Estimating the effect of adolescent fertility on educational attainment in Cape Town using a propensity score weighted regression.

Vimal Ranchhod *

David Lam †

Murray Leibbrandt ‡

Leticia Marteleto

§

¶

March 31, 2009

Abstract

We estimate the effect of a teenage birth on the educational attainment of young mothers in Cape Town, South Africa. Longitudinal and retrospective data on youth from

---

*Corresponding author: vimal.ranchhod@gmail.com

†Lam and Marteleto are at the Univ. of Michigan

‡Leibbrandt and Ranchhod are at the University of Cape Town.

§**Preliminary draft: Please do not cite**

¶Support for this research was provided by the U.S. National Institute of Child Health and Human Development (Grants R01HD39788 and R01HD045581) and the Fogarty International Center of the U.S. National Institutes of Health (D43TW000657).

the CAPS dataset is used. We control for a number of early life and pre-fertility characteristics. We also reweight our data using a propensity score matching process to generate an appropriate counterfactual group. Accounting for respondent characteristics reduces estimates of the effect of a teen birth on dropping out of school, successfully completing secondary school, and years of schooling attained to between 24.2% and 43.4% of the unconditional mean difference, but they remain large and significant. We find some support for the hypothesis that there is heterogeneity in the effect of a teen birth, depending on the actual age of the first birth.

1 Introduction

What is the effect of adolescent fertility on educational outcomes in South Africa? By African standards, South Africa's total fertility rate (TFR) is relatively low. Using 2001 Census data, Moultrie and Dorrington (2004) estimate it to be 2.8 births per woman. Recent declines in the TFR have been driven by declines in fertility at older ages, while adolescent fertility remains relatively high. Thus, South Africa's adolescent fertility rate is only the fifteenth lowest in the continent. (United Nations Population Division 2003). In our dataset, approximately 25% of young African and Coloured women have experienced a teenage birth. The question of what effect, if any, this early life fertility has on the educational outcomes of youth is potentially important in understanding employment patterns, poverty dynamics and other quality of life measures that are affected by educational attainment.

Several researchers have investigated the correlations between education, adolescent sexual initiation and childbearing in developing countries (Bledsoe et al. 1999; Lloyd 2005). The general finding is that educational attainment and early childbearing are negatively correlated (Gupta and Leite 1999, Lloyd and Mensch 2008). In the South African context, Kaufman et al.(2001) find that while young girls are likely to leave school after a birth, many return subsequently to complete their schooling. This return is correlated with familial support and paternal recognition of the child. Madhavan and Thomas (2005) show a similar finding, and emphasize the importance of flexible child care options in successful completion of school. Marteleto et al (2008) use longitudinal data to investigate how household and individual characteristics impact on sexual debut, pregnancy and school dropout. They emphasize the importance of young adults' skills and knowledge in understanding the various inter-relationships in transitioning into adulthood. Grant and Hallman (2008) find that prior scholastic performance is a significant predictor of both adolescent pregnancy and the likelihood of dropping out of school after a pregnancy.

While a considerable body of research exists for South Africa, there is a lack of empirical research that makes a serious attempt at identifying the causal impact of fertility on education. This is the primary contribution of this paper. For methodological guidance, we turn to the corresponding literature from the US, where the causal question has been pursued for several years. Econometrically, the problem is one of endogeneity due to selection into 'treatment'. Girls who experience teen births tend to have poorer measures of socio-economic status and scholastic performance even prior to the birth. This is likely to extend to unobservable characteristics as well. Thus, the girls who did not experience a teen birth would, in expectation, attain higher levels of education than the young mothers do, even in the absence of the birth in the group of young mothers. This makes estimating the counterfactual educational attainment problematic.

The literature on the effects of teen births in the US is vast. An excellent review can be found in Hoffman (1998). Some studies have attempted to account for measures of family background and parental involvement in the girl's education (e.g. Lee et al, 1994 and Hernstein and Murray, 1994). An alternative has been to use a siblings fixed effects estimation method, as in Geronimus and Korenman (1993). Some have used almost-natural experiments such as miscarriages in estimation, such as Hotz et al (1997) and Hotz et al (1999). An alternative has been to use the age of menarche as an instrument in an instrumental variables method, such as Ribar (1994) and Klepinger et al (1995). Levine and Painter (forthcoming), make use of a within-school propensity score matching estimator. Broadly speaking, the research suggests that a large proportion of the observed educational differential between teen mothers and non-teen mothers is a function of other environmental features, although there remains considerable debate as to the magnitude of this proportion.

