THE IMPACT OF A CONDITIONAL EDUCATION SUBSIDY ON SCHOOL ENROLMENT IN COLOMBIA

Orazio Attanasio
Emla Fitzsimons
Ana Gomez

EDePo

Centre for the Evaluation of Development Policies

THE INSTITUTE FOR FISCAL STUDIES
REPORT SUMMARY FAMILIAS 01

The impact of a conditional education subsidy on school enrolment in Colombia*

Orazio Attanasio

University College London and Institute for Fiscal Studies, London

Emla Fitzsimons

Institute for Fiscal Studies, London

Ana Gomez

Econometría Consultores, Bogotá

March 2005

1. Introduction

Familias en Acción (FA) is a welfare programme that was designed to enrich a number of constituents of human capital. The targeted components include health, nutrition and education, and the targeted population comprises individuals living in the poorest 20% of households in selected areas in rural Colombia. The implementation of the programme began in 2001, and it was fully in operation in all targeted (treatment) communities – 57 in total – by 2003.

In this report, we assess the short-term impact of the programme on enrolment in education.[1] The education 'treatment' is a monthly subsidy that is offered to eligible mothers conditional on their child(ren) attending school.[2] The

*This report is a summary of the education component of the report *Evaluación del Impacto del Programa Familias en Acción – Subsidios Condicionados de la Red de Apoyo Social. Informe del Primer Seguimiento*, Bogotá, carried out by a consortium consisting of Econometría Consultores, the Institute for Fiscal Studies and Sistemas Especializados de Información.

[1]A summary of the effectiveness of the three dimensions of the programme is to be found in Attanasio et al. (2005).

[2]The monthly value of the subsidy is 14,000 pesos for primary school children and 28,000 pesos for secondary school children, as of 2002. This is equivalent to $10 and $20 per month

eligibility requirements include having a household welfare indicator that is below a predefined level, having at least one child between 7 and 17 years old, and living in a treatment area. The effect of the subsidy on school enrolment is estimated through a careful comparison of enrolment rates across individuals in treatment and selected control areas.[3]

Conditional education subsidies decrease the price of education facing agents, and thus their theoretical effect on enrolment is positive, assuming that education is a normal good. Indeed, this theoretical prediction has gained much empirical support. The success of the PROGRESA programme in Mexico at increasing school enrolment (see Behrman, Sengupta and Todd (2001)) spurred many countries effectively to emulate it, particularly across Latin America. Moreover, such success stories are by no means unique to less developed countries (LDCs). The efficacy of the Education Maintenance Allowance at increasing post-compulsory secondary education in England propelled its national rolling-out in 2004 (see Battistin et al. (2004)). Of course, it is important to bear in mind that the deadweight loss associated with such subsidies may be large, if many individuals would have stayed on in school even in the absence of the subsidy. Fine-tuning such policies so that they specifically target marginal decision-makers would perhaps be a more resourceful use of available funds, yet the design and implementation of such an initiative would be challenging.

This report is laid out as follows. In Section 2, we outline the evaluation methodology that underlies the analysis. We discuss the importance of investigating pre-programme differences in school enrolment between treatment and control areas and provide evidence that motivates the estimation of the impact of the programme using pre- and post-programme data on outcomes, in a differences-in-differences framework.[4] Section 3 presents our

---

respectively, and to put it in context, its value is approximately 3% (6%) of household monthly consumption for primary (secondary) school attendance.

[3]As the programme was not randomly assigned, control areas were selected for the purpose of the evaluation so as to be comparable to treatment areas on the basis of a number of pre-programme dimensions. To select control areas, strata were defined according to geographic location and an index of school and healthcare infrastructure availability. Each stratum contained areas in which the programme was and was not to be implemented, and within each, controls were selected so as to be comparable to treatment areas on the basis of population size and an index of quality of life (see Attanasio et al. (2004) for details). The main requirement for eligibility at the area level included access to basic education and healthcare infrastructure, the presence of at least one bank and a population of less than 100,000 individuals.

[4]At the time of writing, two surveys have been conducted in treatment and control areas. The two surveys – baseline and follow-up – were designed so as to measure pre- and post-programme information respectively. However, the programme was phased in across treatment areas, with the baseline survey for some treatment areas concurring with the post-programme period. As such, we only have post-programme data for these treatment areas. We

estimates of the effect of the programme on enrolment in education for different subgroups of individuals, and Section 4 concludes.

