The Effects of Teenage Pregnancy on Schooling and Labor Force Participation: Evidence from Urban South Africa

Natalia Cantet, UIUC*

Job Market Paper

This version: April 16, 2020
Click here for the latest version

Abstract

Young mothers have worse health, less schooling, and poorer job market performance in adulthood. However, because there is selection into early motherhood, the causal impact of teenage pregnancy on human capital investments is difficult to estimate. While the majority of the literature has focused on high-income settings. This papers examine the impact of teenage pregnancy in a low income setting Cape Town, South Africa. I study the effects on educational outcomes and future labor-force participation using two main identification strategies: an instrumental variable strategy that relies on the number of fertile teenage years as an instrument for teenage pregnancy and sibling differences among a subsample of sisters in which one sister reported a teenage pregnancy and at least one did not. I find that pregnancy before the age of 19 increases approximately 50 % in the likelihood of failing a grade and an increase of 27% (10 percentage points) in the probability of dropping out of school. Teenagers who report a pregnancy obtain 1.8 fewer years of education. Finally, two specific South African characteristics mitigate the negative effects of teenage pregnancy.

Keywords: Teenage pregnancy; Health; Education; Labor Force Participation.

JEL Classification: O1, I2, J1

*Department of Economics, University of Illinois at Urbana-Champaign. www.nataliacantet.com E-mail: cantet2@illinois.edu. Acknowledgements: I am very grateful to Rebecca Thornton and Marieke Kleemans for their advice during the course of the project. This paper also benefited from the helpful comments from seminar participants at the University of Illinois and conference participants at the Midwest Economic Association Conference, the LACEA's Health Network Workshop, WashU EGSC 2019, Illinois Economics Association, CSGEMP 2020, ASSA 2020 and the many comments received from several Professors and fellow students at UIUC.

1 Introduction

Although women around the world are increasingly older when they begin having children, it is estimated that the absolute global number of teenage pregnancies will nonetheless continue to increase until at least 2030 (Monteiro et al., 2019).[1] Policy makers, thus, have expressed concerns about the lasting implications of teenage pregnancy on human capital accumulation. The causal effects of teenage pregnancy on short- and long-term outcomes, however, are difficult to estimate. The main challenge is the potential for endogeneity around selection into early motherhood. Poorer human capital outcomes could be due to the causal impact of teenage pregnancy or because mothers who have teenage births are negatively selected into it, such that they would have had poorer outcomes regardless of their age at pregnancy. To overcome this concern, researchers who study teenage pregnancy in the United States and other high-income countries have taken several empirical approaches: propensity score matching (e.g., Levine and Painter, 2003 and Lee, 2010), with-in family differences (e.g., Herrera, Sahn, and Villa, 2019), and instrumental variables (Ashcraft, Fernández-Val, and Lang, 2013). However, the number of papers that utilize these methods to estimate the impact of teenage pregnancy in low- and middle-income countries is smaller.[2] Importantly, the impact of teenage pregnancy in low- and middle-income settings might be different due to context-specific characteristics.

In this paper, I examine the causal effects of teenage pregnancy on the educational attainment and labor-force participation of young urban women in a low-income setting, Cape Town, South Africa. To identify the effects, I use two main identification strategies: an instrumental variable identification and a sibling differences estimation. The data for this paper comes from the Cape Area Panel Study (CAPS; Lam, Seekings, and Sparks, 2006).

The CAPS data set is a longitudinal study of a randomly selected sample of young adults living in the Cape Town Metropolitan Area, the second-largest city in South Africa. The sampled individuals, aged 14 to 22 in 2002, were surveyed five times between 2002 and 2009 on demographic characteristics, sexual behavior, schooling, and employment. The first wave of the survey also included a set of retrospective questions from the year in which they were

[1]The average global teenage birth rate (aged 15 to 19 years) decreased from 65 per 1000 girls in the mid-1990s to 49 in 2011 (WHO, 2018 and Sedgh, Finer, Bankole, Eilers, and Singh, 2015). Yet, because the teenage population is growing is growing, the absolute number will grow.

[2]Exceptions include Branson and Byker (2018), Ranchhod, Lam, Leibbrandt, and Marteleto (2011), Ardington, Menendez, and Mutevedzi (2014), Urdinola and Ospino, 2015, Azevedo, Lopez-Calva, and Perova (2012), Narita and Diaz (2016) and Herrera et al. (2019).

born, allowing me to construct a panel in which the unit of observation is a women-year and a collapsed panel data structure.

I first instrument teenage pregnancy using the self-reported fertility status, measured by the time after the age at menarche (first menstruation). The approach was proposed by Ribar (1994) and Klepinger, Lundberg, Plotnick, et al. (1997) who estimate the effects of teenage childbearing in the United States, and later used by Field and Ambrus (2008) to estimate the effects of early marriage in Bangladesh.[3] The intuition behind the identification is that the variation in the timing of first menstruation between the ages of 10 and 17 generates quasi-random differences in the minimum age at which girls have the physical capability of getting pregnant. Importantly, the variation is independent of schooling investment, which allows for causal effect estimation. I find that each additional year that menarche is delayed decreases the likelihood that a teen will get pregnant between 2.4 and 3.2 percentage points.

An important concern in the instrumental variable identification is to what extent the age at menarche is independent of the confounders that affect educational, labor and fertility outcomes.[4] If so, the exclusion restriction would be violated. To examine this possibility, I first test the exogeneity of the instrument in the context of South Africa and consider the possible social consequences of the onset of menarche on educational outcomes. I also consider the exclusion restriction by utilizing event study analysis for school dropout rates.

The second methodological strategy I utilize relies on with-in family differences between sisters using sibling fixed effects in a subsample of the CAPS dataset. In this strategy, I compare a sister who reported a teenage pregnancy to at least one who did not.[5] Sisters are a reasonable counterfactual group because they share the same background and socioeconomic characteristics. If the variation in pregnancies is conditionally independent of unobserved familial differences, the estimation of the effects of teenage pregnancy will be unbiased.

[3]There is a small but growing literature studying the effects of teenage marriage in developing countries. Sekhri and Debnath (2014) and Chari, Heath, Maertens, and Fatima (2017) examine the consequences of female early marriage in India, a country where there is a high-prevalence of teenage pregnancy. Asadullah and Wahhaj (2016) consider differences in age of menarche between sisters to analyze the effects of early marriage on gender attitudes in Bangladesh. For Africa, Sunder (2018) and Hicks and Hicks (2015) utilize the same approach to examine teenage marriage in Uganda and Western Kenya.

[4]Khanna (2019) considers the effects of early menarche on schooling in India. Using a difference-in-difference approach, which takes advantage of the differences in the age at menarche within cohorts, her findings suggest that reaching the first menstruation before the age of 12 is associated with a 13.4-percentage-point decrease in school enrollment rate. The paper highlights the importance of examining the exclusion restriction when using the age of menarche as an instrument for gender issues since it points out that it may directly affect schooling. The setting in my study is different because it is urban and homogeneous.

[5]In this sample, sisters are defined as sharing at least one parent and report living in the same households.

Consistent with the notion that teenage pregnancy has a negative effect on education, women who report a pregnancy in their teens are more likely to lag behind their peers in their education and have lower school attainment. Using the instrumental variable strategy, I find that teenage pregnancy increases the likelihood of grade failure by 55 pp. and increases by 10% the risk of dropping out. Teenagers reporting a pregnancy lag of 0.3 years. The overall results suggest that women who report a pregnancy during their schooling years are, on average, 1.8 years less educated.

Similarly, using the sibling differences strategy, I find that women who report a pregnancy in their teens are 12.8% more likely to fail a grade and are 0.5 years behind their peers. The risk of dropping out increases by 49.4 percentage points. The school attainment effect is smaller in the sibling approach. Sisters who report a pregnancy are 0.774 years less educated, 44% (15.9 percentage points) more likely to sit for the matriculation exam, and 5 percentage points less likely to enroll in training or formal institution after the end of school.

The second group of outcomes studied in this paper describes the labor-force participation following the probable end of school. I study whether reporting an early pregnancy has an impact on the labor-force participation at the ages of 19, 20, 21, and 22. Intriguingly, across both strategies, I find small positive, though not statistically significant, effects. The findings are inconsistent with the evidence found for Latin America. Urdinola and Ospino (2015) find that teenage mothers have increased willingness to accept a lower paying job in Colombia while (Narita & Diaz, 2016); Arceo-Gómez and Campos-Vázquez (2014) and (Azevedo et al., 2012) find positive effects in Brazil and Mexico.[6] For Sub-Saharan Africa, although Branson and Byker (2018) do not find intensive margin estimates that are consistent with the these findings, the paper does find an increase in the monthly earning. The fact that these results are different from the previous literature highlights the importance of considering impact of teenage pregnancy within the specific contexts of the women.

Indeed, specific characteristics might cause the relative opportunity cost of teenage pregnancy to be different in South Africa than in other high-income countries. South Africa had extremely high youth employment rates (Nattrass & Walker, 2005), strong familial networks (Duflo, 2003 and Magruder, 2010); and extremely high rates of grade repetition (Anderson, Case, & Lam, 2001).[7] To test the possibility there are context specific attenuation effects, I use a control function approach in the instrumental variable strategy.

[6]One exception is Berthelon and Kruger (2017) who find a null impact on employment in Chile.

[7]As for the labor force, women were found to be less likely to be in the labor force and more likely to be unemployed at the turn of the century.

First, I consider the importance of child care within the households. A bigger network will change the cost of childcare and may change the relative cost of being enrolled in school. In particular, I study whether having a grandmother at home changes schooling attainment of the teenagers. The findings in this paper provide suggestive evidence that having a mother mitigates the effect of teenage pregnancy by 0.5 years. Next, South Africa is notable for its high rates of grade repetition despite its high schooling attainment. The class composition of grades is such that the spread in ages within grades is wider than in high-income countries (Anderson et al., 2001; Lam, Marteleto, and Ranchhod, 2013). The social stigma of teenage pregnancy might decrease and encourage students to stay in school. I find that attending a school with an above-average grade failure rate leads to a positive mitigation effect of 0.41 additional years of education.

Methodologically, this paper speaks to the literature of the effects of early pregnancy on short and medium term human capital investments.[8] The instrumental variables methodology exploits exogenous variations in teenage pregnancy to identify the causal effect. In this strategy, causal identification is made possible by considering factors that affect the likelihood of teenage pregnancy but are not directly linked to educational outcomes. The literature have utilized miscarriages (Hotz et al., 2005; Ashcraft and Lang, 2006; Fletcher and Wolfe, 2009; Azevedo et al., 2012 and Ashcraft et al., 2013), abortion laws (Bitler & Zavodny, 2001), the age at menarche (Ribar, 1994; and Ribar, 1999) and family planning policy changes (Almanza and Sahn, 2018 and Branson and Byker, 2018). The second approach I use, the with-in family differences strategy, has also been explored by the literature. The main papers that use this strategy typically yield educational effects that are between 0 and 1 fewer years of education (Geronimus and Korenman, 1993; Ribar, 1999; (Ardington et al., 2014); Berthelon and Kruger, 2017 Duncan, Lee, Rosales-Rueda, and Kalil, 2018 and Heiland et al., 2019). The magnitude of the effects is consistent with the magnitude found in this paper.[9]

My paper is closer to two papers that study the effects of teenage pregnancy in South Africa. Ranchhod et al. (2011) analyzed teenage pregnancy using a propensity-score match-

[8] In Table B3 in the appendix, I present a comprehensive review of the literature in high-income countries. See Ribar, 1999; Hotz, McElroy, and Sanders, 2005; Ashcraft and Lang, 2006; Diaz and Fiel, 2016 and Heiland, Korenman, and Smith, 2019 for reviews of the literature. For low- and middle-income countries, I present a complete review of the literature in Table B2.

[9] A third commonly used strategy was the propensity-score matching identification. Generally, these papers find modest negative effects on educational attainment and life satisfaction (Levine and Painter, 2003; Lee, 2010; and Zito, 2018).

ing strategy in the CAPS dataset. The evidence in this paper suggests a moderate impact of early pregnancy on high school completion and school attainment. Ardington et al. (2014) considers a similar research question in a rural area with a sister-differences approach. In this paper, I find that teenagers who report a pregnancy are more likely to drop out before the birth and, compared to their peers, do not lag behind in schooling.

This paper contributes to the literature in several ways. First, I add to the thread of literature that examines the effects of teenage pregnancy in low income countries using two novel approaches. Second, the rich data set allows me to study labor participation and training decisions after high school. The number of studies who consider the question for Sub Saharan Africa is low. Finally, an additional empirical contribution of this paper is that I take advantage of South Africa's unique characteristics to provide suggestive evidence of the factors that may help attenuate the negative effects of teenage pregnancy.

Understanding whether and how much of an impact teenage pregnancy has on educational attainment and early-life labor-force participation has important policy implications. Lower investments in the short run may affect women's long-run health and earnings as well as the well-being of their children. At the same time, the findings that stronger kinship networks and a reduction of the social stigma attenuate the effects of teenage pregnancy on education might inform policy makers who are concerned about human capital investments.

The rest of the paper is structured as follows. In section 2, I provide background information on South Africa and in section 3 I provide a detailed description of the chosen data set, the data structure, and summary statistics. In section 4 I discuss the non-evidence-based approaches used to consider teenage pregnancy and I outline empirical strategies. In section 5, I present the results of both sections. In Section 6, I examine the attenuating effect estimation. The final section concludes with a discussion.

2 Background and Setting

2.1 Sexual Debut, Early Childbearing and Marriage

Teenage pregnancy is regarded as a social issue that impacts families for generations. Understanding its determinants has, therefore, been an important concern across many settings. The debate is frequently centered around the perceived perverse incentives created by social benefits programs (Moultrie & Dorrington, 2004).[10] However, this discussion omits the importance of factors that are specific to teens in low income settings.

An earlier age at first sexual intercourse increases the likelihood of getting and transmitting sexually transmitted diseases, HIV, and teenage pregnancy (Marteleto, Lam, & Ranchhod, 2008) Teenagers usually posses low knowledge safe sexual practices and access to contraception is sometimes restricted by strict provision policies (Branson and Byker, 2018; WHO, 2018). Abortion services, for example, are less likely to be used by teens (Panday, Makiwane, Ranchod, & Letsoala, 2009).

Teenage pregnancy can also be a planned event. Women who get married early are often less able to effectively negotiate their sex practices and face pressure to have children (WHO, 2018). In Nepal, Unicef (2014) find that women who marry before the age of 15 are 33 percentage points more likely to have three or more children by the age of 24. Planned extramarital pregnancies have also been linked to improvements in the social status of teenagers in the lower socioeconomic strata in Brazil (Faisal-Cury et al., 2017).