We make use of data from the Cape Area Panel Study (CAPS) to investigate this question. We employ a propensity score matching method to reweight observations in our regression. We further estimate a separate treatment effect for young mothers who experienced their first birth at ages 16, 17, 18 and 19. Our findings are similar to those in the US. Teen mothers attain fewer years of schooling on average, but they tend to come from disadvantaged backgrounds. Accounting for this reduces the estimated educational cost of adolescent motherhood by more than fifty percent for each outcome measure. The matching process brings all the estimates closer to zero. However, these smaller negative effects remain statistically significant in all cases. By age 22, the estimated effect of a teen birth on the probability of graduating from secondary school is -7.54%. Our findings suggest that in addition to their relatively disadvantaged background, adolescent childbearing further restricts the attainment of young mothers.

2 Data

The data for this study comes from Waves 1 to 4 of the Cape Area Panel study. CAPS is a longitudinal survey of youth in the Cape Town metropolitan area. The first wave was conducted in 2002, with a sample of about 4800 respondents aged 14 to 22. Wave 2a was conducted in 2003, wave 2b in 2004, wave 3 in 2005 and wave 4 in 2006. Details are contained in Lam et al (2006). It has detailed information about early life environment, schooling progress, age of menarche, and various questions about the circumstances in which the girls experienced their sexual debut. Topics such as employment, school and neighborhood characteristics and data on other members of the household are also captured. CAPS also includes a life-history calendar that provides retrospective information on schooling enrolment and progress, timing of pregnancies, timing of births, and parental co-residency.

The sample design was a two-stage probability sample of households, with an over-sampling of white and African households. To take this into account, all results are weighted using the sampling weights from wave 1. For our study, we exclude all males and white females from the analysis. White females have very low levels of observed fertility in our sample, and are very different from the African and coloured subpopulation groups on a number of socio-economic dimensions. Including them as potential counterfactual observations in our analysis would likely confound our results with an upward bias. Of the 2296 remaining observations from wave 1, we have 1932 observations in either wave 3 or wave 4. This represents an attrition rate of approximately 16%. We make no direct corrections for attrition in the sample.

The attrition rate is not particularly problematic for us for two reasons. First, it is relatively small. Second, our entire analysis is based on a mixture of early life characteristics and the life-calendar data. All of the early life characteristics were obtained from the wave 1 data, prior to any attrition. These include the girls' parents' education levels, whether there was a someone in the household growing up with a drinking problem, with a drug addiction problem and whether the household has five or more books.[1] From the life-calendar, we obtain information at each year of age about the respondent's grade attainment, her enrolment status, her pregnancy status, whether she has had sex or not and whether she had a birth or not. Thus a number of girls who were not observed in later waves are still included in the analysis.

We do, however, use the additional waves to supplement the calendar. For example, consider a girl aged 16 in wave 1 who has not had a teen birth yet, but that this has changed by wave 3. Thus, her life calendar by wave 1 is only completed up to age 16, and by including the information at age 19 (in wave 3), we get more data to estimate our parameters. In addition, we use a question that describes the first sexual experience, namely whether she was 'willing', 'persuaded', 'tricked' or 'forced'. We also use a variable that indicates whether contraception was used during her first sexual experience. These questions were only asked if the girl had already had her sexual debut. Information from subsequent waves was used only if this information was not available from an earlier wave.

[1] The 'books' question is a contemporaneous question at the time of the wave 1 interview.

As outcome variables, we use a number of different outcomes:

- *educ18, educ20* and *educ22* are the number of years of primary and secondary schooling attained at ages 18, 20 and 22 respectively. It is bounded above at 12.[2]
- *matric20* and *matric22* are indicator variables that indicate whether the person has successfully completed high school or not, which is equivalent to twelve years of schooling.
- *dropout* is an indicator variable that takes a value of one if at any point a respondent was not enrolled in school prior to successfully completing secondary school. In this context, this variable might be better named as 'interrupt', since 56.7% of girls who do drop out subsequently return to school.

There remains a truncation problem in our data, since we do not observe all the girls up to at least age 20. While these observations are simply dropped from the estimation sample for the outcomes corresponding to ages 20 or greater, they might be included in the analysis for 'educ18' and are certainly included in the 'dropout' outcome. We chose not to exclude these girls as we would then lose 647 observations in our sample. This truncation problem makes the separation into 'treated' and 'untreated' groups problematic. Of the respondents whose last age we observe as 17, those with a birth already will always be teen mothers, but those who are not yet mothers might still have a birth at 18 or 19. This might induce some downward bias into the estimates corresponding to the aforementioned variables.