2. Evaluating the impact of *Familias en Acción* on school enrolment

2.1 Evaluation methodology

In order to estimate the impact of the subsidy on school enrolment, one would like to observe average enrolment in treatment areas both with and without the programme. The difference between the two would be entirely attributable to the programme, and the parameter of interest, Δ, would be estimated as

$$\Delta = E(Y_{1,A} - Y_{0,A} \mid T = 1) \tag{1}$$

where $Y_{j,k}$ is a 0,1 indicator of school enrolment if undergoing treatment (j=1) or not (j=0) before (k=B) or after (k=A) the programme. T=0 (1) denotes control (treatment) areas and E denotes the expected value.

In the absence of observing enrolment in treatment areas without the programme, the general strategy is to make use of control areas to estimate this counterfactual. One could thus begin by comparing *post-programme* school enrolment across treatment and control areas, conditional on a range of observed characteristics that are likely to affect enrolment rates, X:

$$\Delta = E(Y_{1,A} \mid T = 1, X) - E(Y_{0,A} \mid T = 0, X)\,. \tag{2}$$

Δ yields an unbiased estimate of the programme impact under the assumption that conditional on observed characteristics X, there are no unobserved factors, u, differentially affecting enrolment rates in treatment and control areas. That is, under the assumption

$$E(Y_{1,A} - Y_{0,A} \mid T, X, u) = E(Y_{1,A} - Y_{0,A} \mid T, X)\,, \tag{3}$$
$$\text{i.e. } u \perp T \mid X\,.$$

This implies that the difference estimated in (2), Δ, is entirely attributable to the programme. (3) is of course a potentially strong assumption. No matter how similar treatment and control areas are on the basis of observed characteristics, it is always of concern that there are non-programme-related unobserved differences across areas that may also affect school enrolment. Thus one risks confounding programme effects with non-programme-related unobserved factors that affect enrolment, in which case a comparison of post-programme rates only can be very misleading.

---

return to this point in Section 2. We use both surveys in the analysis to estimate the impact of the programme on school enrolment.

If enrolment rates *before* the programme has started are observed in both treatment and control areas, one can estimate the impact of the programme under less stringent assumptions. The first of these is that there are common time effects between treatment and control areas, i.e.

$$E(Y_{0,A} - Y_{0,B} \mid T = 1) = E(Y_{0,A} - Y_{0,B} \mid T = 0)\,. \tag{4}$$

We want to estimate

$$\begin{aligned} \Delta &= E(Y_{1,A} \mid T = 1, X) - E(Y_{0,A} \mid T = 1, X) \\ &= E(Y_{1,A} \mid T = 1, X) - E(Y_{0,B} \mid T = 1, X) + E(Y_{0,B} \mid T = 1, X) - E(Y_{0,A} \mid T = 1, X). \end{aligned} \tag{5}$$

Using assumption (4),

$$E(Y_{0,B} \mid T = 1, X) - E(Y_{0,A} \mid T = 1, X) = E(Y_{0,B} \mid T = 0, X) - E(Y_{0,A} \mid T = 0, X). \tag{6}$$

So

$$\Delta = \underbrace{\left[E(Y_{1,A} \mid T = 1, X) - E(Y_{0,A} \mid T = 0, X)\right]}_{A'} - \underbrace{\left[E(Y_{0,B} \mid T = 1, X) - E(Y_{0,B} \mid T = 0, X)\right]}_{B'}. \tag{7}$$

The second assumption underlying this estimator is that any difference in enrolment between treatment and control areas due to unobserved factors is fixed over time. One can think of this difference as being represented by B′ in (7).[5] By netting this out from A′, one obtains the effect of the programme on school enrolment.

It is necessary for both of these assumptions to hold in order for the differences-in-differences estimation of the programme impact to be unbiased. Assumption (4) is likely to be violated if individuals living in treatment areas change enrolment *in anticipation of* the programme. This would mean that $E(Y_{0,B} \mid T = 1, X)$ is a biased estimate of enrolment in treatment areas in the absence of the programme. Even if the effect of unobserved factors on enrolment is fixed over time (i.e. the second assumption holds), B′ in (7) confounds both the effects of unobserved variables and programme (anticipation) effects. (7) would thus estimate the change in enrolment after the programme is implemented (the parameter of interest), net of the change that was induced by anticipation of the programme.

