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 29, 2011

*Corresponding author: vimal.ranchhod@gmail.com

†Lam is at the University of Michigan.

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

§Marteleto is at the University of Texas at Austin.

¶Support for this research was provided by the Hewlett Foundation, 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). Leibbrandt acknowledges the Research Chairs Initiative of the Department of Science and Technology and National Research Foundation for funding his work as the Research Chair in Poverty and Inequality. Useful comments have been provided by seminar participants at the University of Michigan, the University of Cape Town, the African Econometric Society 2008 conference, the PAA 2009 conference and the PopPov 2011 conference. All errors and omissions remain the responsibility of the authors.

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 the CAPS dataset are 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 a more 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. Our best estimates of the effect of a teen birth on high school graduation by ages 20 and 22 are -5.9 and -2.7 percentage points respectively. The former is significant at the 5% level, while the latter is not statistically significant. Thus, there appears to be some 'catching up' in educational attainment by teen mothers. We find only limited support for the hypothesis that there is heterogeneity in the effect of a teen birth, depending on the actual age of the first birth. By age 22, none of the estimates for high school graduation or years of schooling are statistically significant, regardless of the specific age at which the teen birth occurred. Despite this, we do find evidence that a teen birth does correlate with reduced educational expectations. The proportion of teen mothers who report an expected final educational attainment of high school graduation or greater is about 15 percentage points lower than the matched set of non-teen mothers, but this is not manifest amongst the girls whom we know will subsequently become teen mothers at some point after these expectations are measured.

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 22% 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 schooling. 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. 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 interrelationships between sexual activity, fertility and educational outcomes that young adults experience in the process of transitioning into adulthood. Thus, there is a considerable body of research investigating the causes and consequences of adolescent fertility in South Africa.

Nonetheless, there is a lack of empirical research that makes a serious attempt at identifying the causal impact of fertility on educational attainment. This is the primary contribution of our paper. For methodological guidance, we turn to the corresponding international literature, where the causal question has been pursued for several years. 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. Econometrically, the problem is one of endogeneity due to selection into 'treatment'.

The literature on the effects of teen births 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 attempt to control for home background has been to use a 'siblings fixed effects' estimation method, as in Geronimus and Korenman (1993). Some (Hotz et al, 1997 and Hotz et al, 1999) have used 'almost-natural' experiments such as miscarriages in estimation. An alternative (Ribar, 1994 and Klepinger et al, 1995) has been to use the age of menarche as an instrument in an instrumental variables method. Levine and Painter (2003), make use of a within-school propensity score matching estimator. Collectively this 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. This is a longitudinal dataset of young adults from the Cape Metropolitan area. The first wave was conducted in 2002 with a sample of about 4800 young adults aged 14 to 22. We considered various econometric methods that could be employed to inform our empirical strategy. Many of the methods that have been employed previously in different contexts were not satisfactory, primarily due to data constraints. The data has too few female sibling pairs who were old enough to yield a sample size large enough to reasonably employ a siblings fixed effects estimator. There are also very few reported miscarriages. We further considered what observed alternatives we could use in an instrumental variables approach. Most of these would fail a strict exogeneity requirement. A candidate instrument is the introduction of a Child Support Grant in 1999, which is a means tested unconditional cash transfer program designed to assist mothers of young children. However, the cohort in our sample, combined with the timing of the introduction of the policy, resulted in the vast majority of the respondents having been teenagers after the introduction of the policy. Moreover, the introduction of the policy does not correlate strongly with adolescent fertility rates in our sample. This leads to potential estimation problems due to the combination of small and finite samples together with weak instruments.[1]

In this paper, we employ a propensity score matching method to reweight observations in our regression. In addition, we 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 international literature. 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. These smaller negative effects remain statistically significant at ages 18 and 20, but decrease with age. By age 22, the estimated effect of a teen birth on the probability of having graduated from secondary school is -2.7 percentage points and the estimate is not statistically significant.

From a family planning and reproductive health policy perspective, there are two possible and non-mutually exclusive interventions. First, one could attempt to raise the age of sexual debut, thus changing the age at first birth. Second, one could try to reduce existing fertility rates conditional on existing patterns of sexual behavior. This could include information campaigns and access to contraception, amongst other things. The methods used in this paper relate to the second of these interventions, as we restrict our estimation sample to only include girls who were sexually active as teenagers. Our findings suggest that a reduction

[1] Bound et al (1995) provide a discussion of the econometric problems that arise in this context.

in adolescent childbearing would not enhance the educational attainment of young mothers by very much, as up to 90% of the mean difference in high school graduation rates can be attributed to other factors.