2.2 The South African Context

Teenage pregnancy and schooling

Overall fertility levels in South Africa are low compared to other African countries (WHO, 2014). Teenage fertility rates, however, relatively high through the late 1990s and early 2000.[11] The proportion of women aged 19 who had reported a pregnancy in the 1998 and 2003 South African DHS was 35.1% (as seen in Figure 1)and 27.1%, respectively 2007).[12]

In the Cape Area, teenage pregnancy was estimated to be approximately 22% in 2002. The average is similar to the national level, which is 25%.[13] Interestingly, many South

[10] In the particular case of South Africa, President Jacob Zuma's 2008 election campaign included a proposal that teenage mothers be separated from their babies and forced to get an education (Ardington et al., 2014).

[11] Contrary to many other sub-Saharan countries, contraceptive usage in South Africa was high during the apartheid regime due to the government's plan to control the non-white population (Cooper et al., 2004). Public clinics, hospitals and mobile services largely provided contraceptives for free. However, not everybody benefited from this access to family planning, as evidenced by the surge in unintended pregnancy rates among adolescents. The findings point to social barriers to access to family planning for teenagers.

[12] Abortion was legalized in the country in in 1996. However, even when public and private facilities increased progressively, teenagers have not reported using pregnancy termination services.

[13] Compared to young adults in other regions, however, those who live in the Cape Area are 24% less likely to report a teenage birth than those in rural areas.

African mothers return to complete their schooling after giving birth because they receive support from their families and partners (Madhavan and Thomas, 2005).

General education in South Africa is divided into three periods: primary, middle school and secondary school. Schooling is compulsory until grade 9, and spans 12 grades in total. Literacy and numeracy are taught in primary school, which lasts six years. Then, middle school is three years long, at the end of which students get a basic education-and-training certificate. Secondary school spans grades 10 to 12.[14] In order to continue on to higher education, students have to sit for a matriculation exam.

School enrollment rates are high in South Africa as primary school attainment is nearly universal and secondary school enrollment remains high through the teenage years (Anderson et al., 2001; Lam et al., 2013). There are striking gaps in knowledge gains across students, however, which is shown by poor national standardized test performance compared to other countries (Van der Berg, Louw, et al., 2007).

Lam, Ardington, and Leibbrandt (2011) study differences in grade repetition in the CAPS data set for the period following the education reform that happened in 1994. Out of all of the students enrolled in grades 8 or 9 in 2002, 82% of white students successfully advanced three grades by 2005, whereas only 34% of coloured students and 27% of black students attained the same level of advancement.[15] Two thirds of the students, however, chose to continue their schooling despite this fact. Coloured teenagers are enrolled in secondary school after the age of 20 (Anderson et al., 2001; Marteleto et al., 2008).

Labor Market Characteristics and Youth Unemployment

A key consequence Youth unemployment was structural in South Africa at the time of the study. According to the September 2003 Labour Force Survey, the unemployment rates range between 28% and 42% depending on whether discouraged job seekers are included or only active ones are (Nattrass & Walker, 2005).The rate was concentrated among the young and among blacks, followed by the coloured population (Magruder, 2010).

There are two outstanding facts about South Africa's labor market. First, formal sector wages were high at turn of the century, compared to similar countries.[16] As a consequence

[14]The emphasis is placed on academics and vocational training.

[15]In South Africa, the term "coloured" refers to people of mixed ethnic ancestry, including Khoisan, Bantu, Afrikaner, Whites, Austronesian, East Asian, and South Asian. I keep the term in its South African spelling.

[16]Historically, the trend began under the apartheid regime, when the system was designed to increase

of the high pay-off of formal sector jobs, seekers had longer search spells, and the cost of unemployment shrank (Magruder, 2010). Second, the informal sector is small compared to other countries with similar development levels. In rural areas, agriculture is mostly undertaken in large farms that were established before apartheid, such that there is no subsistence agriculture in rural settings. In more urban settings, the informal sector is relatively small compared to similar countries.

Women are less likely to participate in the labor market and more likely than men to be unemployed (Banerjee et al., 2008). Female labor supply increased drastically in the early 2000s but the demand did not accompany the influx as employment in south Africa's bigger industries, agriculture and mining employment steadily fell.

3 Data

3.1 Cape Area Panel Study

To study the effects of teenage pregnancy on education, and labor supply, I use the Cape Area Panel Study (Lam et al., 2006). This is a longitudinal study, which follows young men and women who lived in the Cape Town Metropolitan Area in 2002.[17] Individuals were sampled using a stratified two-stage sample of households, from sample clusters first and then through households within these clusters. Clusters were selected according to the breakdown of ethnic groups available in 1996 census, where white and black clusters were oversampled to achieve a representative sample (Lam et al., 2006).

Five waves were conducted to create the CAPS data set: first in 2002, then in 2003–2004, in 2005, in 2006 and in 2009. The first wave, administered in August-December of 2002, surveyed 2,612 young women aged 14 to 22 in 2002 in 2,045 households. This paper draws heavily from the young-adult questionnaire in wave 1. The young-adult questionnaire includes questions on race, religion, place of birth, and education of the parents. Wave 1 also includes the complete pregnancy and birth histories, prenatal care, and partners' informa-

white people's wages in capital-intensive production industries and continued for several years (Nattrass and Walker, 2005 and Banerjee, Galiani, Levinsohn, McLaren, and Woolard, 2008). Unionization structure generated high minimum wages in all firms in many industries and later spread to all industries not covered by union arbitration.

[17] Cape Town is the second largest city in the country after Johannesburg, with 2,785,032 inhabitants, and it is the provincial capital of the Western Cape. It is located in the southwestern corner of South Africa.

tion. This information is updated in the subsequent surveys. The self-reported information on pregnancy provides the basis for the early pregnancy history of the older cohorts, who are older than 18. Next, the age at menarche was first asked in 2002.[18]

In wave 2, the total number of sampled women is 2,140 (748 were interviewed in 2003 and 1,410 in 2004). Pregnancy questions and birth history are not included, but the surveys inquire about reasons for not attending school. I am thus able to determine whether women reported pregnancy during those years. Wave 3 includes questions on schooling, employment, fertility, and personal health to 1,911 women now aged 17 to 25 in 2006.[19]

Wave 4 includes questions on the school, work, and childbearing histories of CAPS young adults to 1,877 individuals.[20] An important feature of wave 4 is that it includes a set of health outcomes for young adults and their parents, which limits the sample size. Finally, wave 5 includes a young-adult questionnaire, a young-adult telephonic questionnaire, and a young-adult proxy questionnaire to 1,799 women, aged 21 to 29.

3.2 Variable Construction

This paper takes advantage of the data available in the various waves and retrospective information questions asked in the first CAPS wave to conduct the analysis. Using the available information described in Table B1, the analysis is undertaken in samples with one of these two formats: (I) a panel data set in which each observation represents one individual-year observation and (II) a collapsed panel in which the information is aggregated at the sampled-woman level. I use both the panel and the collapsed-panel samples to conduct the instrumental variable analysis and the sibling differences approaches. In the next sections, I explain how I construct the key variables required for both approaches.

Teenage Pregnancy and Age at Menarche

I begin by providing a description of how the likelihood of a pregnancy and the fertility variables are created for the panel and the collapsed panel samples. First, teenage pregnancy is measured differently at the aggregated level and at the year level. In the panel sample, I define the variable "Pregnant_{isjt}" as an indicator for whether sampled woman i who lives

[18] Almost all sampled women, 96.8%, provided the age in this wave.

[19] Particularly useful for this study was the update on age at menarche for those who did not answer the question in the first wave.The question extends the age at menarche to 103 additional women.

[20] The targeted populations are young adults, their biological children, and older residents (age 50+).

in sampling cluster s of cohort j reported a pregnancy in year t. It describes reports from 1990 until 2009. An important exception is the year 2008, when it is not possible to narrow pregnancies using the fifth wave of the CAPS data.

In the collapsed panel, I define "$Pregnant_{isj} < 18$" as an indicator variable for whether the sampled woman i who lives in sample s of cohort j reported a pregnancy before the age of 18. Figure 2 describes the percentage of women who report a pregnancy between the ages of 14 and 26. As seen in the figure, the percentage of women who report a pregnancy increases significantly until the age of 18 and then flattens out. The percentage of sample women who experienced a teenage pregnancy in my sample is 20%.

Because women become fertile after they reach menarche, the age at menarche is an important threshold for my study. As described in Table B1, the question regarding the age at menarche was asked in wave 1, and later re-asked in wave 3 for those who reported have not reaching it at the time of the first interview.[21] Figure 3 shows the distribution of the age at menarche for the analytical sample of women. The average age at menarche among all of the women studied in this paper is 13.445 years of age.

Empirically, the fertility variables are created differently in the panel and the collapsed panel. In the former, fertility is captured by "$Post\ Menarche_{it}$" which is a variable that is equal to one for the years in which the sampled woman is at least her age at menarche, and zero otherwise. Alternatively, in the collapsed panel, I measure how many years a sampled woman pass between menarche and the age of 17 ("$Fertile\ Years_{ics}$").[22]

Education Outcomes

The CAPS data set includes a set of questions on whether individuals were enrolled in any schooling institution and whether they dropped out before completing a grade, failed, or passed each year. Thus, by combining grade data with year data, I study two important sets of outcomes: (I) the sampled women's progression through their stages of schooling and, (II) the sampled women's overall educational attainment.

In the panel analysis, I consider the likelihood of failing the grade (conditional on being enrolled in a particular year). "$Failure_{it}$" is an indicator variable that is equal to one if

[21]The script for the question reads as follows: "As girls begin to mature into women, certain changes occur in their bodies, such as the start of menstrual periods. At what age did you have your first menstrual period or have you not had one yet? (Please look at the calendar, if that will help you remember.)"

[22]Mechanically, this is equal to 17 minus the age at menarche.

woman i reports having failed a grade in year t or having dropped out midyear, and zero otherwise. Second, "$Grade\ for\ age_{it}$" captures the lag in education until grade 12.[23] Third, "$Drop\ Out_{it}$" is defined for woman i as one for those who report not being enrolled in schooling in year t, and zero if they were.

I study educational attainment in the collapsed panel. "$Years\ of\ education_i$" is a continuous variable measuring years of education completed during the last observed period. Next, I present two indicator variables: "$Took\ Matric_i$" (or NSC exam) is equal to one if the individual sat for the NSC and "$Post\ secondary\ Education_i$" is equal to one if the individual kept studying after high school.

Labor Force Participation Outcomes

I study the effects of teenage pregnancy on their labor-force participation in the static model. Specifically, I consider the labor-force participation for ages 19, 20, 21, and 22.

3.3 Two samples

The analytical "full sample" is comprised of 1,741 women in two sample formats: a panel and a static samples. The former is constructed using a combination of the retrospective self-reported information available in the first wave, and updated using the questions asked in the following waves. The resulting data follows the sampled women from birth to 2009 such that the unit of observation is at the woman-year level. Since the goal of the paper is to study the effects of teenage pregnancy, the analysis is conducted between the ages of 10 and 20. Second, for the static analysis, I collapse the information into a "collapsed panel" where each observation is one sampled woman.

Selection of the women is based on three main criteria: age at menarche, health module availability, and sampling location. I limit my study to women who have undergone menarche between the ages of 10 and 17, have health information, and live in clusters where there is at least one other person in the cluster.

First, medical researchers have established that menarche is delayed when girls report its onset two standard deviations (years) after girls of similar background (Hillard, 2013). Among the women in the full sample, the average age at menarche for those who report the

[23] I follow (Glynn et al., 2018) by setting a measure of how much behind in school and censoring the lags up to two grades behind.

onset was 13.433 (1.667 SD). I thus analyze women who have undergone menarche until the age of 17. The percentage of women reporting reaching menarche after that age, however, is small: only 1.32%.it is chosen as the lower bound for the age at menarche in this study. The sampled women interviewed in the first wave who report reaching menarche before this age is small (0.85%).The percentage of women who fell outside the bound amounts to 2.17%.

Two additional requirements relate to health and sampling clusters. The health module was only asked in wave 4, to a total of 1,790 out of the 1,851 women who were surveyed in 2006 with in the chosen age at menarche range. Finally, the sampling location limitation requires women living in the same sampling cluster with at least one other woman in it. The last criteria censors an additional 1.8% of the women sampled in wave 1.

A second the "sister sample" is comprised of women who share a parent, report living in the same household, and have differing teenage pregnancy reports. Given the criteria, the sample of siblings is small, 418 women, and they come from bigger households than the composition in the full sample.

3.4 Summary Statistics

In Table 1, I provide summary statistics for the sampled woman in the full sample and in the sisters sample. The two samples are different in a number of key characteristics.

The average age in 2002 of the all of the women studied in this paper is 17.723. The sisters in the comparison data set are slightly younger, their average age is 17.902. 46.1% of the individuals are coloured and 48.6% are black; whereas in the sibling sample, the percentages are 40.9 and 56.7, respectively. The education of the mothers of the women is lower in the sister sample, as they are on average, 0.3 years more educated than in the sister sample (8.271 vs. 8.773 years). In contrast, as can be seen in Table 2, because the sample of siblings belong to households where there are at least two young adults in 2002, the average household size is greater by 0.54 individuals (5.808 versus 6.348). By construction, the women in the sister sample come from bigger households and have more full siblings: the number of full siblings is 2.312 in the full sample, but it is 2.685 in the sibling sample.

Given the differences in the composition of the sister sample, the estimation of the effects of teenage pregnancy will yield different results from those found by the instrumental variable approach. First, the sample of sisters is selected to include women from poorer and less educated households, rendering smaller effects. Second, most of the non-teenage mothers in the sample were sharing a household at the time of the pregnancy or birth. Educational attainment might thus be hindered by the presence of the child in the home. In this case, the analysis of the impact of pregnancy would find smaller effects than on the full sample.

4 Empirical Methodology

Researchers have used several approaches to deal with the issue of selection into motherhood when estimating the casual effect of teenage pregnancy on educational attainment. The main goal of choosing a strategy is to minimize bias. In this section, I discuss the two methodologies utilized in this paper, and the potential threats to the validity of the identification strategies.

4.1 Correlation between Teenage Pregnancy and the Outcomes

The main relationship in this study, the structural relationship between teenage pregnancy and the chosen outcomes, is the presented in Equation 1:

$$Outcome_{itc} = \beta_1 + \beta_2 Pregnant_{itc} + \varepsilon_{itc} \tag{1}$$

In this equation 1, $Outcome_{itc}$ is the outcome of woman i observed in year t who lives in cluster c. Pregnant_{itc} is an indicator variable that denotes whether woman i was pregnant in year t. In this equation, β_2 describes the relationship between the educational outputs and teenage pregnancy. I also control for a vector of age fixed effects to reflect the fact there are different probabilities of pregnancy occurring at different ages. ε_{ikc} is the error term.