In column 2 of Table 1 we present the means of the variables used in our analysis. Racially, 36.8% of the sample is African, with the remainder being coloured. The proportion that has experienced a dropout is very high, at 72.4%. By age 22, only 50.9% have completed high school, although the fact that this is considerably larger than 27.6% attests to the observed pattern of dropping out often being temporary rather than terminal. Educational attainment

[2]We do not use completed years of secondary schooling, nor do we include any college level schooling. This is due to the data which is restricted to a relatively young population, as well as the frequency with which people return to school at relatively late ages.

between ages 18 and 20 increases from 9.961 to 10.45, an increase of about 0.5 years, and continues to increase by a smaller amount between ages 20 and 22. Roughly 80% have five or more books, 20% had someone with an alcohol problem in their household growing up, and 8.3% grew up with someone who had a drug addiction problem. Over a third of the girls do not have a valid response for their father's education, while only 11.7% do not have a valid response for their mother's education.[3] The mean levels of parental education from the sample is biased downwards, as those with missing information were assigned a value of zero. The mean mother's and father's education, conditional on a valid response, is 8.35 and 8.53 years respectively. This is relatively low, but not unusual for older African and coloured groups of that generation. The girls lived a large proportion of their early years with their mothers, and a smaller proportion with their fathers. Grade progression from ages 8 to 14 is fairly high at 0.923, although if interpreted as a probability, a significant fraction of the girls will repeat a grade during primary school.

The mean age of menarche is 13.298, while the proportion who have had sex is 71.4%. Of these, the mean age of sexual debut is 17.04. Of interest is the proportion who used contraception during their first sexual experience, at only 59%. The majority of respondents were willing or persuaded, although 1.6% report being forced.

We then compare the means of these variables for the group of teen mothers and the group of non-teen mothers. The difference in means and corresponding t-statistic are also presented. The outcome measures differ by a large amount, always adversely for the teen mothers, and have highly significant t-statistics. The groups also differ in their early childhood characteristics, the educational attainment of their parents and their age of sexual debut. There is a very large and highly significant difference in the proportion that used contraception during their first sexual experience. All of these suggests that the girls who have a teen birth are indeed quite different from those who do not.

[3] The survey captured parental educational attainment if known by the respondents, regardless of co-residency or not.

3 Empirical methodology

Our analysis consists of a combination of propensity score matching and weighted OLS regressions. In the first part of our analysis, the coefficient of interest pertains to a 'teen birth' variable. This takes on a value of 1 if the girl is observed to have had a teen birth, and a value of zero otherwise. This definition is applied to girls whom we observe only up to some age less than 20 as well.

We first estimate the probability that a girl has had a teen birth using probit models. That is, we estimate the propensity score of 'treatment' following Rosenbaum and Rubin (1983). We include as regressors the variables discussed above, as well as a race dummy variable, and separate indicator variables for whether the father's or mother's education is missing. All regressors are entered linearly, and we restrict the sample to girls who had their first sexual experience before the age of 20. The prediction in only done for those observations in the estimation sample.

Once we have the propensity score, we perform a kernel matching procedure on the girls. We imposed a common support condition, which drops treatment observations whose propensity score is higher than the maximum or less than the minimum of the controls. We use an Epanechnikov kernel with a bandwidth of 0.06. This yields a set of matching weights for the control group[4], which allows us to obtain an appropriate set of counterfactual girls. Intuitively, the procedure selects girls who did not have a birth but look like the set of girls who did have a birth (in terms of their propensity score), and gives them a greater weighting.

These weights are then used in our regressions. We apply a composite weight which equals the product of the matching weights and the sampling weights. All the covariates from the probit regression as well as the 'teen birth' variable are included in the final specification. We also maintain the same sample restriction. The relevant coefficient represents our estimate of the effect of a teen birth on the various measures of educational attainment. We estimate a separate regression for each outcome measure. Where the dependent variable is an

[4]The weight for the treated group is set to 1.

indicator variable, these are effectively linear probability models, and the coefficients should be interpreted as probabilities. In order for this method to provide unbiased estimates, we need to believe that conditional on the sample restriction, common support condition and matching weights, the regressors are not correlated with the error term. If this assumption is satisfied, then our estimate represents an 'average treatment effect on the treated' (ATT).

In the second part of the analysis, we investigate the effect of a teen birth at particular ages. We perform essentially the same analysis, but change the way we define the 'treatment' group. We separately investigate the effect of a first birth at ages 16, 17, 18 and 19 respectively. We use the same outcome measures, and define the potential counterfactual group in a corresponding fashion. For example, where the treatment is defined as a first birth at age 16, the sample is restricted to girls who were sexually active by age 16 and who had not yet had a birth by age 15. The counterfactual group thus potentially includes girls who subsequently have their first birth at age 17. This is desirable because some first time mothers at age 16, had they not experienced that birth, would have had their first birth at age 17. Performing the analysis for these separately allows us to observe potential heterogeneity in the effects of births at various ages. *A priori*, it seems reasonable to expect that the effect of a birth at age 19 differs from that of a birth at age 16, particularly when the outcomes we are concerned with is educational attainment.

The results from the probits on the various treatment variables are presented in Table 2. The coefficients and their magnitudes are not of particular interest. In general across the different models, only the race and contraception variables seem to be significant. For the teen birth variable, we also observe that age of sexual debut and parental characteristics affect the probability of a teen birth. Note that the sample sizes are considerably smaller, due to a combination of missing data and the sample restrictions.