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 (2008). The data has detailed information about respondents' early life environment, schooling progress, expected final educational attainment, age at 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 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 in a number of socio-economic dimensions. In particular, white females have much higher levels of educational attainment and grade progression rates. Including whites as potential counterfactual observations in our analysis would thus likely confound our results. Of the 2294 remaining observations from wave 1, we have 1934 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. 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 someone in the household with a drinking problem when the respondent was growing up , with a drug addiction problem and whether the household has five or more books.[2] 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 use in estimating 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.

We use a number of different outcome variables:

- *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.[3]
- *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 school-

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

[3]We do not use respondents' final years of completed secondary schooling, nor do we include any college level schooling. This is due to the data which is restricted to a relatively young population. It is not uncommon for people in this environment to return to school at relatively late ages, which results in our data being incomplete in terms of its measures of final educational attainment.

ing.

- *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.

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 can be included in the analysis for 'educ18' and in the 'dropout' outcome. We chose not to exclude these girls as we would then lose 512 observations in our sample. This truncation problem makes the separation into 'treated' and 'untreated' groups problematic. For example, 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 ages 18 or 19. This causes a type of measurement error problem, as some girls who ultimately would be teen mothers get classified as non-teen mothers, because we do not observe them for a long enough period. This will probably 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. By age 22, only 49.5% have completed high school. Educational attainment between ages 18 and 20 increases from 9.97 to 10.48 years of schooling, an increase of about 0.5 years, and continues to increase by a smaller amount between ages 20 and 22. The proportion that has experienced a dropout is very high, at 47%. Racially, 36.8% of the sample is African, with the remainder being coloured. 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.[4] 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

[4] The survey captured parental educational attainment if known by the respondents, regardless of co-residency.

with their mothers, and a smaller proportion with their fathers. Grade progression from ages 8 to 14 is fairly high at 0.917, although if interpreted as a probability, a significant fraction of the girls will repeat a grade during primary school.

The mean age at menarche is about 13.2 years, while the percentage who have ever had sex is 72.5%. Of these, the mean age of sexual debut is 17.11. Of interest is the percentage who used contraception during their first sexual experience, at only 56.7%. The majority of respondents report that they were willing or persuaded to engage in their first sexual experience, although 1.6% report being forced into their sexual debut.

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 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 also applied to girls whom we observe only up to some age less than 20.

We first use a probit model to estimate the probability that a girl has had a teen birth. 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 is only done for those observations who satisfy this sample restriction and do not have any missing values for any of the covariates.[5]

Once we have the propensity score, we perform a kernel matching procedure on the 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. We impose 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[6], which allows us to obtain an appropriate set of counterfactual girls.

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 linearly in the final specification. We again restrict the estimation sample to only include girls who have had their first sexual experience before the age of 20. 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 indicator variable, these are effectively linear probability models, and the coefficients should be interpreted as marginal 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 unbiased estimate of the 'average treatment effect on the treated' (ATT).

[5]For girls with no parental education level, a value of zero was included and the relevant indicator variable takes on a value of one.

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

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 that 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. This is a very different counterfactual to the same first time mother who would have had a first birth much later. 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 that we are concerned with are age specific measures of 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 the age of sexual debut affects 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.[7] 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. Most of the difference in means and the related t-statistics are much smaller in absolute value than the corresponding statistics in Table 1, and very few of the differences are significant at the 5% level of significance.

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

A final piece of exploratory analysis that we undertake is to calculate and compare the mean expected final educational attainment for the groups of teen and non-teen mothers. These expectations are only captured in wave 1 and wave 3, and are not age specific. We classify these into a binary variable that takes a value of 1 if the final expected attainment is greater than or equal to high school graduation. By comparing the wave 1 and wave 3 measures, we have some idea about how these expectations evolve with the passing of time. For the teen mothers, we further separate the observations into 3 sub-groups; those who were already mothers by wave 1, those who became mothers between wave 1 and wave 3, and those who we know will subsequently become teen mothers but who are not yet mothers by the time of the wave 3 measure. This allows us to observe how the expectations change in conjunction with the event of giving birth. For the non-mothers, we calculate the mean expectation first for the entire group, then for the subset of girls who were sexually active before age 20, and finally we calculate the mean by applying the matching weights. By doing so, we are able to see what the effect of the sample restriction and the matching process is on the mean of expected educational attainment.