Pre-programme outcomes cannot in general be considered to be uncontaminated by programmes, and indeed are likely to be affected by them if they relate to a period after the programme has been announced. For this

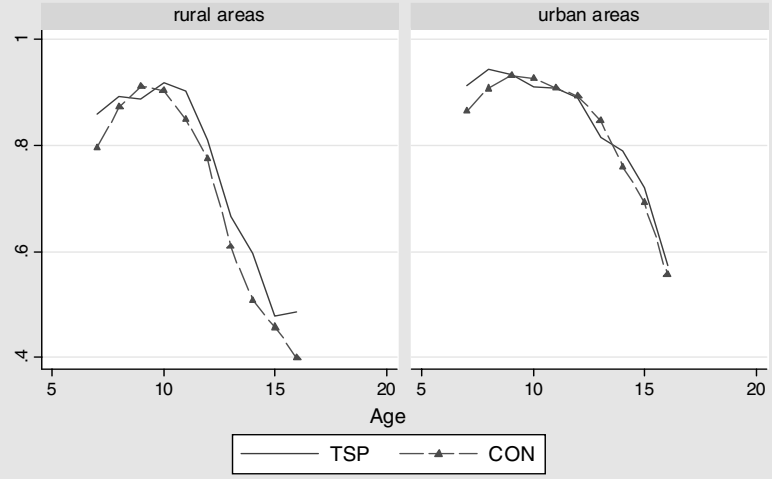
[5]One can think of B′ as representing pre-existing and persistent differences in enrolment between treatment and control areas.

reason, it is important to investigate the existence of anticipation effects in pre-programme choices and to obtain measures of pre-programme outcomes that are unaffected by programme anticipation.

2.2 School enrolment in treatment and control areas one and two years prior to the programme

In this section, we analyse school enrolment rates in rural Colombia *before* the FA programme had started. In light of the discussion above, ideally one would like access to at least two pre-programme periods, in order to discern whether differences between treatment and control areas, if they exist, are due to anticipation effects and/or fundamental factors that vary across areas. Indeed, this is particularly important in Colombia, because for half of our treatment areas (TCP – 'tratamiento con pago' or 'treatment with payment'), the programme had already begun by the time the baseline data were collected. This exacerbates concerns about anticipation effects of the programme for the remaining half (TSP – 'tratamiento sin pago' or 'treatment without payment').[6]

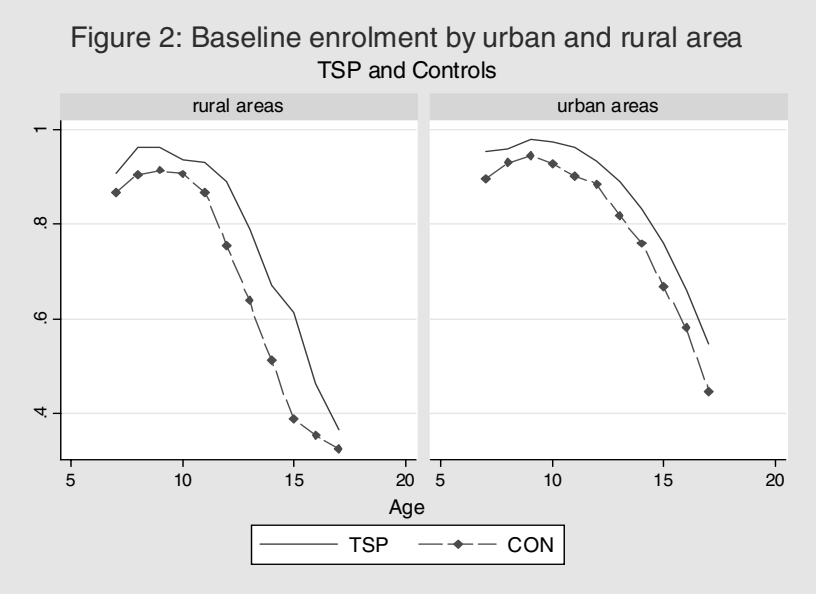
Figure 1: Pre-baseline enrolment by urban and rural area
TSP and Controls