Table 3 shows the effect of the matching and re-weighting on the same set of variables from Table 1.[5] The outcome measures are not of interest here. What we care about is whether the treatment and counterfactual samples are balanced in terms of their covariates. It is striking

[5]We do not present similar tables for the 'birth at age 16' etc for brevity.

that most of these t-statistics are much smaller and none of the differences are significant at the 5% level of significance. In addition, the difference in means in generally much smaller in absolute value.

4 Results

4.1 Teen Births

The results for the teen birth analysis are presented in Table 4. For each outcome variable, we present the coefficient from a regression with no covariates. This is analogous to the difference in means in Table 1. We then show results where we include the household and socio-economic covariates, but do not include those related to sexual activity or contraceptive usage. We next include the sample restriction and the remaining covariates. Finally, we incorporate the weights from the matching process in the fourth specification.

Across the four specifications, the estimated effect on high school graduation by age 20 decreases from -0.303 to -0.208 to -0.125 and finally to -0.10. All of these are significant at the 1% level. By age 22, the corresponding estimates on high school graduation are similar in magnitude and significance, although the propensity score matching estimate is smaller and only significant at the 10% level. The estimated effect on dropping out of school also decreases from 0.19 to 0.147 to 0.102 to 0.0825. We find that teen birth significantly affects the probability of graduating from high school and of dropping out of school. That said, the decrease in the magnitude of the coefficients is also important, and highlights the importance of controlling for additional characteristics which correlate positively with adolescent childbearing and negatively with school performance.

For the educational attainment at ages 18, 20 and 22, we observe a similar pattern. At age 18, the teen mothers have 0.93 fewer years of schooling on average, and the difference is highly significant. This estimate is reduced to 0.62 and then to 0.382 when we control for other factors, and is reduced further to 0.348 using the matching weights. The estimated decreases in educational attainment across specifications by age 20 are 1.331, 0.921, 0.568 and 0.51 years respectively. All of these estimates are significant at the 5% level. By age 22, the corresponding estimates are 1.13, 0.801, 0.358 and 0.274 fewer years of schooling. The first three estimates are significant at the 1% level, while the fourth is marginally significant with a t-statistic of 1.73.

Another interesting pattern is observed in terms of educational attainment measured in years of schooling. The gap between the mothers and non-mothers increases between ages 18 and 20, but decreases considerably between ages 20 and 22. This is true in all four specifications. This suggests that the teen mothers do experience some element of 'catching up' in terms of secondary schooling.[6]

4.2 Births at particular ages

Table 5 presents results from the second part of our analysis. We estimate the effect of a first birth at a specific age on the various outcome measures. This allows us to explore potential heterogeneity in terms of the effects of teen births. We only present results from the propensity score weighted regressions. On aggregate, the evidence is mixed. The sign of the estimates are consistent with those discussed above, but significance is weaker. This might be partly a result of smaller estimation samples and subsequent lack of power.[7]

Having a birth age 16 seems to have only a modest effect on educational attainment, and these mothers appear to catch up to their peers between ages 20 and 22. A similar statement can be made for first time mothers at age 17. At age 22, the estimated effect on high school graduation is -0.0428 and is not significant, and the difference in years of schooling is very small at 0.0483 fewer years. A first birth at age 18 seems to have a significant and

[6]Of course, this does not explore potential differences in college attendance, nor differences in accumulated work experience.

[7]There are 85, 97, 130 & 113 'treated' observations in the estimation samples for a birth at age 16, 17, 18 and 19 respectively.

negative effect. These mothers are less likely to have completed high school by age 22 by 12.7 percentage points. They also have 0.402 fewer years of schooling on average by age 22. Instead of 'catching up', they seem to experience a 'falling behind' relative to their peers in terms of completing secondary school. For first time mothers at age 19, the pattern is qualitatively similar to first time mothers at age 18. However, the coefficient on dropping out of school is much larger and significant.

The results suggest that there is some heterogeneity in the effects of a birth at various ages within the teenage years. Whereas the younger teen mothers seem to catch up with their peers as they age, the older ones seem more likely to drop out of school, thus falling behind. This might reflect different labor market opportunities for 19 year olds as compared to 16 year olds, where the 19 year olds also probably have more education at the time of their first birth as well. At the same time, the first time mothers at ages 18 and 19 are potential counterfactual observations for the first time mothers at ages 16 and 17. Indeed, the matching procedure is likely to enhance the weight of these counterfactual observations. The estimated heterogeneity might thus reflect that girls who experience a first birth at age 16, had they avoided that birth, were likely to experience a birth at age 17 or 18, and would therefore still have experienced a negative educational effect.