Nevertheless, teenage pregnancy might be endogenous if teenage mothers come from backgrounds that are different from non-mother teenagers. In this case, Equation 1 might not account for the fact that there might be observed and unobserved confounding factors that affect teenage pregnancy and the chosen schooling, labor supply, and fertility outcomes. To illustrate this point, in Table A2 I provide the differences in a set of demographic characteristics between the women who report a teenage pregnancy and those who do not.

Women who report a teenage pregnancy are significantly different in age in 2002 from non-teenage mothers in a number of significant points. The race composition is different between these groups. Teenage mothers are two centimeters shorter, less educated (0.9 years) and live in households that are 0.57 greater in size than teenagers who are not mothers.

As a consequence of the differences established in Table A2, the coefficient of interest, β_2, in Equation 1 will be biased. Hence, the estimation of the causal effects of teenage pregnancy requires the use of strategies that reduce the selection bias. I thus turn to an explanation of the instrumental-variable and sibling-differences approaches taken to address endogeneity.

4.2 Instrumental Variable Approach

I first examine the instrumental variable econometric specification, which exploits exogenous variations in teenage pregnancy to identify the causal effect of teenage pregnancy on schooling and labor outcomes. The identification is made possible by considering factors that affect the likelihood of teenage pregnancy but are not directly linked to educational outcomes.

In the presence of heterogeneous treatment effects, the instrumental variable estimation coefficients identify the local average treatment effect (LATE), capturing the average effect for those induced to report a pregnancy in their teens by reaching menarche earlier. That is, the instrument captures the effects of teenage pregnancy through the compliers.[24]

4.2.1 Panel Dataset

I estimate the following two-stage instrumental variable model:

$$\text{First Stage: } Pregnant_{icst} = \sigma_1 + \sigma_2 Post\ Menarche_{icst} + \beta_3 X_{icst} + \varsigma_t + \lambda_i + \epsilon_{icst} \quad (2)$$

$$\text{Second Stage: } Outcome_{icst} = \beta_1 + \beta_2 \widehat{Pregnant_{icst}} + \beta_3 X_{icst} + \varsigma_t + \lambda_i + \varepsilon_{icst} \quad (3)$$

In the first-stage equation, 2, $Pregnant_{icst}$ is an indicator variable that denotes whether the sampled woman i of cohort c from the sampling cluster s reported a pregnancy in year t. $Post\ Menarche_{icst}$ indicates whether individual i was fertile in that year. In this equation, ϑ_t, and λ_c denote time and individual fixed effects, respectively. The X_{icst} is a set of time varying controls such as the age in year t. Finally, ϵ_{itc} is the error term.

[24]The instrumental variable methodology has some important caveats. It is often difficult to find a strong instrument and that the resulting estimate can be imprecise. Duncan et al. (2018), for example, chose a sibling-and-cousin fixed-effects methodology for the impact of maternal age on child development in the United States. Using the same data, this paper also finds that the age at menarche is a weak instrument for maternal age at first birth. Additionally, the authors suggest that miscarriages (Hotz et al., 2005) and state abortion laws (Bitler and Zavodny, 2001) are stronger predictors of teenage pregnancy, but resulted in standard errors that were too large and unable detect significant effects.

In the second-stage equation, $Outcome_{itxc}$ is the educational, labor-supply, or fertility outcome estimated for the sampled woman i of cohort c from the sampling cluster s, who reported a pregnancy in year t. The coefficient of interest, β_2, describes the relationship between the chosen outputs and teenage pregnancy. The controls and fixed effects are the same as in the first-stage equation. The error term is represented by ε_{icst}.

Hazard Estimation

To estimate the likelihood of dropping out, I utilize the same instrumental variable identification strategy in a hazard model which can be described in the following equation:

$$Drop\ Out_{icst} = \beta_1 Pregnant_{icst} + \beta_3 X_{icst} + \varsigma_t + \lambda_i + \epsilon_{icst} \tag{4}$$

In Equation 4, $Dropout_{icst}$ is coded as an indicator variable for whether individual i dropped out in that period. The reminder variables in Equation 4 are defined in the same way as in Equation 3. Also in this equation, h_o represents the baseline risk, defined as defined as $h_o(t) = pt^{p-1}$. The Weibull model estimates an expected survival time.[25]

To address endogeneity in the context of the hazard function, I take a control-function approach in which the first stage is estimated linearly (Wooldridge, 2015). The variance-covariance matrix is also corrected to accurately estimate the standard errors. The first stage allows me to generate a predicted likelihood of teen pregnancy that is used to tease out the confounding factors of teenage pregnancy.

4.2.2 Collapsed Panel

In the Collapsed Panel sample, the two-stage instrumental variable is estimated as follows:

$$\text{First Stage: } Pregnant \leq 18_{ics} = \sigma_1 + \sigma_2 Fertile\ Years_{ics} + \sigma_3 X_i + \vartheta_s + \lambda_c + \epsilon_{ics} \tag{5}$$

$$\text{Second Stage: } Outcome_{ics} = \varphi_1 + \beta_2 \widehat{Pregnant \leq 18_{ics}} + \beta_3 X_i + \vartheta_s + \lambda_c + \upsilon_{ics} \tag{6}$$

In Equation 5, $Fertile\ Years_{ics}$ is defined as a continuous measure of the number of fertile years between the age of menarche and the age of 17 for a sampled woman i of cohort c who

[25] Women who drop out before the age of 10 are left censored and women who drop out at the age of 20 are right censored.

lives in the sampling locations.[26] $Pregnant \leq 18_{ics}$ is equal to one if the same individual reported a pregnancy before the age of 18 and zero otherwise. $Outcome_{ics}$ is the educational or labor outcome estimated. The coefficient of interest, β_2, describes the relationship between the chosen outputs and teenage pregnancy.

X_i is the set of individual controls including height, literacy status of the mother of the teenage mother, language spoken at home, race, whether the woman was born outside the Western Cape, her self-reported religion, the normalized literacy exams, number of full siblings, and asset characteristics index. The error term is represented by ε_{ics}.

The inclusion of the sampling location fixed effects (λ_c) accounts for the adverse events that may have occurred at the year of birth of the sampled women that will also be linked to both the age at menarche and the outcomes of interest. They are included in order to account for any geographical condition that may affect the age at menarche.

The key identifying assumption in this methodology is that, within a given cohort year, location, the demographic and socioeconomic status controls, and the age at menarche are orthogonal to potential confounders. The impact of teenage pregnancy on the probability of lagging behind in schooling is identified from within-location and within-year-of-birth variation in the age at menarche. I address potential concerns in Section 4.2.4.1.

I also generate an alternative specification for the instrumental variables. To allow for non-linearities in the effect of menarche on teenage pregnancy, I specify it as an indicator for an early age at menarche. The threshold for a late menarche is defined as girls who have reached menarche after the age of 15, which is a slightly more conservative definition than the one used by the literature (Lawn, Lawlor, and Fraser, 2017; and Kim and Je, 2019).

4.2.3 First-stage Estimation

In this section I present the first-stage coefficients and a graphical representation of the relationship between the fertility and pregnancy indicator variables in the panel, and between the number of years post menarche and the indicator for teenage pregnancy, in the collapsed panel.[27] The coefficients for Equations 2 and 5 are presented in Table 2.

First, Columns (1) and (2) includes the coefficients for the panel specification with and

[26] Mathematically, I define this continuous variable as 17 minus the age at menarche.

[27] The approach requires strong positive correlations: $Corr(Pregnant_{icst}\ ;\ Post\ Menarche_{icst}) \neq 0$ in Equation 2 and $Corr(Pregnant \leq 18_{ics}\ ;\ Fertile\ Years_i) \neq 0$ in Equation 5.

without controls. The coefficient of interest, σ_2, is equal to 0.023 (0.002 SE, pp<0.01) and 0.024 (0.002 SE, pp<0.01), respectively. The first-stage F-Statistics are 200.6 and 171.6.

Columns (3) and (4) describe Equation 5. With out the controls, or fixed effects, σ_2 takes the value of 0.012 (0.006 SE, p< 0.1) but the instrument is weak, as the F-statistic is equal to 3.771. However, with the inclusion of both controls, the coefficient is 0.032 (0.008 SE, p< 0.01). The F-statistic is equal to 17.61. An additional fertile year during the teenage period increases the probability of reporting a pregnancy before the age of by 3.2 percentage points. A graphical representation of the relationship is presented in Figure 5.

Columns (5) and (6), present the first stage coefficients for the indicator variable for the onset of menarche with and without controls. In column (5), φ_2 is equal to 0.045 (0.018 SE, p< 0.01) and the F-statistic is 5.956. Nevertheless, once the controls are included, reaching menarche before the age of 14 is associated with an increase in the likelihood of adolescent pregnancy of 8.1 percentage points (0.022 SE, P<0.01) and the F-statistic is 13.43.

In Tables A6 and A7 in the appendix, I present two variations of the identification strategy for the collapsed panel instrumental variable approach. First, in Table A6 I modify the definition of teenage pregnancy to the ages of 16 and 21. Younger ages are associated with greater coefficients and stronger F-statistics. In Table A7, I present three variations of the instrument. Instead of utilizing 17 minus the age at puberty, I change the threshold to 18, 19 and 20. The coefficients are similar in size and F-statistic is greater than 10.

The results across all specifications describe a positive relationship between the number of fertile teenage years and teenage pregnancy. Each additional fertile year during a woman's teenage years increases the likelihood that a sampled woman will report a teenage pregnancy. Table 4 clearly describe the positive sign described by the coefficients of interest in both samples.[28] In the next section, I turn to issues associated with the validity of the instrument.

4.2.4 Validity Issues

4.2.4.1 Validity of the Instrument

Since the identification strategy relies heavily on the correlation between age at menarche and likelihood of teenage pregnancy, I examine the instrument's validity. Importantly, for the instrument to successfully tease out endogeneity, the exclusion restriction requires that

[28]The pattern is consistent with previous findings (for example, Ribar, 1994 and Klepinger et al., 1997).

the relationship between teenage fertility and adult outcomes is fully mediated by changes in the likelihood of teen pregnancy.[29] There are, however, two important concerns regarding the exclusion restriction of the age at menarche as an instrument for the likelihood of teen pregnancy. The first relates to the exogeneity of the instrument, and the second concerns the social consequences of the onset of menarche.

Exogeneity of the age at menarche

The medical debate over whether the age at menarche is an exogenous event is voluminous. There is a large body of evidence suggesting that long-run health is determined by shocks that happen in the prenatal and perinatal period (Almond, Chay, and Lee, 2005; Alderman, Hoddinott, and Kinsey, 2006 and Almond and Mazumder, 2011) and in early life (Akresh, Bhalotra, Leone, and Osili, 2012; Mahmud, Shah, and Becker, 2012 and Almond, 2006). The debate also has implications for whether early childhood nutrition determines the age at menarche. There is a strand of the literature that argues that a random genetic component explains the age at first menstruation (Mpora et al., 2014; Jahanfar, Lye, and Krishnarajah, 2013, Sørensen et al., 2013, Adair, 2001), which would suggest that the age at menarche is quasi-random. In contrast, a group of authors have found that early childhood nutrition and socioeconomic status are linked to the timing of puberty among girls (Karapanou and Papadimitriou, 2010; Dahiya and Rathi, 2010; and Rah et al., 2009). In this case, childhood nutrition may affect long-run well-being through human capital development or through its impact of the timing of menarche.[30]

To address the debate over the exogeneity of the age at menarche, I proceed in three ways. My first approach is to examine an indicator of adult health that has been linked to prepubescent health status and nutrition: adult height (Martorell and Habicht, 1986; Fogel, 1993; Silventoinen, 2003). Since in 2006 the younger cohort sampled in the CAPS Data set was 18 years of age, this height variable is the best available proxy of degree of stunting caused by poor nutrition or health issues in childhood. Figure 6 displays the relationship

[29] It is worth mentioning the monotonicity condition. Given the set-up of the study, girls who experience a late menarche, even when using the chosen conservative definition, are very unlikely to experience a pregnancy during their schooling years. The decrease is largely due to the biological impossibility of having a child without undergoing menarche. It is thus possible that the setting rules out the existence of potential defiers. The monotonicity property of the age at menarche as an instrument for teenage pregnancy is therefore satisfied, if we account for the health shocks.

[30] More recently, Khanna (2019)'s main corollary is that the age at menarche is a poor instrument for the age at marriage in India because it does not satisfy the exclusion restriction. Using data from the Young Lives panel, reaching menarche early (defined as before the age of twelve) decreases school enrollment by 13%. It has been argued that nutrition is linked to the timing of menarche among girls.

between age at menarche and adult height. The relationship is weakly positive but not statistically significant. Following Field and Ambrus (2008), Sekhri and Debnath (2014) and Chari et al. (2017), I include the height of the women among the vectors of controls to account for any remaining within-sampling location variation in environmental conditions.

A further test of the exogeneity can be found in Table A3. In this table, I present some descriptive characteristics for girls who have reached menarche by the age of 14, and those who report reaching it later. As seen, most of the demographic characteristics remain balanced across groups. The race composition and the education of the mothers, however, are different among those who reached menarche in different age groups.[31]

Second, I examine whether environmental factors, such as toxins that are specific to a location, cause delays in menarche.[32] The CAPS data is mostly homogeneous: 80% of the population report having lived most of their lives in a formal or informal urban setting. I control for sampling location fixed effects in the static model and test whether the sampling clusters faced different environmental factors. This relationship is not statistically significant.

A third feasible concern would be the possibility of recall bias. The age at menarche is established using a self-reported account of the year in which the sampled women reached menarche. In cases of misremembrance, the first-stage coefficients would not correctly estimate the likelihood of teenage pregnancy. Although unlikely, given the nature of the event in a girl's life, I examine this possibility in Figure 7 where there is a graphical representation of the distribution of the responses to age at menarche in waves 1 and 3. As seen in the figure, there are no differences in the distribution.[33]

Exclusion Restriction

The the onset of menarche could lead girls or their families to engage in actions that might affect the girls' schooling progression. If so, the exclusion restriction would be violated. The

[31]In Table A9 in the appendix, I present the estimates for the largest race group in my sample: black women. This group is driving the results.

[32]This concern is particularly important because girls raised in urban environments have been found to display earlier age at menarche than those who grew up in rural environments.

