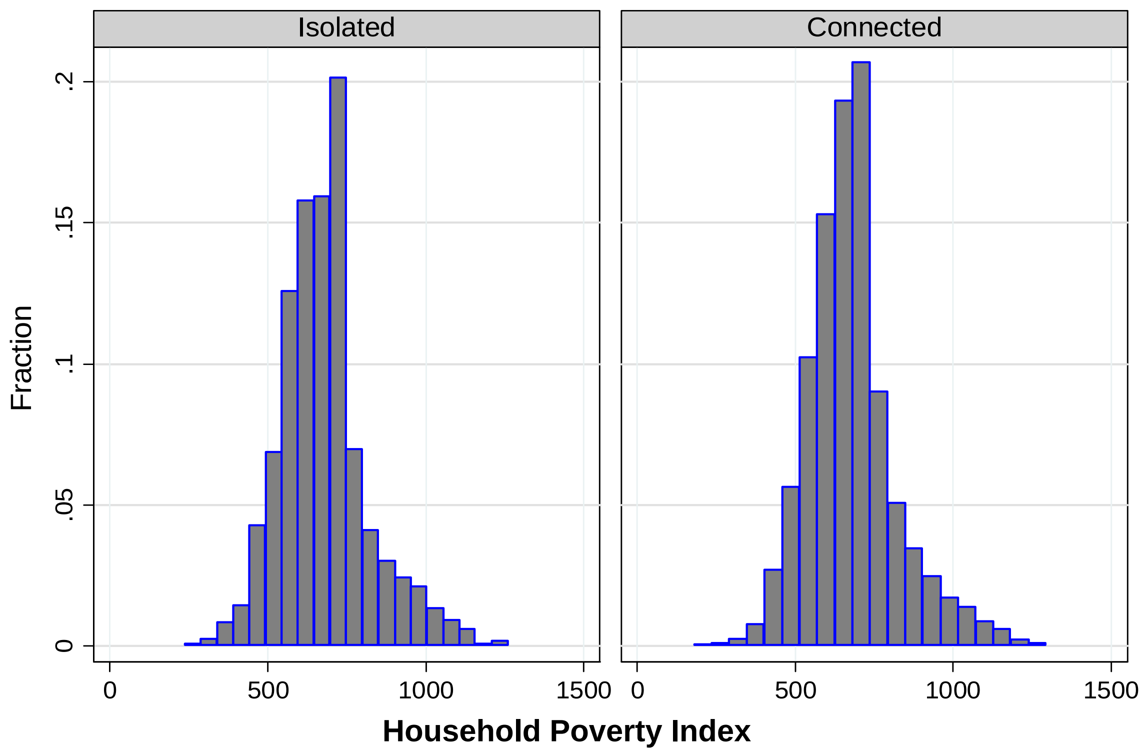


Figure A2: The Household Poverty Index, for Connected and Isolated Households



Notes: In Figure A1, the propensity score is based on the following observable characteristics - the ages of the head and spouse, dummy variables for whether they are working, literate, and speak an indigenous language, whether the household owns its home, whether any land is owned, whether any member of the household temporarily migrated in the last year, the household poverty index, the number of male (females) aged 0 to 5, 6 to 9, 10 to 12, 13 to 16, 17 to 29, 40 to 55, 56 and older, and village fixed effects. For Figure A2, at baseline, households were classified as either being eligible (poor) or non-eligible (not poor) for Progresa transfers according to a household poverty index. This index is a weighted average of household income (excluding children), household size, durables, land and livestock, education, and other physical characteristics of the dwelling. The index is designed to give relatively greater weight to correlates of permanent income rather than current income.