5 Discussion

We investigate the causal effects of adolescent fertility on educational outcomes in Cape Town, South Africa. We make use of a rich dataset that includes several variables on early life socio-economic characteristics, grade progression in the pre-pubescent years, schooling enrolment and educational attainment. We also use information about contraceptive usage on sexual debut, age of sexual debut and a description of the girl's willingness to engage in her initial sexual experience. We allow for heterogeneity both in the timing of the first teen birth, as well as the possibility that educational attainment is affected differently at different ages in the life cycle. We employ propensity score matching methods to reweight our sample. This allows us to obtain a more appropriate counterfactual group which is used to estimate the average treatment effect on the treated.

Our findings are somewhat similar to those obtained in the US. Teen mothers in South Africa do exhibit significantly lower levels of education, whether measured in years of schooling, the probability of high school graduation or the probability of dropping out of school. However, they also tend to have lower socio-economic status growing up. Accounting for this reduces the estimated effect by approximately 75% when considering the probability of high school graduation or years of completed schooling by age 22.

We find some evidence that heterogeneity exists by age at first birth. For example, younger teen mothers at age 16 or 17, do not have significant estimated effects on educational outcomes at age 22. Moreover, the estimates suggest that the negative effects of an adolescent birth decrease with age relative to their peers, some of whom will themselves become teen mothers at ages 18 or 19. First time mothers at age 18 and 19, however, do have some significant estimates. The point estimates on the probability of high school graduation are negative and increase with age (in absolute value). Understanding the time path of schooling interruption depending on the age of first birth would be useful in gaining a clearer understanding of the cumulative effects of adolescent fertility on educational attainment.

Our results suggest only nuanced policy recommendations. Family planning and reproductive health policy might well benefit young girls in terms of their educational outcomes. The efficacy of such policy may also benefit from specific targeting of different age groups. On the other hand, the overall finding is that most of the observed differences in outcomes is attributable to pre-existing adverse characteristics.

6 References

Bledsoe, Caroline, John Casterline, Jennifer Johnson-Kuhn, and John Haaga (eds.) 1999. Critical Perspectives on Schooling and Fertility in the Developing World. Washington, DC: National Academy Press.

Geronimus, Arline and Sanders Korenman. 1993. "The socioeconomic costs of teenage childbearing: Evidence and interpretation." Demography 30(2): 281-290.

Grant, Monica and Kelly Hallman. 2006. "Pregnancy-related school dropout and prior school performance in KwaZulu-Natal, South Africa." Studies in Family Planning 39(4): 369-382.

Gupta, Neeru and Iuri da Costa Leite. 1999."Adolescent fertility behavior: Trends and determinants in northeastern Brazil." International Family Planning Perspectives 25(3): 125-130.

Hernstein, Richard and Charles Murray. 1995. The Bell Curve. New York: Free Press.

Hoffman, Saul. 1998. "Teenage childbearing is not so bad after all . . . or is it? A review of the new literature." Family planning Perspectives, 30(5): 236-236.

Hotz, V. Joseph, C. Mullin and Seth G. Sanders. 1997. "Bounding causal effects using data from a contaminated natural experiment: Analyzing the effects of teenage childbearing." Review of economic studies 64(4): 575-603.

Hotz, V. Joseph, Susan Williams McElroy and Seth G. Sanders. 1999. "Teenage childbearing and its lifecycle consequences: Exploiting a natural experiment." NBER w.p. 7397. Cambridge, MA.

Kaufman, Carol, Thea de Wet and Jonathan Stadler. 2001. "Adolescent pregnancy and parenthood in South Africa." Studies in Family Planning 32(2): 147-160.

Klepinger, Daniel, Shelly Lundberg and Robert Plotnick. 1995. "Adolescent fertility and the educational attainment of young women." Family Planning Perspectives 27(1): 23-28.

Lam, David, Jeremy Seekings and Meredith Sparks. 2006. The Cape Area Panel Study: Overview and Technical Documentation for Waves 1-2-3. Cape Town: University of Cape Town.

Lee, Valerie, David Burkam, Herbert Zimiles and Barbara Ladewski. 1994. "Family structure and its effect on behavioral and emotional problems in young adolescents." Journal of Research on Adolescence 4(3): 405-437.

Levine, David I. and Gary Painter. Forthcoming. "The schooling costs of teenage out-of-wedlock childbearing: Analysis with a within-school propensity score matching estimator." Review of Economics and Statistics, forthcoming.

Lloyd, Cynthia B. (ed). 2005. Growing Up Global: The Changing Transitions to Adulthood in Developing Countries. Washington, DC: National Academies Press.

Lloyd, Cynthia B. and Barbera S.Mensch. 2008. "Marriage and childbirth as factors in school dropout: An analysis of DHS data from sub-Saharan Africa." Population Studies 62(1): 1-13.

Madhavan, Sangeetha and Kevin Thomas. 2005. "Childbearing and schooling: New evidence from South Africa." Comparative Education Review 49(4): 452-467.