[33]To further test the possibility of recall bias, I compare the average age at menarche for the women in my study sample (13.392 years old) with other similar countries in Africa. Sunder (2018) and (Prentice, Fulford, Jarjou, Goldberg, & Prentice, 2010) find that the average age at menarche for Ugandan and Gambian girls was 14.4 and 14.90 years, respectively. In Nigeria, (Adebara & Munir'deen, 2012) estimated that girls reached menarche at the age of 13.21 and in Mozambique Padez (2003), the average amounted to 13.2. The estimates are similar to the average in the current study sample, which further lessens the concerns of inaccurate reporting of the age at menarche in the CAPS data set.

concern is particularly important in many low-income countries because the lack of knowledge of menstruation within families results in the exclusion of menstruating women and girls from daily routines or public spaces. If so, these girls face additional challenges associated with menstrual management and/or public shaming.

To test for the potential consequences of the onset of menarche, I undertake an event study analysis, as shown in Figure 8. The event is defined as the year in which the sampled women reached menarche. The selected variables are the pregnancy, dropping out of school, the interaction between dropping out and pregnancy, and the interaction between dropping out and not being pregnant. The estimation includes five lags and five leads and controls for age, time, and year fixed effects.

By definition, teenage pregnancy is virtually zero before reaching menarche and there is an upward trend the year after its onset. The event study for dropout rates show that dropouts did not increase significantly until two years after the onset of menarche. At the same time, dropout does not react in the four periods prior to the onset of menarche. In the left bottom panel, the analysis of the interaction of teenage pregnancy and dropout indicates a similar pattern. Finally, the pattern of the interaction of dropping out and being pregnant increases after the onset of menarche, but the coefficients are not statistically different from zero for the first four lags.

4.2.4.2 Attrition

Attrition rates are important in the CAPS dataset. As shown in Table A4, the percent of women in wave 4 represents 72% of the women originally interviewed and only 70% of women with age at menarche form the full sample. It is thus important to consider whether the teenage pregnancy or age at menarche are different between my sample and the individuals not considered in the paper. If women interviewed in the first wave but not in the forth are more educated, then the estimated found in this paper would be biased upward.

An important consideration in this study is the overall demographic characteristics of individuals who drop from the sample. The attrition rates differ significantly by race. The black population attrition rate is 20%. Marteleto et al. (2008) explain that this attrition level is due to return migration to the rural Eastern Cape province. The colored population attrition level is 10%. Finally, attrition is higher among older cohort groups, which is indicative of a positive association between age and attrition in the CAPS data.

To test whether there is selective attrition, I present the coefficients of regression where the left-hand side variable is an indicator variable equal to one if the woman was not included in the full sample and zero otherwise, and the right hand side is comprised of indicator variable for the age at menarche or the indicator variable for pregnancy before the age of 18. These estimates are presented in in Table A4.

In columns (1) and (2), the main regressor is the age at menarche. In column (1) the estimation does not include controls and in column (2) the estimation includes indicator variables for race, cohort, religion, household size, the literacy level of the mother of the teenage mother, the teenage mother's place of birth, and the household size. Column (2) also includes sampling cluster fixed effects. Columns (1) and (2) provide evidence that attrition is not statistically associated with the age at menarche. Both coefficients are negative, close to zero and not statistically significant.

In columns (3) through (6), the regression includes an indicator variable for pregnancy before the age of 18. As with the age at menarche, I analyze the relationship between attrition and teenage pregnancy with and without controls and sampling cluster fixed effects. The sample in columns (3) and (4) is composed of all of the women surveyed in the first wave. The regressions presented in columns (5) and (6) limit the sample to those who were already 18 years of age in the first wave because they it was already defined whether they were teenage mothers in 2002 before dropping out of the study.

The teenage pregnancy coefficient in column (3) is equal to -0.088 (statistically significant at the one-percent level). However, when controls and fixed effects are added, the association between teenage pregnancy and attrition diminishes to 0.044 and is no longer statistically significant. Similarly, when the sample is restricted to women who were older than 18 in 2002, the coefficients are smaller but are only significant when additional variables are included. The information available in Table A4 suggests that teenage pregnancy is not a key determinant of attrition.

4.3 Sibling differences Approach

The second empirical approach utilized in this study to reduce selection bias by comparing outcomes among family members. The key assumption is that there are factors that are shared by members of the same family but are unobservable to the researcher. The inclusion of sibling fixed effects thus allows me to control for common socioeconomic factors (such as genetics, school quality, and economic resources). I then test whether women who share the same background but have different pregnancy outcomes perform differently in their educational paths. If this variation is conditionally independent of unmeasured within-household differences, this would also affect the outcomes, and thus the estimate is unbiased.

4.3.1 Panel Data Estimation

Equation 7 describes the econometric specification for the panel dataset:

$$Outcome_{icht} = \varphi + \varphi_2 Pregnant_{icht} + \varphi_3 X_i + \psi_h + \varsigma_c + \varepsilon_t + \upsilon_{itjh} \tag{7}$$

In equation 7, $Outcome_{icht}$ represents the outcome of the sampled woman i of cohort c, who lives in household h and is observed in year t. Also, $Pregnant_{icht}$ indicates whether that same woman got pregnant in year t. X_i is a set of individual-level adult health controls including individual i's height and normalized grade on the literacy exam.

Additionally, ψ_h is the sibling fixed effect, ς_j is the birth year fixed effect and ε_t are time fixed effects. Finally, υ_i is the error term. In Equation 7, φ_2 represents the coefficient of interest, which captures the effect of individual i reporting a pregnancy in year t.

4.3.2 Collapsed Panel Data Estimation

I turn next to the Collapsed Panel estimation in Equation 8:

$$Outcome_{ich} = \varphi_1 + \varphi_2 Pregnant \leq 18_{ijh} + \varphi_3 X_i + \psi_h + \varsigma_c + \upsilon_{ijh} \tag{8}$$

Where $Outcome_{ijh}$ represents the outcome of interest of individual i, of cohort c and who belongs to household h. Also, $Pregnant \leq 18_{ijh}$ is an indicator variable for whether the same individual reported a pregnancy before the age of 18. X_i is a set of individual level adult health controls including the individual i's height. Also, in Equation 8, ψ_h is the sibling fixed effect and ς_j is the birth year fixed effect. Finally, υ_{ijh} is the error term. φ_2 is the coefficient of interest and measures the effect of an early pregnancy for woman i.

Siblings may have been exposed to different geographical and environment characteristics. I thus control for the potential effects of the environment by including fixed effects for the respondent's region of birth and sampling clusters.

4.3.3 Validity Issues

A key undertaking in this methodological approach is that it controls for unobserved factors using sibling fixed effects. In sibling differences strategy the variation that identifies the effects of teenage pregnancy is derived from families in which at least one sister reported a teenage pregnancy and at least one did not. The variation is independent (conditionally) of unmeasured sibling differences in order for the estimates to be unbiased.[34]

An important concern, thus, is that sisters who grew up together but at different points in time may have been exposed to different shocks that affect nutrition and stress levels differently. In such cases, I would expect significant gaps in sisters' age at menarche. To test whether sisters are similar, I look at intra-cluster correlation within the sister sample. The correlation coefficient for sisters in the sibling sample is 0.742 (0.035 se) which suggests that sisters have very similar age at menarche. To investigate the possibility, I present the variation in the age at menarche by the respondent's birth order in the sample in Figure 9.

5 Educational Results

5.1 Schooling Progression

Table 3 reports the effects of teenage pregnancy on schooling progression. The top panel displays instrumental variable estimates for the effect of early pregnancy on the likelihood of failing a grade, the age-for-grade measure and the dropout risk; the bottom panel shows the reduced-form estimation.[35] The effects of self-reported teenage pregnancy between sisters are described in Table 4.

In the instrumental variable approach, the estimation of the effects of teenage pregnancy renders a coefficient of 55.7 (0.028 SE, $p<0.01$) percentage points decrease in the probability of grade failure. In contrast, in the sibling difference the coefficient for the same outcome is smaller: 13.2 (0.037 SE, $p<0.01$). Teenagers who report a pregnancy also lag behind

[34] The manner in which the sibling identification strategy deals with family is to control with fixed effects. Hence, what this strategy does not do is to control for individual characteristics, that is, characteristics not shared by sisters. However, since sisters resemble each other in their age at menarche, the variation is small. This fact, combined with the relatively small size of the 39-sister sample, renders instrumenting teenage pregnancy using the age at menarche among sisters not possible. The first stage is weak.

[35] In Table A8 in the appendix, I extend the sample until the age of 24.

their peer by 0.284 years (0.010 SE, p<0.01) in the instrumental variable strategy and by 0.521 years (0.122 SE, p<0.01) in the sibling differences approach. Finally, the coefficient for the likelihood of dropping out is equal to a 0.097% in the instrumental variable strategy. However, it more than (p <0.01) in the sisters approach.

The reduced-form estimation of the association between teenage fertility and the grade failure is equal to 55.7 (0.028 SE). The age-for-grade measure is estimated to be 0.315 (0.009 se). Finally, the dropout risk coefficient is equal to 0.423 (0.201 SE). All coefficients are statistically significant at the one-percent level.

5.2 Schooling Attainment

I present the effects of reporting a pregnancy before the age of 18 on the school attainment using a static instrumental variables identification strategy in Table 3 and the sibling differences approach in Table 4.[36]

Teenage pregnancy decreases the number of years of completed schooling by 1.8 years (0.922 SE, p<0.05) in the instrumental-variable specification. Using the sibling differences, the coefficient is smaller and amounts to 0.78 years (0.194 SE, p<0.01). The differences in size could be potentially explained by the sample as the non pregnant sister is approximately a year less educated that non pregnant teens in the full sample.

Using the instrumental variable strategy, I find that teens who report a pregnancy are not more likely to sit for the matriculation exam but are not reporting that they continue to post-secondary school education, as the coefficients are 0.040 (0.289 SE) and -0.259 (0.199 SE), respectively. None of these coefficients is statistically significant. In contrast, in the sibling difference strategy, teenage mothers are not less likely to sit for the matriculation exam but do not report that they continue to post secondary school education, as the coefficients are 0.211 (0.042 SE, p<0.01) and -0.100 (0.024 SE, p<0.05), respectively.

The bottom panel of Table 3 presents the reduced form estimation coefficients. The estimates are equal to 0.057(0.030 SE, p<0.01), 0.001 (0.001 SE), and -0.008 (0.001 SE) for the years of completed years of education, taking the matric exam, and the post secondary education variables respectively. However, only the first coefficient is statistically significant.

[36]Table A5 in the appendix contains the OLS results for the educational attainment variables. Teenage pregnancy decreases the number of years of completed schooling by 1.05, likelihood of sitting for the matriculation exam decreases the number by 0.233 percentage points, and post-secondary education decreases it by 4.3 points.

Across the paper, I have presented the estimates of women who report a pregnancy before the age of 18. However, the CAPS data set allows me to identify which pregnancy resulted in a birth. Hence, in Appendix Table A11, I present the estimation of the effects of teenage birth on the schooling attainment of the sampled women. The results are consistent with those found in Table 3.

6 Labor Force Participation

I present the effects of teenage pregnancy on the labor force participation of the full and the sister samples. The coefficients presented in Table 5 represent coefficients that describe employment at ages 19, 20, 21 and 22. The top panel includes the coefficients estimated using the instrumental variable identification strategy, and Panel B the reduced form estimation coefficients. The bottom panel the sibling differences estimation results.[37]

First, the instrumental variable results, presented in panel A, amount to: 0.192 (0.260 SE), 0.338 (0.294 SE), 0.229 (0.264 SE) and -0.114 (0.266 SE) at the ages of 19, 20, 21 and 22. Nevertheless, none of the instrumental variable coefficients are statistically significant.[38] The reduced-form coefficients (the relationship between teenage fertility and the outcomes of interest) are close to zero, have wide standard errors, and are not statistically significant.

In the following panel, the sibling differences estimation coefficients follow a similar pattern. The likelihood of being employed at the ages of 19, 20, 21, and 22 is, on average, 0.034 (0.037 SE), 0.006 (0.042 SE), and 0.010 (0.045 SE) and 0.008 (0.046 SE) additional percentage points, respectively. None of the coefficients is statistically significant.

7 Mitigating factors

As previously discussed, the consequences of teenage pregnancy in South Africa might be different from other higher-income countries. The main reasons for the differential include

[37] Also, in the appendix Table A13, I examine a different in the sample for the year 2006. The information available in wave 4 allows me to study the number of hours women devotes to working and studying, if she attends school, and her willingness of to accept positions as domestic workers and security guards.

[38] In Table A5 in the appendix, I present the coefficients for the OLS estimation. The estimates are positive: 0.115, 0.059, 0.048, and 0.078 for the ages of 19, 20, 21, and 22, respectively. Three coefficients are statistically significant, but the one for the age of 21 is not.

higher youth unemployment rates, the existence of strong familial networks, and high levels of grade repetition. The uncertainty created by the lack of prospects in the labor market may distort the relative human capital costs of teenage motherhood. Similarly, the strength familial networks can lead to changes in the cost of child care, while the increased grade repetition may lead to a reduction in the stigma of teenage pregnancy in schools. The first item affects every sampled woman in my sample equally, but the CAPS data set contains demographic information that allows me to consider the last two possibilities. I test for the possibility of a change in the individual opportunity cost in the the static model.

I select the collapsed panel model for the mitigation analysis because the information is collapsed at the woman level and it is mostly common to siblings. Econometrically, in order to undertake the analysis I interact the indicator for teenage pregnancy with a indicator variable for either the having the grand mother living at home or attending a school with above average repetition rates I take a two-step control function estimation approach in which the endogeneity is corrected by modelling the endogeneity in the error term (Wooldridge, 2015). I correct the variance-covariance matrix using bootstrapping techniques because the second-stage regression includes the first-stage generated regressors.

7.1 Familial Networks

I begin the analysis by testing whether having a grandmother alive during the sampled women's teen years with the pregnancy variable had an impact on the schooling of the teenager. The results for educational attainment are presented in Table 6. The effect of teenage pregnancy on the total completed years of education is estimated to be 2.138 fewer years (0.692 se, $p<0.01$). The coefficient for the grandmother is 0.285 (0.157 se, $p<0.1$). The interaction between the two variables renders a positive coefficient of 0.521 (0.316 se, $p<0.1$).The findings is suggestive of the fact that having additional child care might mitigate the negative effects of teenage pregnancy.