Foreseeing this and its potential importance for the evaluation, contemporary and retrospective data on school enrolment were collected at the baseline (henceforth 'baseline' and 'pre-baseline' data, for one and two years prior to the programme, respectively) so as to obtain relatively clean pre-programme

[6]In addition, the programme had been widely publicised in TSP areas and registration had already commenced. Anticipation effects are less likely to be evident in TCP areas. Even though the programme had been publicised by the pre-baseline period, it had not yet been implemented in any other areas. One potential worry is that anticipation effects may *appear* to exist if TCP individuals misreported pre-baseline education levels, in the belief that it would increase their chances of programme entitlement. This would result in a biased estimate of the effect of the programme on TCP areas in the baseline, were we to estimate the effect using differences-in-differences.

measures, at least for individuals living in TSP areas. A cursory look at these TSP pre-programme data suggests that anticipation effects are large. Figures 1 and 2 show school enrolment in TSP and control areas for ages 7–16 and ages 7–17 in the pre-baseline and baseline periods respectively, by urban and rural areas. This graphical inspection points towards anticipation effects of the programme: in the pre-baseline period, school enrolment rates are very similar in TSP and control areas (see Figure 1). However, by the time the registration for the programme is underway in TSP areas (and the programme is already operative in TCP areas), school enrolment in TSP areas is higher than in control areas (see Figure 2).



These differences are borne out after controlling for a set of individual, household and area characteristics. Table 1 shows the differences in enrolment between TSP and control areas in the baseline and pre-baseline periods, separately for both 8- to 13- and 14- to 17-year-olds.[7] It also shows the anticipation effect of the programme, i.e. the difference between baseline and pre-baseline enrolment.

Table 1 provides strong evidence that baseline differences in enrolment between TSP and control areas are unlikely to be representative of differences over time in the absence of the programme: enrolment in education is significantly higher in TSP than in control areas at the baseline, particularly for the older age group. The amount of this difference that is due to anticipation effects may be obtained by netting out differences in pre-baseline enrolment, under the assumption that in this earlier period there was no effect of the programme in TSP areas. Column (3) shows that anticipation effects are large, especially for the older age group.

[7]These age groups are chosen because from 14 onwards, school drop-out rates increase sharply.

Table 1: Pre-programme differences in enrolment between TSP and control areas, and programme anticipation

	Pre-programme differences		Anticipation
	Baseline enrolment	**Pre-baseline enrolment**	**Baseline – Pre-baseline enrolment**
	(1)	**(2)**	**(3)**
8- to 13-year-olds			
Urban	0.0388**	0.0176	0.0408**
	(0.0072)	(0.0151)	(0.0078)
Rural	0.0497**	0.0336**	0.0363*
	(0.0151)	(0.0116)	(0.0165)
14- to 17-year-olds			
Urban	0.1024**	0.0251	0.0692**
	(0.0240)	(0.0208)	(0.0172)
Rural	0.1746**	0.0559	0.1019**
	(0.0272)	(0.0328)	(0.0207)

Notes: Figures refer to differences between TSP and control areas, as estimated using a parametric specification controlling for the variables listed in the Appendix. Standard errors, adjusted for clustering at the municipality level, are given in parentheses. * denotes statistical significance at the 1–5% level and ** denotes statistical significance at the 1% level or less.

This analysis points to the importance of using pre-baseline enrolment in the differences-in-differences estimation. If one were instead to proxy pre-existing differences using baseline enrolment rates, the effect of the programme on enrolment in education would likely be remarkably underestimated.

3. Estimation and results

3.1 Estimation

Programme impacts can be estimated either parametrically or non-parametrically. As is well known, the assumption underlying parametric estimation – that the effect is linear in observable characteristics – can be restrictive, particularly when invoked to extrapolate beyond the region of 'common support'. An alternative is to estimate the effect non-parametrically using matching, which does not impose linearity in estimating the parameter of interest and for this reason is generally considered to be more robust to misspecification.

On balance, we estimate the impact of the programme on school enrolment using a linear parametric regression framework. First, common support problems are practically non-existent within each of our samples, and thus we

are safeguarded against misleading inferences that may arise from extrapolation beyond the common support.[8] Second, the imposition of linearity seems to provide a very reasonable representation of the effects – the effects that are estimated using matching are very similar.[9] As the parametric specification leads to a large gain in efficiency compared with matching, we focus on these results throughout. It is of course important to bear in mind that whilst efficiency gains are achieved through parameterisation, the unbiasedness of both estimators relies on the differences-in-differences assumptions discussed in Section 2.