Marteleto, Leticia, David Lam and Vimal Ranchhod. 2008. "Sexual behavior, pregnancy and schooling among young people in urban South Africa." Studies in Family Planning 39(4): 351-367.

Moultrie, Tom and Rob Dorrington. 2004. Estimation of Fertility from the 2001 South Africa Census data. Cape Town: Centre for Actuarial Research for Statistics South Africa.

Ribar, David C. 1994. "Teenage fertility and high school completion." Review of Economics and Statistics 73(3): 413-424.

United Nations Population Division.2003. World Fertility Report 2003. New York: United Nations.

Rosenbaum, P. and D. Ruben. 1983. "The central role of the propensity score in observational studies for causal effects." Biometrika 70:41-55.

Table 1: **Summary statistics in sample, and by teen birth**

Variable	N	Mean	teen birth = 0		teen birth = 1		Diff in	std. err.	t-stat
			N	Mean	N	Mean	means	(of diff)	(of diff)
dropout	2295	0.724	1766	0.680	529	0.870	-0.190	0.020	-9.407
matric22	1129	0.509	847	0.584	282	0.282	0.302	0.035	8.759
matric20	1735	0.443	1294	0.518	441	0.215	0.303	0.026	11.462
educ22	1129	10.657	847	10.935	282	9.805	1.130	0.139	8.148
educ20	1735	10.450	1294	10.778	441	9.447	1.331	0.122	10.869
educ18	2224	9.961	1701	10.177	523	9.245	0.931	0.101	9.259
books in hh	2295	0.801	1766	0.819	529	0.739	0.079	0.022	3.676
drinker in hh	2295	0.200	1766	0.179	529	0.273	-0.094	0.024	-3.941
drugs in hh	2295	0.083	1766	0.066	529	0.138	-0.072	0.019	-3.793
educ father missing	2295	0.363	1766	0.345	529	0.423	-0.078	0.026	-2.974
educ father	2295	5.435	1766	5.764	529	4.312	1.452	0.262	5.539
educ mother missing	2295	0.117	1766	0.113	529	0.130	-0.018	0.018	-0.971
educ mother	2295	7.375	1766	7.627	529	6.518	1.109	0.211	5.246
prop. yrs with mom (0-14)	2295	0.860	1766	0.862	529	0.853	0.010	0.015	0.618
prop. yrs. with dad (0-14)	2295	0.600	1766	0.606	529	0.578	0.028	0.024	1.184
prop. grades passed (8-14)	2292	0.923	1764	0.925	528	0.916	0.010	0.005	1.785
contraception 1st sex	1674	0.590	1174	0.680	500	0.397	0.283	0.028	10.116
age 1st sex	1761	17.041	1236	17.455	525	16.147	1.309	0.108	12.163
had sex	2295	0.714	1766	0.630	529	1.000	-0.370	0.013	-27.990
age 1st period	2261	13.298	1734	13.297	527	13.302	-0.005	0.106	-0.049
1st sex forced	1773	0.016	1247	0.019	526	0.009	0.010	0.006	1.668
1st sex tricked	1773	0.043	1247	0.047	526	0.034	0.012	0.010	1.281
1st sex persuaded	1773	0.084	1247	0.080	526	0.093	-0.013	0.015	-0.838
1st sex willing	1773	0.857	1247	0.854	526	0.864	-0.010	0.018	-0.543
African	2295	0.368	1766	0.369	529	0.364	0.006	0.024	0.229
Coloured	2295	0.632	1766	0.631	529	0.636	-0.006	0.024	-0.229

Notes:

1. This calculation includes all girls in the sample, including those who are observed only before age 20. For example, if we observe a girl only up to age 18 and she has had a birth, 'teenbirth'==1, otherwise 'teenbirth'==0.

2. Sampling weights are included in the calculations of means.

3. 'dropout' is better described as an 'interrupt' variable, as 56.7% of girls in the sample who do drop out of school subsequently return at some point.