4 Results

4.1 Teen Births

The results for the teen birth analysis are presented in Table 4. For each outcome variable, we start with the coefficient from a regression with no covariates. This is analogous to the difference in means in Table 1. Specification 2 introduces the sample restrictions. This excludes girls who were not yet sexually active, or who made their sexual debut after the age of 19. It also imposes the common support restriction from the matching procedure. Specifications 2 to 5 are all estimated on data from the same sub-sample of respondents. We then show results where we add 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 fifth specification.

Across the five specifications, the estimated effect on high school graduation by age 20 decreases from -0.294 to -0.166 to -0.117 to -0.076 and finally to -0.059. The first four are significant at the 1% level, and the fifth one is significant at the 5% level. By age 22, the corresponding estimates on high school graduation are smaller in magnitude and not always significant, relative to the estimates at age 20. To begin with, the coefficient estimate in specification 1 is -0.28, and is significant at the 1% level. After incorporating the matching weights and the full set of covariates, the coefficient estimate is reduced by about 90% to -0.027, and is no longer statistically significant. The estimated effect on dropping out of school also decreases from 0.382 to 0.29 to 0.224 to 0.19 to 0.164, but remains significant in each specification. We find that a teen birth significantly affects the probability of graduating from high school by age 20 and of dropping out of school, but does not significantly affect the probability of graduating from high school by age 22. This suggests some element of 'catching up' by the teen mothers. The decrease in the magnitude of the coefficients across the specifications 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 0.945 fewer years of schooling on average, and the difference is highly significant. From specifications 2 to 5, this estimate reduces to 0.60, 0.40, 0.265 and finally stabilizes at 0.284 fewer years of schooling for the teen mothers. All of these estimates are significant at the 1% level. At age 20, the trend in the coefficients across specifications is very similar to the trend at age 18, but the coefficients are all larger in absolute value. By age 22, the corresponding estimates are 1.06, 0.663, 0.317, 0.139 and 0.144 fewer years of schooling. The first three estimates are significant at the 1% level, but the last two are not. The educational gap between the teen mothers and non-teen mothers increases between ages 18 and 20, but decreases considerably between ages 20 and 22. This is true in all the specifications. This reinforces the point that teen mothers do experience some element of 'catching up' in terms of secondary schooling.[8]

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 by age in terms of the effects of teen births. The 'treatment' here is a first birth at age i, and the counterfactual group is drawn from the set of girls who were sexually active by age i and had not yet had a birth at that age. We only present results from the propensity score weighted regressions. On aggregate, the evidence is mixed. The sign of the estimates, when the estimates are significant, are similar to those discussed above. A number of the coefficients are not significant. This might be partly a result of smaller estimation samples and subsequent lack of power.[9]

Having a first birth at age 16 seems to have only a modest effect on the probability of graduating from high school, relative to other girls who were also sexually active by age 16 but did not have a first birth by age 16. First time mothers at age 17, however, are 12.4 percentage points less likely than their peers to have completed secondary school by age 20. By age 22, this coefficient decreases to -0.059 and is not significant. For these first time mothers at age 17, the estimated effects on years of schooling attained is about -0.5 of a year at ages 18 and 20, but this decreases to -0.12 by age 22. In general, the coefficient estimates on educational attainment are not significant for the first time mothers at age 18 and 19. At the same time, the estimated effect on dropping out of school is large and significant for the first time mothers at ages 17, 18 and 19. This is somewhat puzzling, given the lack of significance of the estimates for attainment and high school graduation.

[8] Note that this does not consider potential differences in college attendance, nor differences in accumulated work experience. Accounting for these would likely increase the estimated adverse effects of a teen birth.

[9] There are 65, 103, 143 & 122 'treated' observations in the estimation samples for a birth at age 16, 17, 18 and 19 respectively. Note that the estimation samples differ depending on the dependent variable. This is because of the distribution of the most recently observed age in the data.

This combination would be possible if the counterfactual group is enrolled but not passing their grades at a high rate, or the mothers who drop out subsequently return to high school. These explanations are not mutually exclusive, and both may provide a partial explanation for the observed results.