The estimated effect of teenage pregnancy on the likelihood of sitting for the matriculation exam, presented in column (1), is similar in size to the coefficients presented in Table 3. The coefficients for teenage pregnancy and living grandmother are -0.059 (0.211 se) and 0.124 (0.057 se), respectively. However, only the second is statistically significant (at the five-percent level). The interaction coefficient, although positive and equal to 0.017 (0.073 se), is not statistically significant. The coefficients for teenage pregnancy, living grandmother, and the interaction are -0.165 (0.142 se), 0.023 (0.028 se) and -0.015 (0.054 se), respectively.[39]

7.2 Schools with High Levels of Grade Repetition

To examine whether there is any reduction in the stigma in an environment where grade repetition is high, I take advantage of the school data available in wave 1 of the CAPS data set for all of the young adults, men and women.[40] For each reported institution, I estimate the percentage of young adults who reported failing a grade in 2002. I then create an indicator variable that takes the value of one if the sampled woman attended a school with high grade-repetition rate and zero otherwise. The coefficients for these regressions are presented in Table 6.[41]

In column (3), I present the impact of teenage pregnancy on the total completed years of education, including the interaction between teenage pregnancy and attendance at a school with a high rate of grade repetition. The coefficient for teenage pregnancy is 1.889 fewer years of completed education (0.095 se, $p<0.05$). Attending a school with a high level of grade repetition leads to an increase of 0.143 (0.104 se) years of education. However, the interaction coefficient is positive 0.410 (0.225 se, $p<0.1$). Again, the result provides support to the idea that reducing the stigma mitigates the negative effects of teenage pregnancy.

The coefficients for the regression for when the left-hand side variable is an indicator variable for whether the individual sat for the matriculation exam are 0.123 (0.333 se) for teenage pregnancy and 0.110 (0.040 se, $p<0.1$) for attendance at a school with high grade repetition. The interaction estimated coefficient is -0.006 (0.052). As seen in column (6), when the outcome variable is post secondary education, the coefficients for teenage pregnancy, living grandmother, and the interaction are all negative: 0.255 (0.216 se), 0.004 (0.029 se), and 0.016 (0.029 se), respectively. Neither coefficient is statistically significant.

[39] The results for the presence of the grandfather can be found in found in Appendix Table A14. As shown, there are no statistically significant effects when the father is alive.

[40] They reported the last institution they have attended.

[41] The correlation between attending schools with high grade repetition and teenage pregnancy in the sample is -0.0415.

8 Conclusion

Evidence of the causal effects of teen pregnancy is scarce in low-income countries, and particularly in sub-Saharan Africa. The paper provides empirical evidence by addressing whether early pregnancy causally affects the education and labor outcomes of young women in Cape Town, South Africa. Estimating the economic consequences of teen pregnancy is difficult because it requires disentangling the causal effects of the confounding selection into early motherhood. Using a rich data set of South Africa, I explore the issue using two identification strategies: instrumenting teenage pregnancy using teenage fertility and a sibling differences.

My findings suggest that teen pregnancy decreases the pace at which girls progress in their schooling. In particular, I find that teenage mothers are more likely to fail a grade, lag behind in their education, and drop out of school. In the static model, my findings indicate that women who report a pregnancy are less educated and less likely to continue to post secondary education. The results are mostly consistent between identification strategies. Taken together, the results contribute to existing evidence that delayed pregnancy leads to beneficial outcomes for teenage girls in the short run.

Nevertheless, I find two factors that attenuate the negative effects of teenage pregnancy. First, the presence of the teen's mother during the teenage years attenuates the negative effects of early pregnancy on schooling. Second, attending a school with above-average failure rates also mitigates the lag in education effect. Estimating these specific attenuating characteristics is an important step towards designing policies can be successfully assist women continue their education.

In South Africa, where a third of the women report a birth by the age of 19, the evidence presented in this paper suggests that the timing of the first birth is important for the educational outcomes of young women. This paper speaks to the debate on reproductive health policies in low-income countries (Herrera et al., 2019). My findings imply that postponing the age at first birth can have important consequences for women's educational attainment and may delay entrance into the labor force or improve overall well-being (Ardington et al. (2014) and Urdinola and Ospino, 2015). The mitigating factors found in this paper would also inform these policies. Additionally, policies that aim at assisting teenage mothers should focus on the factors that can alleviate the duties related to childcare and the stigma.[42]

[42] In future research, I will study the overall fertility decisions of South African women. A secondary result, presented in Table A15 in the appendix, suggests that women who report a teenage pregnancy are not more

References

Adair, L. S. (2001). Size at birth predicts age at menarche. *Pediatrics*, *107*(4), E59.
Adebara, I. & Munir'deen, A. I. (2012). Relationship of biosocial factors and hemoglobin genotype with age of menarche among school girls in ilorin, nigeria. *Nepal Journal of Obstetrics and Gynaecology*, *7*(1), 19–24.
Akresh, R., Bhalotra, S., Leone, M., & Osili, U. O. (2012). War and stature: Growing up during the nigerian civil war. *American Economic Review*, *102*(3), 273–77.
Alderman, H., Hoddinott, J., & Kinsey, B. (2006). Long term consequences of early childhood malnutrition. *Oxford economic papers*, *58*(3), 450–474.
Almanza, C. H. & Sahn, D. E. (2018). Early childbearing, school attainment, and cognitive skills: Evidence from madagascar. *Demography*, *55*(2), 643–668.
Almond, D. (2006). Is the 1918 influenza pandemic over? long-term effects of in utero influenza exposure in the post-1940 us population. *Journal of political Economy*, *114*(4), 672–712.
Almond, D., Chay, K. Y., & Lee, D. S. (2005). The costs of low birth weight. *The Quarterly Journal of Economics*, *120*(3), 1031–1083.
Almond, D. & Mazumder, B. (2011). Health capital and the prenatal environment: The effect of ramadan observance during pregnancy. *American Economic Journal: Applied Economics*, *3*(4), 56–85.
Anderson, K. G., Case, A., & Lam, D. (2001). Causes and consequences of schooling outcomes in south africa: Evidence from survey data. *Social dynamics*, *27*(1), 37–59.
Arceo-Gómez, E. O. & Campos-Vázquez, R. M. (2014). Teenage pregnancy in mexico: Evolution and consequences. *Latin american journal of economics*, *51*(1), 109–146.
Ardington, C., Menendez, A., & Mutevedzi, T. (2014). Early childbearing, human capital attainment, and mortality risk: Evidence from a longitudinal demographic surveillance area in rural kwazulu-natal, south africa. *Economic development and cultural change*, *63*(2), 281–317.
Asadullah, M. N. & Wahhaj, Z. (2016). Early marriage, social networks and the transmission of norms. *Economica*.
Ashcraft, A., Fernández-Val, I., & Lang, K. (2013). The consequences of teenage childbearing: Consistent estimates when abortion makes miscarriage non-random. *The Economic Journal*, *123*(571), 875–905.
Ashcraft, A. & Lang, K. (2006). *The consequences of teenage childbearing.* National Bureau of Economic Research.
Azevedo, J. P., Lopez-Calva, L. F., & Perova, E. (2012). *Is the baby to blame? an inquiry into the consequences of early childbearing.* The World Bank.

likely to have more children and are more likely to report a smaller subsequent fertility. These results hint at a degree of substitution between teenage fertility and subsequent fertility that should be explored in more detail

Banerjee, A., Galiani, S., Levinsohn, J., McLaren, Z., & Woolard, I. (2008). Why has unemployment risen in the new south africa? 1. *Economics of Transition*, *16*(4), 715–740.

Berthelon, M. & Kruger, D. I. (2017). Does adolescent motherhood affect education and labor market outcomes of mothers? a study on young adult women in chile during 1990–2013. *International journal of public health*, *62*(2), 293–303.

Bitler, M. & Zavodny, M. (2001). The effect of abortion restrictions on the timing of abortions. *Journal of health economics*, *20*(6), 1011–1032.

Branson, N. & Byker, T. (2018). Causes and consequences of teen childbearing: Evidence from a reproductive health intervention in south africa. *Journal of health economics*, *57*, 221–235.

Chari, A., Heath, R., Maertens, A., & Fatima, F. (2017). The causal effect of maternal age at marriage on child wellbeing: Evidence from india. *Journal of Development Economics*, *127*, 42–55.

Cooper, D., Morroni, C., Orner, P., Moodley, J., Harries, J., Cullingworth, L., & Hoffman, M. (2004). Ten years of democracy in south africa: Documenting transformation in reproductive health policy and status. *Reproductive Health Matters*, *12*(24), 70–85.

Dahiya, M. & Rathi, V. K. (2010). Relationship between age at menarche and early-life nutritional status in india. *British Journal of Sports Medicine*, *44*(Suppl 1), i43–i43.

Department of Health/South Africa and Macro International. (2007). South africa demographic and health survey 1998. https://dhsprogram.com/pubs/pdf/FR206/FR206.pdf.

Diaz, C. J. & Fiel, J. E. (2016). The effect (s) of teen pregnancy: Reconciling theory, methods, and findings. *Demography*, *53*(1), 85–116.

Duflo, E. (2003). Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in south africa. *The World Bank Economic Review*, *17*(1), 1–25.

Duncan, G. J., Lee, K. T., Rosales-Rueda, M., & Kalil, A. (2018). Maternal age and child development. *Demography*, *55*(6), 2229–2255.

Faisal-Cury, A., Tabb, K. M., Niciunovas, G., Cunningham, C., Menezes, P. R., & Huang, H. (2017). Lower education among low-income brazilian adolescent females is associated with planned pregnancies. *International journal of women's health*, *9*, 43.

Field, E. & Ambrus, A. (2008). Early marriage, age of menarche, and female schooling attainment in bangladesh. *Journal of political Economy*, *116*(5), 881–930.

Fletcher, J. M. & Wolfe, B. L. (2009). Education and labor market consequences of teenage childbearing evidence using the timing of pregnancy outcomes and community fixed effects. *Journal of Human Resources*, *44*(2), 303–325.

Fogel, R. W. (1993). New sources and new techniques for the study of secular trends in nutritional status, health, mortality, and the process of aging. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *26*(1), 5–43.

Geronimus, A. T. & Korenman, S. (1993). The socioeconomic costs of teenage childbearing: Evidence and interpretation. *Demography*, *30*(2), 281–290.

Glynn, J. R., Sunny, B. S., DeStavola, B., Dube, A., Chihana, M., Price, A. J., & Crampin, A. C. (2018). Early school failure predicts teenage pregnancy and marriage: A large population-based cohort study in northern malawi. *PloS one*, *13*(5), e0196041.

Heiland, F., Korenman, S., & Smith, R. A. (2019). Estimating the educational consequences of teenage childbearing: Identification, heterogeneous effects and the value of biological relationship information. *Economics & Human Biology*, *33*, 15–28.
Herrera, C., Sahn, D. E., & Villa, K. M. (2019). Teen fertility and female employment outcomes: Evidence from madagascar. *Journal of African Economies.*
Hicks, J. & Hicks, D. L. (2015). Lucky late bloomers? the consequences of early marriage for women in western kenya. *The Consequences of Early Marriage for Women in Western Kenya (June 23, 2015).*
Hillard, P. J. A. (2013). *Practical pediatric and adolescent gynecology.* John Wiley & Sons.
Hotz, V. J., McElroy, S. W., & Sanders, S. G. (2005). Teenage childbearing and its life cycle consequences exploiting a natural experiment. *Journal of Human Resources*, *40*(3), 683–715.
Jahanfar, S., Lye, M.-S., & Krishnarajah, I. S. (2013). Genetic and environmental effects on age at menarche, and its relationship with reproductive health in twins. *Indian journal of human genetics*, *19*(2), 245.
Karapanou, O. & Papadimitriou, A. (2010). Determinants of menarche. *Reproductive Biology and Endocrinology*, *8*(1), 115.
Khanna, M. (2019). The precocious period: The impact of early menarche on schooling in india. *Available at SSRN 3419041.*
Kim, Y. & Je, Y. (2019). Early menarche and risk of metabolic syndrome: A systematic review and meta-analysis. *Journal of Women's Health*, *28*(1), 77–86.
Klepinger, D. H., Lundberg, S., Plotnick, R., et al. (1997). *How does adolescent fertility affect the human capital and wages of young women?* Institute for Research on Poverty, University of Wisconsin.
Lam, D., Ardington, C., & Leibbrandt, M. (2011). Schooling as a lottery: Racial differences in school advancement in urban south africa. *Journal of development economics*, *95*(2), 121–136.
Lam, D., Marteleto, L. J., & Ranchhod, V. (2013). The influence of older classmates on adolescent sexual behavior in cape town, south africa. *Studies in family planning*, *44*(2), 147–167.
Lam, D., Seekings, J., & Sparks, M. (2006). The cape area panel study (caps): Overview and technical documentation for waves 1–2–3. *Cape Town, South Africa: CAPS, University of Cape Town.*
Lawn, R. B., Lawlor, D. A., & Fraser, A. (2017). Associations between maternal prepregnancy body mass index and gestational weight gain and daughter's age at menarche: The avon longitudinal study of parents and children. *American journal of epidemiology*, *187*(4), 677–686.
Lee, D. (2010). The early socioeconomic effects of teenage childbearing: A propensity score matching approach. *Demographic Research*, *23*, 697–736.
Levine, D. I. & Painter, G. (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*(4), 884–900.

Madhavan, S. & Thomas, K. J. (2005). Childbearing and schooling: New evidence from south africa. *Comparative Education Review*, *49*(4), 452–467.
Magruder, J. R. (2010). Intergenerational networks, unemployment, and persistent inequality in south africa. *American Economic Journal: Applied Economics*, *2*(1), 62–85.
Mahmud, S., Shah, N. M., & Becker, S. (2012). Measurement of women's empowerment in rural bangladesh. *World development*, *40*(3), 610–619.
Marteleto, L., Lam, D., & Ranchhod, V. (2008). Sexual behavior, pregnancy, and schooling among young people in urban south africa. *Studies in family planning*, *39*(4), 351–368.
Martorell, R. & Habicht, J.-P. (1986). Growth in early childhood in developing countries.
Monteiro, D. L. M., Martins, J. A. F. d. S., Rodrigues, N. C. P., Miranda, F. R. D. d., Lacerda, I. M. S., Souza, F. M. d., … Trajano, A. J. B. (2019). Adolescent pregnancy trends in the last decade. *Revista da Associação Médica Brasileira*, *65*(9), 1209–1215.
Moultrie, T. A. & Dorrington, R. (2004). *Estimation of fertility from the 2001 south africa census data*. Centre for Actuarial Research, University of Cape Town Cape Town.
Mpora, B. O., Piloya, T., Awor, S., Ngwiri, T., Laigong, P., Mworozi, E. A., & Hochberg, Z. (2014). Age at menarche in relation to nutritional status and critical life events among rural and urban secondary school girls in post-conflict northern uganda. *BMC women's health*, *14*(1), 66.
Narita, R. & Diaz, M. D. M. (2016). Teenage motherhood, education, and labor market outcomes of the mother: Evidence from brazilian data. *Economia*, *17*(2), 238–252.
Nattrass, N. & Walker, R. (2005). Unemployment and reservation wages in working-class cape town. *South African Journal of Economics*, *73*(3), 498–509.
Padez, C. (2003). Age at menarche of schoolgirls in maputo, mozambique. *Annals of human biology*, *30*(4), 487–495.
Panday, S., Makiwane, M., Ranchod, C., & Letsoala, T. (2009). Teenage pregnancy in south africa: With a specific focus on school-going learners.
Prentice, S., Fulford, A. J., Jarjou, L. M. A., Goldberg, G. R., & Prentice, A. (2010). Evidence for a downward secular trend in age of menarche in a rural gambian population. *Annals of Human Biology*, *37*(5), 717–721. PMID: 20465526. doi:10.3109/03014461003727606. eprint: https://doi.org/10.3109/03014461003727606
Rah, J. H., Shamim, A. A., Arju, U. T., Labrique, A. B., Rashid, M., & Christian, P. (2009). Age of onset, nutritional determinants, and seasonal variations in menarche in rural bangladesh. *Journal of health, population, and nutrition*, *27*(6), 802.
Ranchhod, V., Lam, D., Leibbrandt, M., & Marteleto, L. (2011). Estimating the effect of adolescent fertility on educational attainment in cape town using a propensity score weighted regression.
Ribar, D. (1994). Teenage fertility and high school completion. *The Review of Economics and Statistics*, 76(3):413–424.
Ribar, D. C. (1999). The socioeconomic consequences of young women's childbearing: Reconciling disparate evidence. *Journal of Population Economics*, *12*(4), 547–565.