Table 2: **Probit regressions to generate the pscores**

	teen birth		birth at 16		birth at 17		birth at 18		birth at 19	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
African	-0.738**	[0.087]	-0.543**	[0.17]	-0.641**	[0.14]	-0.682**	[0.13]	-0.683**	[0.13]
1st sex persuaded	0.238	[0.12]	0.0194	[0.23]	-0.0896	[0.21]	0.457**	[0.16]	0.1150	[0.19]
1st sex tricked	-0.1690	[0.16]	-0.0469	[0.31]	-0.0880	[0.27]	-0.3230	[0.30]	-0.5060	[0.34]
1st sex forced	-1.300**	[0.34]	-	-	-0.5450	[0.44]	-	-	-	-
age 1st period	0.0151	[0.025]	-0.116*	[0.052]	0.0035	[0.042]	0.0067	[0.036]	0.0745*	[0.038]
age 1st sex	-0.169**	[0.023]	0.0803	[0.066]	-0.0688	[0.041]	0.0182	[0.038]	-0.0410	[0.036]
contraception 1st sex	-0.625**	[0.074]	-0.830**	[0.15]	-0.436**	[0.12]	-0.400**	[0.11]	-0.1880	[0.12]
prop. grades passed (8-14)	-0.2860	[0.37]	-0.2750	[0.68]	0.1030	[0.64]	-0.0999	[0.57]	0.1690	[0.60]
prop. yrs. with dad (0-14)	0.179*	[0.096]	0.1610	[0.19]	0.1460	[0.16]	0.0550	[0.14]	0.1390	[0.14]
prop. yrs with mom (0-14)	-0.259*	[0.13]	-0.564*	[0.24]	-0.1520	[0.22]	-0.0738	[0.20]	-0.1230	[0.21]
educ mother	-0.0313*	[0.014]	-0.0006	[0.027]	-0.0255	[0.023]	0.0016	[0.021]	-0.0539**	[0.021]
educ mother missing	-0.378*	[0.16]	-0.1950	[0.31]	-0.2010	[0.25]	-0.2180	[0.25]	-0.450	[0.24]
educ father	0.0000	[0.015]	0.0340	[0.028]	-0.0149	[0.024]	-0.0355	[0.022]	0.0451	[0.024]
educ father missing	0.254	[0.14]	0.2060	[0.26]	-0.0351	[0.22]	-0.1040	[0.20]	0.654**	[0.23]
drugs in hh	0.0487	[0.14]	0.3330	[0.23]	0.0703	[0.21]	-0.2730	[0.21]	0.1870	[0.22]
drinker in hh	0.0833	[0.089]	0.0136	[0.17]	0.0143	[0.15]	0.308*	[0.13]	-0.1530	[0.15]
books in hh	-0.0344	[0.082]	-0.1150	[0.16]	-0.1640	[0.13]	-0.218	[0.12]	0.265	[0.14]
Constant	3.485**	[0.57]	0.3510	[1.35]	0.7630	[1.00]	-0.5500	[0.90]	-1.4450	[0.92]
Observations	1486		655		909		1075		1080	

Notes:

1. Standard errors in brackets, ** $p<0.01$, * $p<0.05$
2. The 'teen birth' sample is restricted to women who first had sex before the age of 20.
3. The 'birth at 16' sample is restricted to women who experience their sexual debut by age 16, and who have not had a live birth by age 15.
4. A corresponding definition is used for the 'birth at 17', 'birth at 18' and 'birth at 19' variables.

Table 3: **Summary statistics by 'teen birth' after sample restriction and re-weighting**

Variable	teen birth = 0		teen birth = 1		Diff in	std. err.	t-stat
	N	Mean	N	Mean	means	(of diff)	(of diff)
dropout	991	0.766	492	0.867	-0.101	0.029	-3.428
matric22	536	0.387	271	0.283	0.104	0.050	2.081
matric20	801	0.336	417	0.212	0.125	0.038	3.299
educ22	536	10.21	271	9.82	0.395	0.199	1.991
educ20	801	10.12	417	9.44	0.675	0.163	4.130
educ18	978	9.701	486	9.228	0.473	0.135	3.508
books in hh	991	0.736	492	0.735	0.000	0.031	0.003
drinker in hh	991	0.271	492	0.270	0.001	0.037	0.035
drugs in hh	991	0.129	492	0.143	-0.014	0.030	-0.450
educ father missing	991	0.446	492	0.424	0.022	0.038	0.572
educ father	991	4.185	492	4.301	-0.116	0.352	-0.329
educ mother missing	991	0.130	492	0.128	0.001	0.029	0.047
educ mother	991	6.548	492	6.543	0.005	0.302	0.018
prop. yrs with mom (0-14)	991	0.825	492	0.849	-0.024	0.025	-0.959
prop. yrs. with dad (0-14)	991	0.541	492	0.576	-0.035	0.035	-1.019
prop. grades passed (8-14)	991	0.929	492	0.916	0.013	0.007	1.861
contraception 1st sex	991	0.398	492	0.396	0.003	0.035	0.076
age 1st sex	991	16.20	492	16.15	0.053	0.127	0.421
had sex	991	1.000	492	1.000	0.000		
age 1st period	991	13.29	492	13.32	-0.027	0.110	-0.244
1st sex forced	991	0.015	492	0.008	0.007	0.007	1.016
1st sex tricked	991	0.034	492	0.035	-0.001	0.011	-0.085
1st sex persuaded	991	0.133	492	0.094	0.039	0.026	1.497
1st sex willing	991	0.818	492	0.863	-0.046	0.028	-1.605
African	991	0.372	492	0.371	0.002	0.033	0.054
Coloured	991	0.628	492	0.629	-0.002	0.033	-0.054

Notes:

1. Propensity score weights were obtained from a kernel matching procedure using the *psmatch2* command in Stata. An Epanechnikov kernel with a bandwidth of 0.06 was used.