4.3 Expected final educational attainment

The final piece of analysis that we undertake is to summarize expectations of educational attainment for the teen mothers and non-teen mothers. We compare these expectations across the two groups and see how they change between wave 1 and wave 3, i.e. 2002 and 2005. We make use of a question that asks *"As it stands now, how much education do you think you will complete?"*. We find this additional analysis useful for a number of reasons. First, there is the possibility of reverse causality between dropping out of school and teen motherhood. Our interpretation of the results thus far has been to assume that adolescent fertility causes dropping out of school, but it could also be that girls who drop out of school start families younger and the results cannot separately identify the relative magnitudes of these two phenomena. By looking at expectations, we can see whether the girls who become teen mothers have lower educational expectations to begin with. Second, we can more directly observe how these expectations change over time, as a function of whether the girl has had her first child yet or not. Third, we can see the effect of the sample restriction and matching on the mean expectations of the non-teen mothers. This enables us, at least in part, to unpack the process by which our main results are obtained in Table 4.

It is worth noting that these expectations appear to be highly optimistic given the realized attainment by age 22 in the sample. For example, in wave 3, 78.6% of all girls expect to attain a high school graduation certificate or greater, yet by age 22, only 45% have already done so. Regardless of which sub-sample one considers, one finds large differences between the expectations and the realized values. Nonetheless, the way they change depending on the group under consideration and across time are potentially informative. In addition, the number of observations used is considerably smaller than the sample size within each cell.

This is due primarily to item non-response, as a sizable fraction of individuals answer "don't know" to the question.

The results for this component of the analysis are presented in Table 6. Several interesting patterns are revealed in the table. First, the expectations of the mothers are almost always lower than those of the non-mothers. Second, expectations become considerably smaller with time. In wave 1, more than 90% of teen mothers and non-teen mothers expect to graduate from high school, but by wave 3 this reduces to 57.5% and 82.2% for the teen mothers and non-teen mothers respectively. Third, the decrease in expectations is more pronounced amongst the teen mothers than the comparison group. For the mothers, we observe a decrease of about 32.6 percentage points in the three years between wave 1 and wave 3, whereas for the non-teen mothers the corresponding statistic is 13.4 percentage points. Fourth, when we restrict our attention to the non-teen mothers, we find that the group of girls who become sexually active by age 19 do not have expectations that are substantially different from those who experience their sexual debut later. In addition, the matching process does not affect the mean expectations in a material way in wave 1. However, the matching weights do change the mean expectations by wave 3, where the mean decreases by about 9 percentage points from 0.801 to 0.711. Thus, by wave 3, the matching process increases the weight of girls who expect to have relatively lower scholastic outcomes, amongst the counterfactual observations. A final and very interesting point that emerges from the table is seen when we compare the girls who will subsequently become teen mothers in wave 4 with the matched set of girls who are not teen mothers. Their expectations, in both wave 1 and wave 3, are virtually identical. This means that the girls who will experience their first birth within the next year do not have any different expectations from their counterparts *a priori*. In conjunction with the rest of the evidence in this table, this suggests that the event of a first birth does indeed affect the educational expectations of young girls.

If a first birth in their teens does affect expectations downwards, then we are left with a somewhat puzzling situation given our main results which suggest relatively small effects on actual attainment which are not significant by age 22. There are three possible reasons, and these are not mutually exclusive. First, it seems that all the groups of girls are overly optimistic about their final educational outcomes. Thus, expectations and outcomes do not need to converge at these ages. Second, for a teen birth to occur in wave 4, which is our final wave, these girls cannot be part of the estimation sample for the educational outcomes at age 22, as they are still too young. There is thus some correspondence to reality as the estimated effects at ages 18 and 20 are larger and statistically significant, as compared to those at age 22. Third, the number of observations we have valid expectations data on for teen mothers who have their first birth between wave 3 and wave 4 is relatively small, and we should be careful about making bold inferences based off of these findings.

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 and expectations. 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 re-weight 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 international literature. 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 90% when considering the probability of high school graduation or years of completed schooling by age 22. We find only limited evidence that heterogeneity exists by age at first birth. This is manifest primarily in the likelihood of dropping out of school, and not in the likelihood of high school graduation.

Our results suggest only nuanced policy recommendations. Family planning and reproductive health policy that reduces adolescent fertility would probably not substantially benefit young girls in terms of their educational outcomes. The overall finding is that most of the observed differences in outcomes is attributable to pre-existing adverse characteristics. Thus, the girls who do become teen mothers, had they not had that birth, would likely have had a relatively low level of educational attainment in any case. We make the conjecture that policy with an objective of educational attainment might be more effective if it concentrated on socio-economic factors and the household environment.