Sedgh, G., Finer, L. B., Bankole, A., Eilers, M. A., & Singh, S. (2015). Adolescent pregnancy, birth, and abortion rates across countries: Levels and recent trends. *Journal of Adolescent Health*, *56*(2), 223–230.

Sekhri, S. & Debnath, S. (2014). Intergenerational consequences of early age marriages of girls: Effect on children's human capital. *The Journal of Development Studies*, *50*(12), 1670–1686.

Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of biosocial science*, *35*(2), 263–285.

Sørensen, K., Juul, A., Christensen, K., Skytthe, A., Scheike, T., & Kold Jensen, T. (2013). Birth size and age at menarche: A twin perspective. *Human Reproduction*, *28*(10), 2865–2871.

Sunder, N. (2018). Marriage age, social status and intergenerational effects in uganda. *Social Status and Intergenerational Effects in Uganda (March 1, 2018)*.

Unicef. (2014). *Ending child marriage: Progress and prospects*. UNICEF.

Urdinola, B. P. & Ospino, C. (2015). Long-term consequences of adolescent fertility: The colombian case. *Demographic Research*, *32*, 1487–1518.

Van der Berg, S., Louw, M. et al. (2007). Lessons learnt from sacmeqii: South african student performance in regional context. *University of Stellenbosch, Department of Economics and Bureau for Economic Research Working Paper*, *16*(07).

WHO. (2018). World health organization. https://www.who.int/news-room/fact-sheets/detail/adolescent-pregnancy.

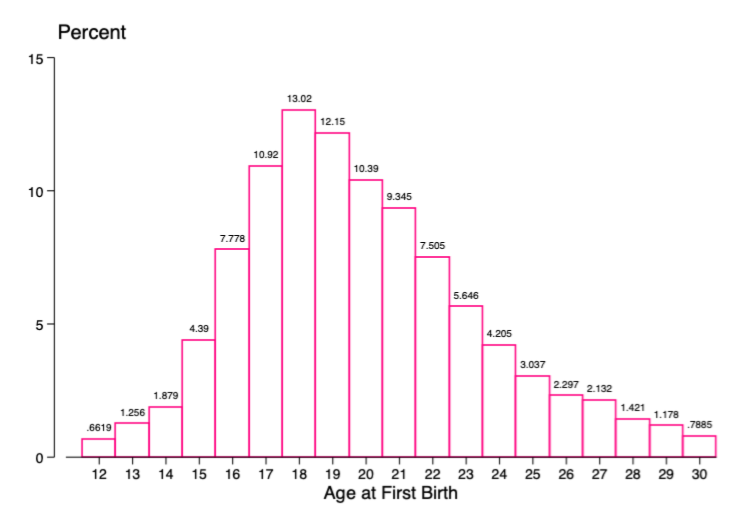
Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, *50*(2), 420–445.

World Health Organization. (2014). Trends in maternal mortality: 1990 to 2013: Estimates by who, unicef, unfpa, the world bank and the united nations population division.

Zito, R. C. (2018). Children as saviors? a propensity score analysis of the impact of teenage motherhood on personal transformation. *Youth & Society*, *50*(8), 1100–1122.

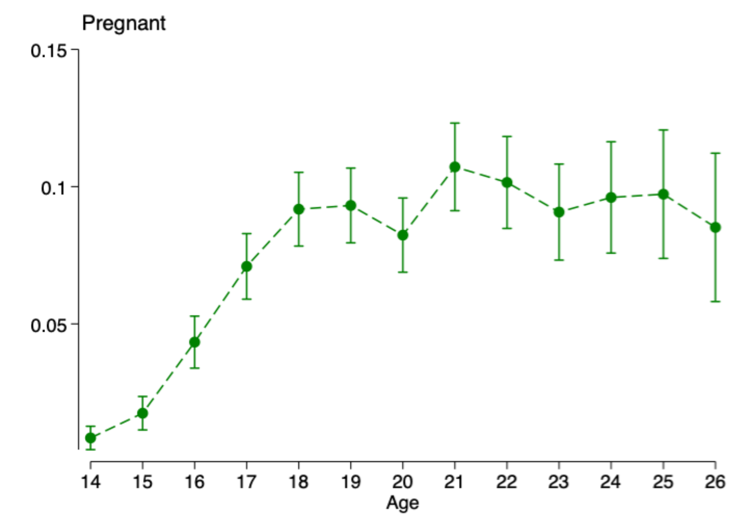
A Graphs and Tables

Figure 1: Age at First Birth - Adult South African Women



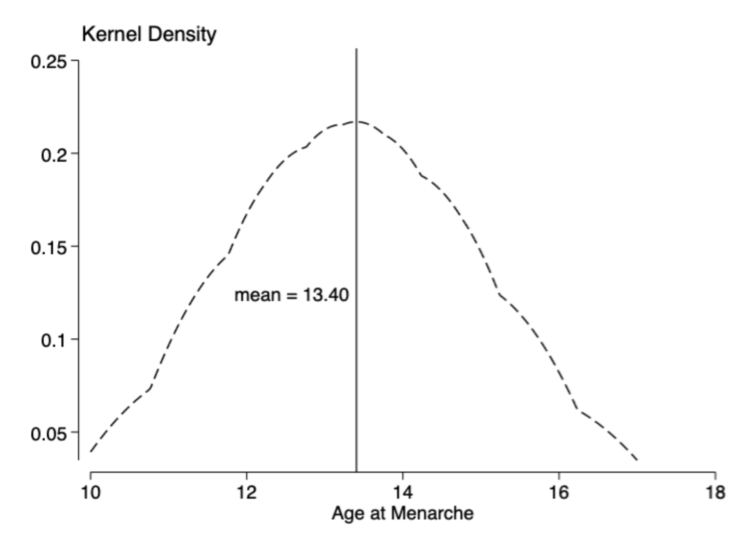
Source: Demographic Health Survey, 1998. Observations=102,073.

Figure 2: Percentage of Reported Pregnancies by Age



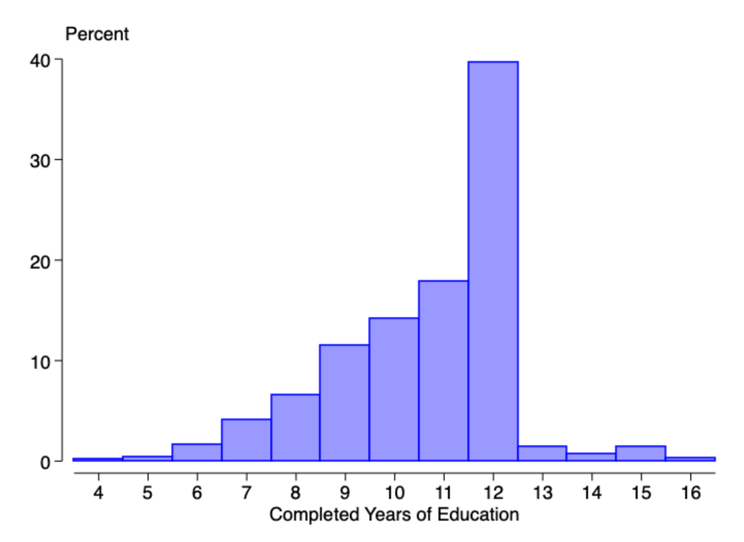
Notes: The sample includes women who reached menarche between the ages of 10 and 17. Source: waves 1-5 of the CAPS data set. Observations=1,741.

Figure 3: Age at Menarche.



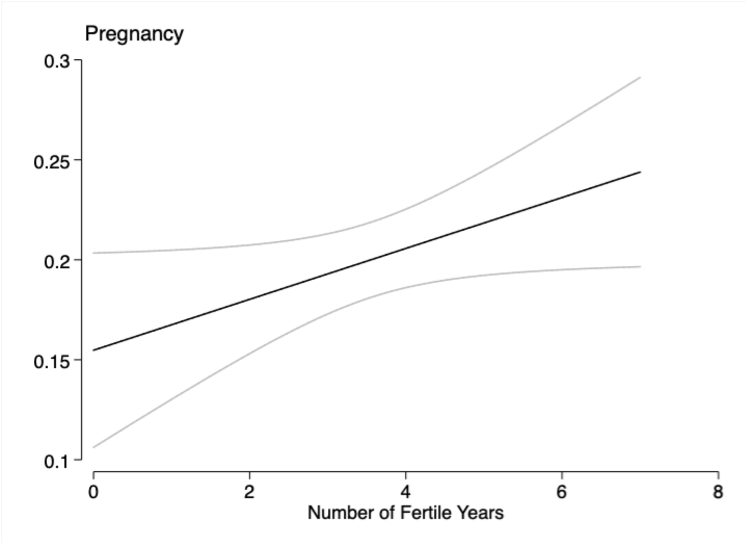
Notes: The sample includes women who reached menarche between the ages of 10 and 17. Source: waves 1 and 3, CAPS. Observations: 1,741.

Figure 4: Completed Years of Education



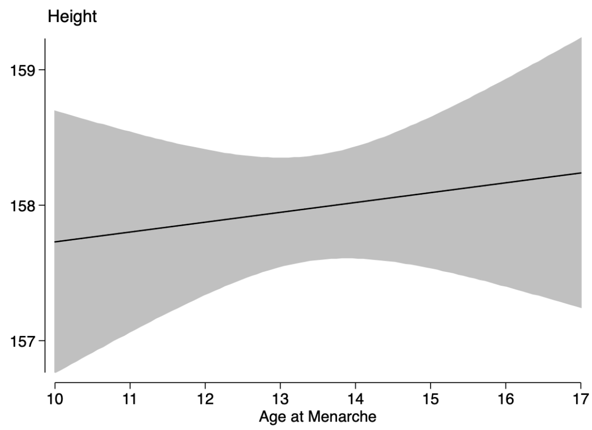
Notes: The sample includes women who reached menarche between the ages of 10 and 17 with complete information of all the outcomes and the control variables. Source: waves 1-5 of the CAPS data set. Observations: 2,612

Figure 5: First Stage: Teenage Pregnancy and Fertility



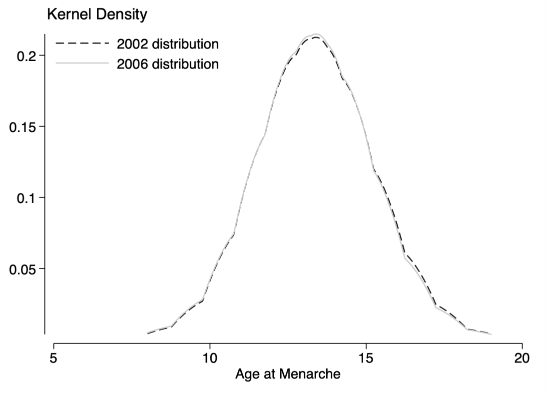
Notes: Relationship between the number of Fertile years and Teenage Pregnancy. The sample includes women who reached menarche between the ages of 10 and 17. Source: CAPS waves 1-5. Observations: 1,741

Figure 6: Adult Height and Age at Menarche



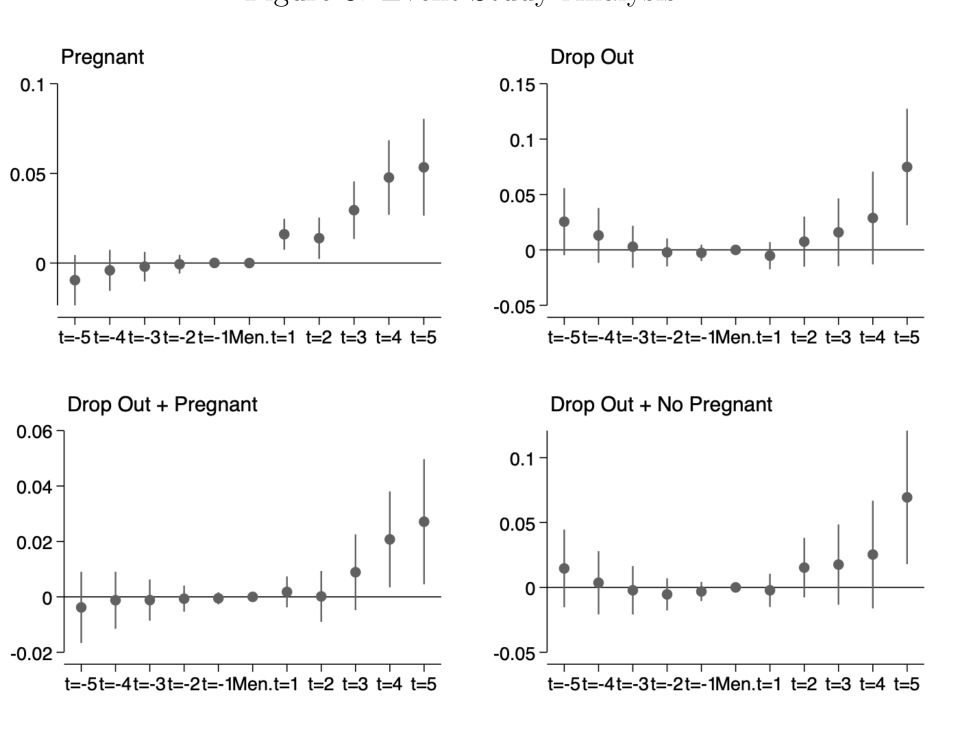
Notes: The sample includes women who reached menarche between the ages of 10 and 17 with complete outcomes and health variables. Source: CAPS wave 4. Observations: 1,741

Figure 7: Age at menarche questions - CAPS Dataset



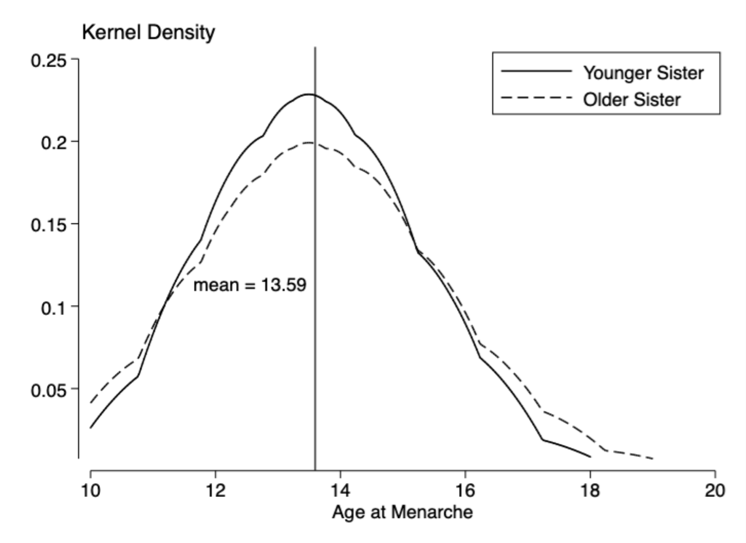
Notes: The sample includes women who reached menarche between the ages of 10 and 17. Epanechnikov, bandwidth = 1. Source: waves 1 and 3 of the CAPS data set. Observations: 2455.