The impact of the programme on school enrolment is estimated across TSP and TCP areas using data on enrolment in the pre-baseline (t=0), baseline (t=1) and follow-up (t=2) periods.[10]

$$Y_t = \alpha_1 + \alpha_2 d_{[TSP,2],[TCP,2,1]} + \alpha_3 d_{[TSP,1]} + \alpha_4 d_1 + \alpha_5 d_2 + \alpha_6 X + \varepsilon_t \qquad (8)$$

where

Y_t	= 1	if attending school in period t, for t=0,1,2
	= 0	otherwise
$d_{[TSP,2],[TCP,2,1]}$	= 1	if (*TSP*, t=2) or (*TCP*, t=2) or (*TCP*, t=1)
	= 0	otherwise
$d_{[TSP,1]}$	= 1	if (*TSP*, t=1)
	= 0	otherwise
d_j	= 1	if t=j, for j=1,2
	= 0	otherwise

X represents the control variables listed in the Appendix.

3.2 Results

3.2.1 Results by urban/rural area

Table 2 shows the estimates of the effect of the programme on school enrolment of 8- to 13- and 14- to 17-year-olds, as estimated using equation (8). The subsidy has the largest impact on the enrolment of 14- to 17-year-olds, leading to an increase in school enrolment of approximately 5½ percentage

[8]We restrict the analysis to individuals who fall within the common support by dropping pilot individuals who have no counterpart in a control area. We estimate the propensity score and drop pilots whose propensity score is above the maximum observed propensity score in control areas. In no cases do we lose more than 5% of the sample.

[9]All corresponding results from the matching estimation are shown in the Appendix.

[10]This assumes that pre-baseline enrolment rates for TCP areas are unrelated to the programme. Note that the omission of TCP areas from the analysis leads to very similar results, which are available from the authors upon request.

points in both urban and rural areas. The effects for 8- to 13-year-old children are lower, at around 3 percentage points in rural areas, and not statistically different from 0 in urban areas. Given the relatively high pre-programme enrolment rates of 8- to 13-year-olds compared with 14- to 17-year-olds, the relatively low impact of the programme on school enrolment rates in this age group is not surprising.

Table 2: Impact of the programme on school enrolment – treatment versus control areas

	All individuals		
	Impact	**Enrolment without subsidy**	**Enrolment with subsidy**
Urban			
8- to 13-year-olds	0.0131 (0.0071)	94.1%	95.4%
14- to 17-year-olds	0.0533** (0.0131)	72.0%	77.3%
Rural			
8- to 13-year-olds	0.0295* (0.0117)	89.2%	92.2%
14- to 17-year-olds	0.0592** (0.0221)	54.4%	60.3%

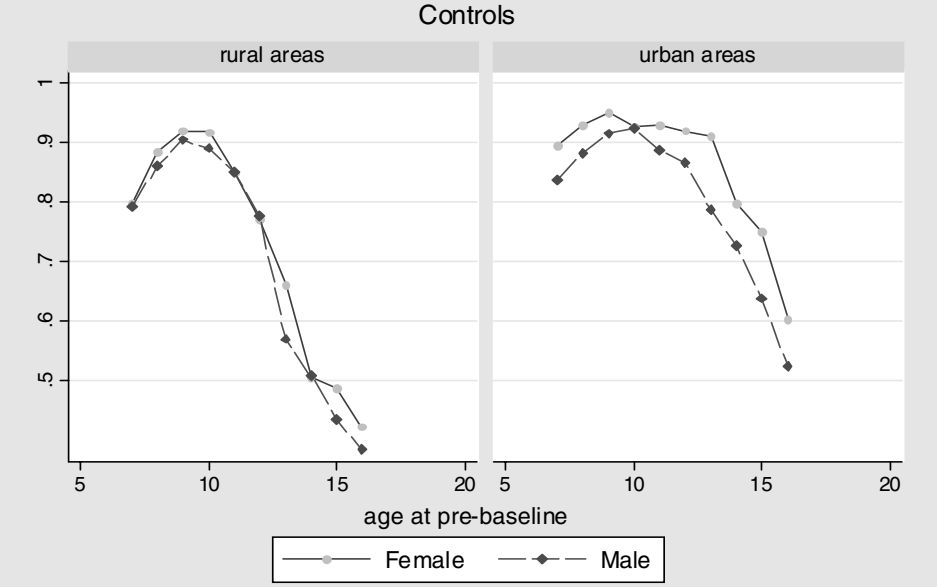
Notes: Impact is estimated parametrically using equation (8), controlling for the variables listed in the Appendix. Treatment areas include both TSP and TCP. Standard errors, adjusted for clustering at the municipality level, are given in parentheses. * denotes statistical significance at the 1–5% level and ** denotes statistical significance at the 1% level or less.