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.

Bound, John, David A. Jaeger and Regina M. Baker. 1995. "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak." Journal of the American Statistical Association 90(430): 443-450.

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, Cally Ardington, Nicola Branson, Anne Case, Murray Leibbrandt, Alicia Menendez, Jeremy Seekings and Meredith Sparks. 2008. The Cape Area Panel Study: Overview and Technical Documentation Waves 1-2-3-4 (2002-2006). The University of Cape Town, October 2008.

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. 2003. "The schooling costs of teenage out-of-wedlock childbearing: Analysis with a within-school propensity score matching estimator." Review of Economics and Statistics 85(2003): 884-900.

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	2291	0.470	1787	0.387	504	0.769	-0.382	0.024	-16.00
matric22	1076	0.495	820	0.560	256	0.280	0.280	0.036	7.76
matric20	1711	0.438	1297	0.507	414	0.213	0.294	0.027	10.88
educ22	1076	10.631	820	10.877	256	9.814	1.063	0.143	7.43
educ20	1711	10.488	1297	10.778	414	9.538	1.240	0.113	10.98
educ18	2214	9.974	1716	10.183	498	9.238	0.945	0.099	9.59
African	2294	0.368	1789	0.370	505	0.360	0.009	0.024	0.38
Coloured	2294	0.632	1789	0.630	505	0.640	-0.009	0.024	-0.38
books in hh	2294	0.801	1789	0.822	505	0.724	0.098	0.022	4.37
drinker in hh	2294	0.200	1789	0.180	505	0.274	-0.095	0.024	-3.89
drugs in hh	2294	0.083	1789	0.066	505	0.141	-0.075	0.019	-3.86
educ father missing	2294	0.363	1789	0.346	505	0.428	-0.083	0.027	-3.11
educ father	1415	8.532	1136	8.788	279	7.465	1.323	0.261	5.07
educ mother missing	2294	0.116	1789	0.111	505	0.132	-0.020	0.018	-1.11
educ mother	2026	8.348	1586	8.583	440	7.476	1.107	0.186	5.95
prop. yrs live with mother (0-14)	2293	0.844	1788	0.846	505	0.837	0.009	0.016	0.54
prop. yrs. live with father (0-14)	2293	0.585	1788	0.592	505	0.559	0.032	0.024	1.33
prop. of grades passed (8-14)	2290	0.917	1786	0.920	504	0.904	0.017	0.007	2.28
contraception 1st sex	1787	0.567	1282	0.644	505	0.388	0.256	0.028	9.22
age 1st sex	1756	17.113	1256	17.525	500	16.157	1.368	0.098	14.00
had sex	2288	0.725	1783	0.648	505	1.000	-0.352	0.013	-26.87
age 1st period	2259	13.239	1756	13.231	503	13.267	-0.036	0.086	-0.42
1st sex forced	1779	0.016	1277	0.019	502	0.009	0.009	0.006	1.50
1st sex tricked	1779	0.044	1277	0.047	502	0.035	0.013	0.010	1.27
1st sex persuaded	1779	0.086	1277	0.081	502	0.096	-0.015	0.015	-0.97
1st sex willing	1779	0.855	1277	0.853	502	0.860	-0.007	0.019	-0.36