Figure 8: Event Study Analysis



The event is the age at menarche, which is denoted by "Men." in the graph. Source: waves 1-5 of the CAPS data set. Notes: Men Observations=15,170

Figure 9: Age at menarche between sisters



Notes: The sample includes sisters. Epanechnikov, bandwidth = 1 Source: waves 1 and 3, CAPS. Observations: 418

Table 1: Summary Statistics - by Sample

	Full Sample (1)	Sibling Sample (2)	Difference (3)
Demographic Characteristics			
Age in 2002	17.723	17.902	0.179
	(2.445)	(2.556)	
% Coloured	0.461	0.409	-0.052*
	(0.499)	(0.492)	
% Black	0.486	0.565	0.070***
	(0.500)	(0.496)	
Asset index	-0.145	-0.560	-0.415***
	(1.981)	(2.012)	
Adult Height - cm	157.979	158.355	0.376
	(8.131)	(8.014)	
Mother's Education	8.276	7.973	-0.303*
	(3.136)	(2.906)	
Not Born in WC	0.283	0.378	0.095***
	(0.451)	(0.485)	
# Full Siblings	2.317	2.557	0.240**
	(1.756)	(1.829)	
Sexual Activity			
Age 1st Partner	19.779	19.656	-0.123
	(3.209)	(3.347)	
Age 1st Active	17.392	17.016	-0.376***
	(2.289)	(2.036)	
Observations	1,741	418	

Notes: "Adult Height"is measured in 2006. "% Coloured" and "% Black" report the percentage of the population who identify with that race (the third category is white). "Asset index" is pca index of the assets of women's household. " Adult Height - cm" is the women's height measured in centimeters. "Mother's Education" describe the total number of years completed by mother and "Not Born in WC" is an indicator variable for individuals born outside the Western Cape. "# Full Siblings" refers to the number of siblings. Finally, "Age of 1st. Partner" and "Age 1st Active" describe the ages of the first sexual partner and the age in which women in the sample became active respectively. The Diff column is the difference in means of Columns (1) and (2), where a T-test where the hypothesis is that the coefficient is equal to zero. *** p≤0.01, ** p≤0.05, * p≤0.1

Table 2: First Stage Estimation

	Panel Analysis		Collapsed Panel			
	Pregnant$_{it}$		Pregnancy $\leq 18_i$			
	(1)	(2)	(3)	(4)	(5)	(6)
Fertile$_t$	0.023***	0.024***				
	(0.002)	(0.002)				
Fertile years$_i$			0.016***	0.032***		
			(0.006)	(0.008)		
Menarche $\leq 14_i$					0.054***	0.090***
					(0.018)	(0.022)
Observations	15,170	15,170	1,741	1,741	1,741	1,741
R-squared	0.009	0.010	0.004	0.011	0.005	0.011
First-stage F	202	171.2	6.914	17.61	8.935	17.09
Sampling Location FE	No	Yes	No	Yes	No	Yes
Controls	No	Yes	No	Yes	No	Yes
Time FE	No	Yes				

Notes: "Pregnant$_{icst}$" is defined 1 for individual i of cohort c, who lives in sampling cluster s and observed in year t if she got pregnant and zero; if she did not get pregnant during that year. In the same Data set, "Fertile$_t$" is a dichotomous variable equal to zero if the sample individual had not reached menarche in year t and equal to one if she had in that same year. "Pregnancy $\leq$ 18" in the collapsed panel Data-set is defined one for those who got pregnant before 18 and zero for those who did not. Also, the "Number of Fertile Years" indicates the number of years passed since menarche until 17. Controls include the race, an asset index, height, number of siblings, the log of the household level and indicator variables for religion, language spoken by her family, place of birth and residence. Cohort and sampling-location fixed effects are included. In the panel data set, Pregnant in year t is an indicator variable for whether the individual got pregnant in year t and Fertile is an indicator variable of whether a woman was fertile in a specific year. In both data sets, standard errors in parentheses, clustered at the sampling location level.*** p $\leq$ 0.01, ** p $\leq$ 0.05, * p$\leq$0.1

Table 3: Estimation of the Effects of Teenage Pregnancy - Schooling Progression

	Panel			Collapsed Panel		
	Failed grade	Age for Grade	Hazard Drop Out	Years of Education	Took Matric Exam	Post Secondary Schooling
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: IV Estimation						
Pregnant_{icst}	0.557***	0.284***	0.097***			
	(0.028)	(0.010)	(0.009)			
$\text{Pregnancy} \leq 18_i$				-1.820**	0.044	-0.259
				(0.922)	(0.289)	(0.199)
Observations	15,170	15,170	14,354	1,741	1,741	1,741
R-squared	0.010	0.023		0.172	0.208	0.019
First stage-F-stat	171.2	171.2	171.2	17.61	17.61	17.61
Panel B: Reduced Form Estimation						
$\text{Post Menarche}_{itcs}$	0.617***	0.315***	0.423*			
	(0.033)	(0.009)	(0.201)			
Fertile Years_i				-0.057*	0.001	-0.008
				(0.030)	(0.009)	(0.006)
Observations	15,170	15,170	14,354	1,741	1,741	1,741
R-squared	0.010	0.023		0.462	0.457	0.405
Mean dependent var.	0.113	1.091	0.359	11.05	0.436	0.146

Notes: "Pregnant_{icst}" is equal to 1 for individual i of cohort c, who lives in sampling cluster s and reported a pregnancy in year t and 0 otherwise. Similarly, "$\text{Pregnant}\leq 18_i$" takes the value of 1 if women i reported a pregnancy before the age of 18 and 0 otherwise. Columns (1) - (6) contain the IV and reduced estimates of the effects on the likelihood of failing a grade in year t and the degree to which a sampled women lagged behind her secondary education in that same year. In panel A, endogeneity is accounted for using a linear first stage of the effects of teenage fertility on pregnancy. The First Stage statistic comes from the a first stage regression where pregnancy is instrumented using either a post menarche indicator or the number of fertile years. In the the static analysis, controls include the race, an asset index, height, number of siblings, and the indicator variables for religion, language spoken by her family. Standard errors in parentheses, clustered at the smallest sampling location level. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$

Table 4: Sibling Differences Estimation of the Effects of Teenage Pregnancy

	Panel			Collapsed Panel		
	Failed grade (1)	Age for Grade (2)	Hazard Drop Out (3)	Years of Education (4)	Took Matric Exam (5)	Post Secondary Schooling (6)
Panel A: OLS Estimation						
Pregnant_{icst}	0.128***	0.501***	0.583***			
	(0.043)	(0.122)	(0.091)			
$\text{Pregnancy} \leq 18_i$				-0.707***	-0.155***	-0.057**
				(0.194)	(0.042)	(0.024)
Observations	3,521	3,521	4,535	418	418	418
R-squared	0.123	0.617		0.172	0.208	0.019
Panel B: Sibling Differences						
Pregnant_{icst}	0.132***	0.521***	0.494**			
	(0.037)	(0.080)	(0.103)			
$\text{Pregnancy} \leq 18_i$				-0.774***	-0.159***	-0.051**
				(0.170)	(0.040)	(0.024)
Observations	3,521	3,521	4,535	418	418	418
R-squared	0.010	0.023		0.075	0.075	0.062
Mean dependent var.	0.101	1.002	0.143	10.52	0.360	0.102

Notes: Notes: "Pregnant_{icst}" is defined as 1 for individual i of cohort c, who lives in sampling cluster s and observed in year t if she got pregnant and zero; if she did not get pregnant during that year. Similarly, "$\text{Pregnant}\leq 18_i$" takes the value of 1 if woman i reported a pregnancy before the age of 18 and 0 otherwise. Columns (1) and (2) contain the OLS and Sibling Differences estimates of the effects on the likelihood of failing a grade in year t and the degree to which a sampled women lagged behind her secondary education in that same year. Column (3) contains the hazard rated for dropping out with and without the housed fixed effects. "Mean dependent var." indicates the average of the outcome if Pregnancy in year t is equal to zero. Controls includes height, Standard errors in parentheses, clustered at the smallest sampling location level. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$

Table 5: Estimation of the Effects of Teenage Pregnancy - Labor Force Participation

	At age 19 (1)	At age 20 (2)	At age 21 (3)	At age 22 (4)
Panel A: IV Estimation				
Pregnancy$\leq 18_i$	0.192	0.338	0.229	-0.114
	(0.260)	(0.294)	(0.264)	(0.266)
Observations	1,741	1,741	1,741	1,741
R-squared	0.215	-0.083	0.003	0.074
First Stage F-stat	17.61	17.61	17.61	17.61
Panel B: Reduced Form Estimation				
Fertile Years$_i$	0.006	0.011	0.007	-0.004
	(0.008)	(0.009)	(0.008)	(0.009)
Observations	1,741	1,741	1,741	1,741
R-squared	0.235	0.036	0.068	0.077
Mean Dependent Var	0.458	0.704	0.764	0.686
Panel C: Sibling Differences				
Pregnancy$\leq 18_i$	0.034	0.006	0.010	0.008
	(0.037)	(0.042)	(0.045)	(0.046)
Observations	418	418	418	418
R-squared	0.029	0.031	0.070	0.086
Mean Dependent Var	0.197	0.282	0.328	0.363

Notes: "Pregnancy $\leq$ 18" is defined one for those who got pregnant before 18 and zero for those who did not. Controls include the race, an asset index, height, number of siblings, the log of the household level and the indicator variables for religion, language spoken by her family, place of birth and residence. In each column, the outcome variable is an indicator variable of whether the sampled women reported being active at the age on the top; and zero otherwise. The "First Stage F-stat" comes from the a first stage regression where "Pregnancy $\leq$ 18" is instrumented with the number of teenage fertile years. The Mean Dependent Var indicates the average of the outcome if "Pregnancy $\leq$ 18" is equal to zero. Cohort and sampling-location fixed effects are included. Standard errors in parentheses, clustered at the sampling location level. *** p $\leq$ 0.01, ** p $\leq$ 0.05, * p$\leq$0.1

Table 6: Sibling Differences Estimation of the Effects of Teenage Pregnancy

	Years of Education (1)	Took Matric Exam (2)	Post Secondary Schooling (3)	Years of Education (4)	Took Matric Exam (5)	Post Secondary Schooling (6)
Pregnancy $\leq$ 18	-2.138***	-0.059	-0.165	-1.889**	0.123	-0.255
	(0.692)	(0.211)	(0.142)	(0.955)	(0.333)	(0.216)
Grandmother	0.285*	0.124**	0.023			
	(0.157)	(0.057)	(0.028)			
Pregnancy$\leq$18 x Grandmother	0.521*	0.017	-0.015			
	(0.316)	(0.073)	(0.054)			
High Failure School				0.143	0.110***	-0.004
				(0.104)	(0.040)	(0.029)
Pregnancy$\leq$18 x High Failure School				0.410*	-0.006	-0.016
				(0.225)	(0.052)	(0.029)
Observations	1,741	1,741	1,741	1,741	1,741	1,741
R-squared	0.169	0.175	0.058	0.184	0.177	0.057
First Stage F-Statistic	17.67	17.67	17.67	16.76	16.76	16.76
Mean Dependent Var	10.94	0.524	0.147	10.98	0.524	0.157

Notes: Pregnancy $\leq$ 18 is defined one for those who got pregnant before 18 and zero for those who did not. "Grandmother" is an indicator variable for whether the teenager's mother was alive in her teen years, and zero otherwise. "High Failure School" is eaiql The variable was constructed using the CAPS data set. "Years of Education" indicates the Years of education completed. "Took Matric" is an indicator variable of whether the individual took the Matric exam and "Post Secondary Education" is equal to one if the sampled woman kept studying after high school, and zero otherwise. The "First Stage F-stat" comes from the a first stage regression where "Pregnancy ¡ 18" is instrumented with the number of teenage fertile years. The Mean Dependent Var indicates the average of the Outcome if "Pregnancy ¡ 18" is equal to zero. Controls include the race, an asset index, height, number of siblings, and the indicator variables for religion, language spoken by her family. Cohort and sampling-location fixed effects are included. Standard errors in parentheses, clustered at the sampling location level. *** p $\leq$ 0.01, ** p $\leq$ 0.05, * p$\leq$0.1

A Appendix Tables

Table A1: Samples of Women across waves

	2002 Wave 1 (1)	2003-2004 Wave 2 (2)	2005 Wave 3 (3)	2006 Wave 4 (4)	2009-2010 Wave 5 (5)
Panel A: All Women Interviewed Per Wave					
Observations	2,612	2,140	1,911	1,881	1,799
Percentage	100%	81.93%	73.16%	72.01%	68.87%
Panel B: Analytical Sample					
Observations	2,455	2,020	1,815	1,741	1,707
Percentage	100%	82.28%	73.93%	70.96%	69.53%

Source: CAPS Data waves 1-5

Notes: Panel A includes all sampled women interviewed in each wave of the CAPS data set. Panel B includes sampled women that where interviewed and also had reached menarche between the ages of 10 and 17 and lived in sampling clusters with at least one other sampled women.