Table 2 also shows our best estimate of the percentage of children attending school had the programme not been in operation. The counterfactual enrolment rates are estimated by subtracting the programme impact from the observed follow-up enrolment rates in treatment areas.

3.2.2 Results by urban/rural area and males/females

In this section, we examine whether the programme has had differential effects by gender. A priori, there is no reason to believe that the programme will be equally effective for males and females, particularly if pre-programme enrolment varies by gender. To examine pre-existing differences in enrolment between males and females, we compare pre-baseline enrolment rates by gender in control areas, separately for urban and rural areas. Figure 3 shows that whilst female and male enrolment rates are very similar in rural control areas below the age of 13, from then on female enrolment is generally higher

Figure 3: Pre-baseline enrolment by gender and urban-rural area



than that of males. In urban areas, female enrolment is substantially higher than that of males at all ages apart from age 10.[11]

We turn now to the causal analysis of the effect of the programme on school enrolment by gender. Table 3 shows the results of the differences-in-differences estimation, separately by gender. The programme has a positive and significant effect on the enrolment of all males. For the younger age group, the effect is between 2 and 3 percentage points, whilst for older males, it is around 7 percentage points. For females, on the other hand, significant effects are only observed for 14- to 17-year-olds in urban areas and for 8- to 13-year-olds in rural areas.

In terms of closing the gender gap, for younger ages, the post-programme difference in enrolment between females and males is 1½ percentage points in urban areas. In the absence of the programme, we estimate that this gap would be 3½ percentage points. In rural areas, the gender enrolment gap for the younger ages remains more or less the same as before the programme. For the older age group, female enrolment continues to be higher, by around 6½ percentage points in urban areas. In the absence of the programme, we estimate that the discrepancy would be 10 percentage points. The gap is 12 percentage points for the older age group in rural areas, and it is estimated to be over 14½ percentage points without the programme. Thus whilst the programme has

[11]This is in fact contrary to what is generally observed in many countries, in which male enrolment tends to be higher than female enrolment (for example, in Brazil, Pakistan, Turkey and Mexico). Indeed, many programmes, such as PROGRESA in Mexico, are designed with the aim of increasing the enrolment rates of females so that they catch up with those of males.

Table 3: Impact of the programme on school enrolment, by gender – treatment versus control areas

	Males			Females		
	Impact	**Enrolment without subsidy**	**Enrolment with subsidy**	**Impact**	**Enrolment without subsidy**	**Enrolment with subsidy**
Urban						
8–13	0.0210* (0.0106)	92.7%	94.8%	0.0052 (0.0070)	96.1%	96.2%
14–17	0.0692** (0.0163)	67.3%	74.2%	0.0342* (0.0148)	77.4%	80.8%
Rural						
8–13	0.0314* (0.0140)	87.4%	90.5%	0.0270* (0.0125)	91.2%	93.9%
14–17	0.0702** (0.0255)	47.9%	54.9%	0.0424 (0.0263)	62.6%	66.9%

Notes: Impact is estimated parametrically using equation (8), controlling for the variables listed in the Appendix. Treatment areas include both TSP and TCP. Standard errors, adjusted for clustering at the municipality level, are given in parentheses. * denotes statistical significance at the 1–5% level and ** denotes statistical significance at the 1% level or less.

reduced the gender disparity in school enrolment, it remains the case that female school enrolment exceeds that of males.