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. 13.6% 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	-0.754**	[0.087]	-0.422*	[0.18]	-0.822**	[0.15]	-0.645**	[0.12]	-0.581**	[0.12]
1st sex persuaded	0.197	[0.12]	-0.052	[0.24]	-0.33	[0.22]	0.339*	[0.15]	0.228	[0.17]
1st sex tricked	-0.246	[0.16]	0.106	[0.28]	0.025	[0.24]	-0.398	[0.27]	-0.306	[0.26]
1st sex forced	-1.194**	[0.34]	0	0	-0.820	[0.47]	-0.763	[0.49]	0	0
age 1st period	0.029	[0.025]	-0.065	[0.058]	-0.06	[0.042]	0.071*	[0.034]	0.031	[0.035]
age 1st sex	-0.222**	[0.026]	-0.117	[0.082]	-0.213**	[0.054]	0.005	[0.042]	-0.032	[0.038]
contraception 1st sex	-0.539**	[0.072]	-0.915**	[0.17]	-0.518**	[0.12]	-0.297**	[0.10]	-0.293**	[0.11]
prop. grades passed (8-14)	-0.406	[0.33]	-1.975**	[0.58]	-0.197	[0.54]	-0.348	[0.47]	0.373	[0.52]
prop. yrs. with dad (0-14)	0.159	[0.095]	-0.09	[0.20]	0.224	[0.17]	0.015	[0.13]	0.176	[0.14]
prop. yrs with mom (0-14)	-0.202	[0.13]	-0.188	[0.25]	-0.581**	[0.21]	0.003	[0.18]	-0.089	[0.19]
educ mother	-0.021	[0.014]	-0.009	[0.029]	0.001	[0.023]	-0.002	[0.019]	-0.039*	[0.019]
educ mother missing	-0.303	[0.16]	-0.301	[0.33]	-0.149	[0.26]	-0.014	[0.22]	-0.475*	[0.23]
educ father	-0.009	[0.015]	0.032	[0.030]	-0.014	[0.024]	-0.039	[0.021]	0.022	[0.022]
educ father missing	0.157	[0.14]	0.095	[0.28]	-0.083	[0.23]	-0.125	[0.19]	0.452*	[0.21]
drugs in hh	0.053	[0.13]	0.059	[0.25]	0.193	[0.20]	-0.219	[0.19]	0.114	[0.20]
drinker in hh	0.119	[0.088]	0.038	[0.17]	-0.196	[0.15]	0.414**	[0.12]	-0.075	[0.14]
books in hh	-0.081	[0.080]	-0.085	[0.16]	-0.091	[0.13]	-0.17	[0.11]	-0.008	[0.12]
Constant	4.175**	[0.54]	3.864**	[1.45]	4.388**	[1.06]	-1.12	[0.85]	-0.937	[0.84]
Observations	1571	.	707	.	1003	.	1206	.	1181	.

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	1078	0.576	489	0.764	-0.188	0.033	-5.67
matric22	544	0.334	248	0.285	0.049	0.047	1.03
matric20	852	0.309	400	0.218	0.091	0.035	2.60
educ22	544	10.08	248	9.84	0.248	0.192	1.29
educ20	852	10.10	400	9.58	0.521	0.146	3.56
educ18	1059	9.680	482	9.270	0.410	0.125	3.29
books in hh	1078	0.750	489	0.731	0.019	0.029	0.64
drinker in hh	1078	0.280	489	0.275	0.005	0.036	0.13
drugs in hh	1078	0.119	489	0.144	-0.025	0.028	-0.89
educ father missing	1078	0.446	489	0.432	0.014	0.036	0.39
educ father	661	7.384	270	7.427	-0.043	0.320	-0.14
educ mother missing	1078	0.136	489	0.133	0.003	0.029	0.11
educ mother	949	7.602	426	7.531	0.072	0.222	0.32
prop. yrs live with mother (0-14)	1078	0.807	489	0.831	-0.024	0.024	-0.97
prop. yrs. live with father (0-14)	1078	0.520	489	0.553	-0.033	0.033	-0.99
prop. of grades passed (8-14)	1078	0.920	489	0.904	0.016	0.008	1.95
contraception 1st sex	1078	0.398	489	0.399	-0.001	0.034	-0.04
age 1st sex	1078	16.16	489	16.16	0.001	0.108	0.01
had sex	1078	1.000	489	1.000	0.000	0.000	
age 1st period	1078	13.29	489	13.28	0.010	0.107	0.10
1st sex forced	1078	0.015	489	0.008	0.008	0.008	0.96
1st sex tricked	1078	0.036	489	0.036	0.000	0.011	0.04
1st sex persuaded	1078	0.143	489	0.096	0.047	0.025	1.87
1st sex willing	1078	0.805	489	0.860	-0.055	0.027	-2.00
African	1078	0.386	489	0.371	0.015	0.032	0.48
Coloured	1078	0.614	489	0.629	-0.015	0.032	-0.48

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 is included in the calculations of means.