2. This calculation only includes girls for whom we have a valid pscore from the probit regression in column 1 of table 2.

3. Girls who had not had sex by age 19 were excluded from the estimation sample.

4. The product of the sampling weights and the weights from the matching algorithm included in the calculations of means.

5. A common support condition was imposed as well.

6. The number of observations for the outcome variables varies due to missing values for some outcomes.

Table 4: **Regression results: Coefficients on 'teen birth' after sample restriction and re-weighting**

		Dependent variable					
Description of specification		matric20	matric22	educ18	educ20	educ22	dropout
Specification 1:	coeff.	-0.303***	-0.302***	-0.931***	-1.331***	-1.130***	0.190***
No sample restriction,	std. err.	[0.026]	[0.035]	[0.10]	[0.12]	[0.14]	[0.020]
sampling weights only,	Obs	1735	1129	2224	1735	1129	2295
no covariates	R-sq	0.07	0.07	0.05	0.09	0.07	0.03
Specification 2:	coeff.	-0.208***	-0.227***	-0.620***	-0.921***	-0.801***	0.147***
No sample restriction,	std. err.	[0.025]	[0.033]	[0.081]	[0.11]	[0.12]	[0.019]
sampling weights only,	Obs	1718	1118	2193	1718	1118	2258
limited covariates	R-sq	0.31	0.28	0.45	0.39	0.37	0.16
Specification 3:	coeff.	-0.125***	-0.112***	-0.382***	-0.568***	-0.358***	0.102***
With sample restriction,	std. err.	[0.029]	[0.037]	[0.096]	[0.12]	[0.13]	[0.024]
sampling weights only,	Obs	1221	810	1467	1221	810	1486
all covariates	R-sq	0.27	0.26	0.37	0.37	0.4	0.14
Specification 4:	coeff.	-0.100***	-0.0754*	-0.348***	-0.510***	-0.274*	0.0825***
With sample restriction,	std. err.	[0.032]	[0.039]	[0.11]	[0.14]	[0.16]	[0.026]
sampling & propensity score	Obs	1218	807	1464	1218	807	1483
matching weights,	R-sq	0.24	0.27	0.36	0.34	0.37	0.14
all covariates							

Notes:

1. Standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1.

2. Propensity score weights were obtained from a kernel matching procedure using the *psmatch2* command in Stata. An epanechnikov kernel with a bandwidth of 0.06 was used.

3. Girls who had not had sex by age 19 were excluded from the estimation sample in Spec. 3 & 4.

4. The product of the sampling weights and the weights from the matching algorithm are used to weight the regression in specification 4.

5. A common support condition was imposed in specification 4.

6. The set of full covariates suppressed is fully described in Table 3. (Spec. 3 & 4)
In specification 2, the variables on age of sexual debut, contraceptive usage and description of first sexual experience were not included in the regression.

7. In specification 4, there are 492 teen mothers out of the 1483 women in the estimation sample.

Table 5: **Regression results: Estimates of the effect of a first birth at various ages.**

age at 1st birth		Dependent variable					
		matric20	matric22	educ18	educ20	educ22	dropout
birth at 16	coeff.	-0.0162	0.02	-0.383	-0.742**	-0.27	0.0502
	std. err.	[0.055]	[0.062]	[0.27]	[0.31]	[0.27]	[0.035]
	Observations	491	307	638	491	307	653
	R-squared	0.25	0.39	0.28	0.3	0.49	0.13
birth at 17	coeff.	-0.103**	-0.0428	-0.285*	-0.334*	-0.0483	0.0291
	std. err.	[0.048]	[0.066]	[0.15]	[0.19]	[0.25]	[0.038]
	Observations	699	453	892	699	453	907
	R-squared	0.21	0.25	0.35	0.34	0.3	0.13
birth at 18	coeff.	-0.100**	-0.127**	-0.299**	-0.502***	-0.402**	0.0576
	std. err.	[0.044]	[0.055]	[0.14]	[0.18]	[0.19]	[0.040]
	Observations	860	570	1060	860	570	1073
	R-squared	0.29	0.32	0.43	0.38	0.43	0.17
birth at 19	coeff.	-0.0932*	-0.115*	-0.00119	-0.319	-0.266	0.103**
	std. err.	[0.050]	[0.061]	[0.13]	[0.20]	[0.19]	[0.042]
	Observations	885	591	1067	885	591	1080
	R-squared	0.25	0.3	0.36	0.25	0.38	0.13

Notes:

1. Standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1 .

2. There are 85, 97, 130 & 113 'treated' observations in the estimation samples
for birth at 16, birth at 17, birth at 18 and birth at 19 respectively.

3. Coefficients omitted for the full set of other covariates. (those included in Table 3).

4.These results are from models analogous to those described as specification 4 in Table 4.