4. Conclusion

In this report, we have considered the impact on school enrolment of a large welfare programme in Colombia, *Familias en Acción*. Our best estimate of the effect takes into account pre-existing differences between treatment and control areas, is careful not to confound anticipation effects of the programme, and controls for a wide range of observable characteristics that are likely to affect enrolment. We find that the programme has been effective at increasing enrolment, particularly amongst 14- to 17-year-olds, in both urban and rural areas. Further, males have generally benefited more than females from the programme. Younger age groups have seen very modest increases in enrolment, but this is largely due to the fact that initial attendance rates amongst these individuals are already relatively high. This is also the case for females relative to males. This is suggestive of a substantial deadweight loss of the programme, and it is worth considering that educational resources might be more effectively redirected at other interventions, such as improving school quality, rather than at increasing school enrolment of young children.

Appendix

A.1 Control variables

Gender
Age of household head
Age of spouse of household head
Dummy variable for single head of household
Dummy variables for education of household head and spouse
Dummy variable for serious illness before age 1
Dummy variables indicating type of health insurance of the head of the household – subsidised, unsubsidised or none
Dummy variables indicating type of health insurance of the spouse of the head of the household – subsidised, unsubsidised or none
Dummy variable for whether live in a house
Dummy variable for whether own or rent house
Dummy variables for material of house walls (brick, wood, cardboard, no walls)
Dummy variable for existence of gas pipe
Dummy variable for existence of water pipe
Dummy variable for whether household has sewage system
Dummy variable for whether household has rubbish collection
Dummy variable for whether household has or has access to a telephone
Dummy variable for whether toilet is connected to sewage system / septic tank or not connected / no toilet
Dummy variable for whether at least one individual in the household has suffered from extreme violence in 2000, 2001 or 2002
Dummy variables for regions
Altitude
Number of urban public schools in the municipality
Number of rural public schools in the municipality
Number of students per teacher in the municipality
Classroom square metres per student in the municipality
Number of public hospitals in the municipality
Index of quality of life in the municipality in 1993
Population in the urban part of the municipality in 2002
Population in the rural part of the municipality in 2002
Number of public puestos (small healthcare providers) in the municipality
Number of pharmacies in the municipality
Proportion of households in the municipality with piped water
Proportion of households in the municipality with sewage system
Dummy variable for whether any IPS (healthcare provider) in the municipality suffered task-force desertion

A.2 Matching estimates

Table A1: Impact of the programme on school enrolment – treatment versus control areas

	Matching
Urban	
8- to 13-year-olds	0.0207 (0.0194)
14- to 17-year-olds	0.0634* (0.0261)
Rural	
8- to 13-year-olds	0.0245 (0.0263)
14- to 17-year-olds	0.0755* (0.0223)

Notes: Treatment areas include both TSP and TCP. Bootstrapped standard errors are given in parentheses. * denotes statistical significance at the 5% level or less.

Table A2: Impact of the programme on school enrolment, by gender – treatment versus control areas

	Matching	
	Males	**Females**
Urban		
8- to 13-year-olds	0.0378 (0.0311)	0.0063 (0.0317)
14- to 17-year-olds	0.0802* (0.0400)	0.0353 (0.0382)
Rural		
8- to 13-year-olds	0.0246 (0.0198)	0.0237 (0.0370)
14- to 17-year-olds	0.0703 (0.0451)	0.0689 (0.0444)

Notes: Treatment areas include both TSP and TCP. Bootstrapped standard errors are given in parentheses. * denotes statistical significance at the 5% level or less.

References

Attanasio, O. et al. (2004), *Baseline Report on the Evaluation of* Familias en Acción, http://www.ifs.org.uk/edepo/wps/familias_accion.pdf.

Attanasio, O., E. Battistin, E. Fitzsimons, A. Mesnard and M. Vera-Hernández (2005), *How Effective Are Conditional Cash Transfers? Evidence from Colombia*, IFS Briefing Note no. 54, http://www.ifs.org.uk/bns/bn54.pdf.

Battistin, E. et al. (2004), *The Evaluation of Education Maintenance Allowance Pilots: Three Years' Evidence. A Quantitative Evaluation*, Department for Education and Skills, Research Report no. RR499, http://www.dfes.gov.uk/research/data/uploadfiles/RR499.pdf.

Behrman, J., P. Sengupta and P. Todd (2001), *Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment*, University of Pennsylvania and the International Food Policy Research Institute, Washington, DC.