5. A common support condition was imposed.

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.	-0.294**	-0.28**	-0.945**	-1.24**	-1.0638**	0.382**
No sample restriction	std. err.	[0.027]	[0.036]	[0.099]	[0.113]	[0.143]	[0.024]
Sampling weights only	Obs	1711	1076	2214	1711	1076	2291
No covariates	R-sq	0.063	0.056	0.052	0.090	0.066	0.099
Specification 2:	coeff.	-0.166**	-0.139**	-0.601**	-0.838**	-0.663**	0.29**
Sample restriction	std. err.	[0.03]	[0.039]	[0.105]	[0.122]	[0.156]	[0.027]
Sampling weights only	Obs	1252	792	1541	1252	792	1567
No covariates	R-sq	0.028	0.019	0.028	0.050	0.030	0.077
Specification 3:	coeff.	-0.117**	-0.09*	-0.404**	-0.483**	-0.317*	0.224**
Sample restriction	std. err.	[0.027]	[0.037]	[0.085]	[0.098]	[0.129]	[0.026]
Sampling weights only	Obs	1252	792	1541	1252	792	1567
Limited covariates	R-sq	0.228	0.200	0.329	0.372	0.346	0.189
(excl. sexual behavior)							
Specification 4:	coeff.	-0.076**	-0.049	-0.265**	-0.315**	-0.139	0.19**
Sample restriction	std. err.	[0.028]	[0.038]	[0.087]	[0.101]	[0.136]	[0.028]
Sampling weights only	Obs	1252	792	1541	1252	792	1567
All covariates	R-sq	0.246	0.222	0.347	0.396	0.372	0.214
Specification 5:	coeff.	-0.059*	-0.027	-0.284**	-0.304**	-0.144	0.164**
Sample restriction	std. err.	[0.03]	[0.038]	[0.097]	[0.116]	[0.161]	[0.029]
Sampling weights &	Obs	1252	792	1541	1252	792	1567
matching weights	R-sq	0.234	0.255	0.362	0.388	0.354	0.195
All covariates							

Notes:

1. Standard errors in brackets, ** p<0.01, * p<0.05
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. The sample restriction excludes girls who had not yet had sex, or had not had sex by age 19. The sample restriction also includes a common support condition from the matching process.
4. The product of the sampling weights and the weights from the matching algorithm are
5. The set of full covariates suppressed is described in Table 3. (Spec. 4 & 5)
6. In specification 3, the variables on age of sexual debut, contraceptive usage and description of first sexual experience were not included in the regression.

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.	0.012	0.06	-0.317*	-0.42*	-0.237	0.052
	std. err.	[0.055]	[0.068]	[0.19]	[0.233]	[0.281]	[0.049]
	Observations	515	303	686	515	303	706
	R-squared	0.22	0.316	0.458	0.478	0.433	0.22
birth at 17	coeff.	-0.124**	-0.059	-0.523**	-0.544**	-0.12	0.20**
	std. err.	[0.046]	[0.064]	[0.152]	[0.181]	[0.234]	[0.041]
	Observations	754	460	982	754	460	1003
	R-squared	0.2	0.325	0.375	0.421	0.478	0.212
birth at 18	coeff.	-0.064	-0.082	-0.203	-0.29	-0.186	0.163**
	std. err.	[0.042]	[0.055]	[0.136]	[0.155]	[0.212]	[0.041]
	Observations	947	599	1187	947	599	1205
	R-squared	0.282	0.285	0.346	0.363	0.318	0.204
birth at 19	coeff.	-0.051	-0.065	-0.025	-0.091	-0.136	0.138**
	std. err.	[0.045]	[0.056]	[0.13]	[0.144]	[0.186]	[0.043]
	Observations	945	605	1159	945	605	1178
	R-squared	0.272	0.306	0.375	0.385	0.386	0.244

Notes:

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

2. There are 65, 103, 143 & 122 'treated' observations in the estimation samples for birth at 16, birth at 17, birth at 18 and birth at 19 respectively.

3. The estimation sample is retricted to girls who were sexually active by the relevant age but had not yet had a first birth by that age..

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

5.These results are from models analogous to those described as specification 5 in Table 4.

Table 6: **Mean proportion that expect to complete matric in Waves 1 and 3**

	Wave 1		Wave 3	
	N	Mean	N	Mean
Teen Mothers				
All combined	350	0.901	364	0.575
1st birth by Wave 1	173	0.885	187	0.559
1st birth by Wave 3	130	0.906	143	0.559
1st birth after Wave 4	47	0.941	34	0.720
Non-Teen Mothers				
All	1599	0.956	1321	0.822
Sample restriction	961	0.957	819	0.801
Sample restriction & matching weights	930	0.951	806	0.711

Notes:

1. Girls who have already completed matric are given a value of 1 in the calculation.
2. Sample restriction for non-teen mothers restricted to girls who were sexually active by age 19.
3. The number of observations changes across waves due to attrition and changes in item non-response.