Table A2: Summary Statistics by Self-reported Early Pregnancy

	No Early Pregnancy	Pregnancy≤18	Difference
	(1)	(2)	(3)
Demographic Characteristics			
Age in 2002	17.689	17.865	0.176
	(2.445)	(2.448)	
% Coloured	0.438	0.555	0.117***
	(0.496)	(0.498)	
% Black	0.497	0.443	-0.055*
	(0.500)	(0.497)	
Adult Height	158.271	156.808	-1.463***
	(7.987)	(8.597)	
Mother's Education	8.447	7.559	-0.888***
	(3.191)	(2.785)	
# Full Siblings	2.286	2.416	0.162
	(1.739)	(1.813)	
Household Size	5.691	6.256	0.565***
	(2.528)	(2.755)	
Not Born in WC	0.288	0.264	-0.024
	(0.453)	(0.442)	
Sexual Activity			
Age 1st Partner	19.923	19.378	-0.546***
	(3.128)	(3.395)	
Age 1st Active	17.811	15.928	-1.883***
	(2.326)	(1.374)	
Observations	1394	348	

Notes: Pregnancy early is equal to 1 if the individual reported a pregnancy before the age of 17, and 0 otherwise. "Adult Height"is the height of the sampled women, measured in 2006. "Not Born in WC" is an indicator variable for individuals born outside the Western Cape, and "Mother's Education" is the total number of years completed by mother. "# Full Siblings" is number of siblings. Finally "Age of 1st. Partner" and "Age 1st Active" describe the ages of the first sexual partner and the age in which women in the sample became active respectively. The Diff column is the difference in means of Columns (1) and (2), where a T-test where the hypothesis is that the coefficient is equal to zero. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$

Table A3: Summary Statistics by Early vs Late age at menarche

	Menarche Late	Menarche Early	Diff.
	(1)	(2)	(3)
	Menarche $\geq$ 14	Menarche <14	Difference
	(1)	(2)	(3)
Coloured	0.298	0.61	0.312***
	(0.458)	(0.488)	
Black	0.667	0.321	-0.346***
	(0.472)	(0.467)	
Height - cm	158.01	157.96	-0.052
	(8.80)	(7.48)	
Married ever	0.175	0.178	0.003
	(0.380)	(0.382)	
Hh Size	5.808	5.799	-0.009
	(2.734)	(2.442)	
Married ever	0.175	0.178	0.003
	(0.380)	(0.382)	
Mother attended school	0.856	0.885	0.028*
	0.351	0.319	
	829	912	

Notes: Menarche Early is defined is equal to one if the sampled individual has reached menarche before the age of 14. "Adult Height"is measured in 2006. "Not Born in WC" is an indicator variable for individuals born outside the Western Cape, and "Mother's Education" describe the total number of years completed by mother. "# Full Siblings" and "Household Size" refer to the number of siblings and household size, respectively. The label ln(Hhold Inc.) estimates the natural log of the household income and "Literacy Exam" is the normalized grade of the Young Adult Literacy and Numeracy Evaluation. Finally "Age of 1st. Partner" and "Age 1st Active" describe the ages of the first sexual partner and the age in which women in the sample became active respectively. The Diff column is the difference in means of Columns (1) and (2), where a T-test where the hypothesis is that the coefficient is equal to zero. *** p$\leq$0.01, ** p$\leq$0.05, * p$\leq$0.1

Table A4: Determinants of Attrition in the Sibling Sample

	Age at Puberty		Teenage pregnancy		Teenage Pregnancy - Age > 18 in 2002	
	(1)	(2)	(3)	(4)	(5)	(6)
Age at Menarche	-0.000	-0.009				
	(0.006)	(0.007)				
Pregnancy ≤ 18			-0.088***	-0.044	-0.078**	-0.011
			(0.023)	(0.022)	(0.032)	(0.033)
Race - Black		-0.345***		-0.338***		-0.418***
		(0.042)		(0.042)		(0.054)
Race - Coloured		-0.324***		-0.318***		-0.369***
		(0.039)		(0.039)		(0.048)
Literate Mother		-0.109***		-0.104***		-0.122***
		(0.030)		(0.030)		(0.036)
Not Migrant		0.146***		0.142***		0.184***
		(0.026)		(0.026)		(0.038)
HH size		-0.012***		-0.011***		-0.012**
		(0.004)		(0.004)		(0.005)
Muslim		0.018		0.021		0.000
		(0.030)		(0.030)		(0.042)
Cohort - 1987		-0.013		-0.046		
		(0.034)		(0.036)		
Cohort - 1986		0.030		-0.018		
		(0.036)		(0.038)		
Cohort - 1985		0.025		-0.026		
		(0.035)		(0.036)		
Cohort - 1984		0.078**		0.027		
		(0.038)		(0.038)		
Cohort - 1983		0.115***		0.069		0.039
		(0.040)		(0.042)		(0.036)
Cohort - 1982		0.179***		0.129***		0.101***
		(0.041)		(0.042)		(0.038)
Cohort - 1981		0.151***		0.099**		0.072**
		(0.041)		(0.040)		(0.035)
Cohort - 1980		0.149***		0.098**		0.069*
		(0.042)		(0.043)		(0.039)
Observations	2,577	2,572	2,612	2,607	1,435	1,432
R-squared	0.000	0.118	0.005	0.112	0.004	0.122

Notes: The sample in columns (1) and (2) includes women who were interviewed in 2002. Columns (3) and (4) include all the women in 2002 and columns (5) and (6) women were at least 18 years of age in 2002. "Age at Menarche" is the age at menarche and "Pregnancy≤18" is equal to 1 if women reported a pregnancy by the age of 18 and zero otherwise. *** p<0.01 ** p<0.05 * p<0.1

Table A5: OLS Estimation of the Effects of Teenage Pregnancy

Panel A: School Progression

	Failed grade	Age for Grade	Hazard Drop Out	
	(1)	(2)	(3)	
Pregnant_{icst}	0.242***	0.687***	0.017***	
	(0.031)	(0.051)	(0.001)	
Observations	1,741	1,741	1,741	
R-squared	0.160	0.178		

Panel B: Schooling Attainment

	Years of Education	Took Matric Exam	Post Secondary Schooling	
	(1)	(2)	(3)	
Pregnancy $\leq$ 18	-1.034***	-0.233***	-0.043***	
	(0.115)	(0.026)	(0.014)	
Observations	15,170	15,170	14,354	
R-squared	0.462	0.457	0.405	

Panel C: Labor Force Participation

	Active at			
	age 19	age 20	age 21	age 22
	(1)	(2)	(3)	(4)
Pregnancy $\leq$ 18	0.115***	0.059*	0.048	0.078***
	(0.030)	(0.033)	(0.033)	(0.029)
Observations	1,741	1,741	1,741	1,741
R-squared	0.071	0.037	0.061	0.083

Notes: Birth $\leq$ 18 is defined one for those who gave before 18 and zero for those who did not. Controls include the race, an asset index, height, number of siblings, and the indicator variables for religion, language spoken by her family. "Pregnant_{icst}" is defined as 1 for individual i of cohort c, who lives in sampling cluster s and observed in year t if she got pregnant and zero; if she did not get pregnant during that year."Years of Education" indicates the Years of education completed. "took Matric" is an indicator variable of whether the individual took the Matric exam and "Post Secondary Education" is equal to one if the sampled woman kept studying after high school, and zero otherwise. Active at each specific age is an indicator variable for whether the sampled women reported being in the labor force at that age.

Table A6: First stage - Different Thresholds for Early Pregnancy

	Pregnancy≤16 (1)	Pregnancy≤17 (2)	Pregnancy≤18 (3)
Fertile Years	0.023***	0.031***	0.032***
	(0.005)	(0.006)	(0.008)
Observations	1,741	1,741	1,741
R-squared	0.014	0.016	0.009
First stage F	20.67	25.53	17.61

Notes: "Pregnancy ≤ 16", "Pregnancy ≤ 17" and "Pregnancy ≤ 18" are defined one for those who got pregnant before the indicated age and zero for those who did not. Also, the "Number of Fertile Years" indicates the number of years passed since menarche until 17. Controls include the race, an asset index, height, number of siblings, the log of the household level and indicator variables for religion, language spoken by her family, place of birth and residence. Cohort and sampling-location fixed effects are included. In the panel data set, Pregnant in year t is an indicator variable for whether the individual got pregnant in year t and Fertile is an indicator variable of whether a woman was fertile in a specific year. In both data sets, standard errors in parentheses, clustered at the sampling location level.*** p ≤ 0.01, ** p ≤ 0.05, * p≤0.1

Table A7: First stage - First stage with different top values in the instrument

	Pregnancy≤18					
	(1)	(2)	(3)	(4)	(5)	(6)
Fertile Years_{18}	0.016**	0.031***				
	(0.006)	(0.008)				
Fertile Years_{19}			0.016**	0.031***		
			(0.006)	(0.008)		
Fertile Years_{20}					0.016**	0.031***
					(0.006)	(0.008)
Observations	1,741	1,741	1,741	1,741	1,741	1,741
R-squared	0.003	0.011	0.003	0.011	0.003	0.011
First stage F	6.626	17.57	6.626	17.57	6.626	17.57

Notes: "Pregnancy ≤ 18" is defined one for those who got pregnant before 18 and zero for those who did not. Also, the "Fertile Years" indicates the number of years passed since menarche until 18, 19 and 20 in each Row. Controls include the race, an asset index, height, number of siblings, the log of the household level and indicator variables for religion, language spoken by her family, place of birth and residence. Cohort and sampling-location fixed effects are included. In the panel data set, Pregnant in year t is an indicator variable for whether the individual got pregnant in year t and Fertile is an indicator variable of whether a woman was fertile in a specific year. In both data sets, standard errors in parentheses, clustered at the sampling location level.*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$

Table A8: Schooling Progression - Until the age of 24

	Failed grade (1)	Age for Grade (2)	Hazard Drop Out (3)
Panel A: IV Estimation			
Pregnant$_{icst}$	0.036	0.317***	0.095***
	(0.092)	(0.012)	(0.007)
Observations	15579	15579	15474
Panel B: Reduced Form Estimation			
Post Menarche$_{itcs}$	0.111***	0.315***	0.423*
	(0.033)	(0.009)	(0.201)
Observations	15,646	15579	15474
Mean Dependent Var	0.113	1.091	0.359

Notes: "Pregnant$_{icst}$" is defined as 1 for individual i of cohort c, who lives in sampling cluster s and observed in year t if she got pregnant and zero; if she did not get pregnant during that year. Columns (1) and (2) contain the OLS and IV estimates of the effects on the likelihood of failing a grade in year t and the degree to which a sampled women lagged behind her secondary education in that same year. Column (3) contains the hazard rated for dropping out. In Panel A Column (3), the coefficient is calculated with out considering endogeneity. In panel B, endogeneity is accounted for using a linear first stage of the effects of teenage fertility on pregnancy. The First Stage statistic comes from the a first stage regression where "Pregnant$_{icst}$" is instrumented with an indicator variable of whether the teenager was fertile in year t. The Mean Dependent Var indicates the average of the outcome if Pregnancy in year t is equal to zero. Controls includes height, Standard errors in parentheses, clustered at the smallest sampling location level. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$

Table A9: Estimation of the Effects of Teenage Pregnancy - Schooling Attainment
Black Women

	Years of Education (1)	Took Matric Exam (2)	Post Secondary Schooling (3)
Panel A: OLS Estimation			
Pregnancy $\leq 18_i$	-2.261*	-0.171	-0.107
	(1.170)	(0.371)	(0.253)
Observations	841	841	841
R-squared	0.080	0.177	0.048
First Stage F-stat	14.71	14.71	14.71
Panel B: Reduced Form Estimation			
Fertile Years$_i$	-0.083*	-0.006	-0.004
	(0.045)	(0.014)	(0.010)
Observations	846	846	846
R-squared	0.117	0.156	0.057
Mean Dependent Var	10.82	0.466	0.118

Notes: Pregnancy $\leq$ 18 is defined one for those who got pregnant before 18 and zero for those who did not. Controls include height and cohort fixed effects. "Years of Education" indicates the Years of education completed. "Took Matric" is an indicator variable of whether the individual sat for Matric exam and "Post Secondary Education" is equal to one if the sampled woman kept studying after high school, and zero otherwise. The Mean Dependent Var indicates the average of the Outcome if "Pregnancy $\leq$ 18" is equal to zero. Cohort and sampling-location fixed effects are included. *** p $\leq$ 0.01, ** p $\leq$ 0.05, * p$\leq$0.1

Table A10: Inverse Probability Weights Estimation

Panel A: School Attainment

	Years Education	Took Matric	Post Secondary Educ.
	(1)	(2)	(3)
Pregnancy $\leq 18_i$	-1.275**	-0.096	-0.122
	(0.629)	(0.168)	(0.106)
Observations	1,735	1,735	1,735
Mean Dependent Var	10.94	0.524	0.147
First Stage F-stat	22.85	22.85	22.85

Panel B: Labor force participation

	At age 19	At age 20	At age 21	At age 22
	(1)	(2)	(3)	
Pregnancy $\leq 18_i$	0.176	0.314*	0.042	0.074
	(0.188)	(0.173)	(0.175)	(0.169)
Observations	1,735	1,735	1,735	1,735
First Stage F-stat	22.85	22.85	22.85	22.85
Mean Dependent Var	0.566	0.461	0.421	0.394

Notes: Pregnancy $\leq$ 18 is defined one for those who got pregnant before 18 and zero for those who did not. Controls include height and cohort fixed effects. "Years of Education" indicates the Years of education completed. "Took Matric" is an indicator variable of whether the individual sat for Matric exam and "Post Secondary Education" is equal to one if the sampled woman kept studying after high school, and zero otherwise. The Mean Dependent Var indicates the average of the Outcome if "Pregnancy $\leq$ 18" is equal to zero. Cohort and sampling-location fixed effects are included. *** p $\leq$ 0.01, ** p $\leq$ 0.05, * p$\leq$0.1

Table A11: IV Estimation of the Effects of Teenage Birth - Schooling Attainment

	Years of Education (1)	Took Matric Exam (2)	Post Secondary Schooling (3)
Panel A: OLS Estimation			
Birth $\leq 18_i$	-1.087***	-0.224***	-0.041***
	(0.117)	(0.027)	(0.014)
Observations	1,741	1,741	1,741
R-squared	0.167	0.166	0.055
Panel B: IV Estimation			
Birth ≤ 18	-2.380**	0.056	-0.337
	(1.196)	(0.378)	(0.267)
Observations	1,741	1,741	1,741
R-squared	0.082	0.117	-0.092
First Stage F-stat	10.64	10.64	10.64
Mean Dependent Var	10.94	0.524	0.147

Notes: Birth $\leq$ 18 is defined one for those who gave before 18 and zero for those who did not. Controls include the race, an asset index, height, number of siblings, and the indicator variables for religion, language spoken by her family. "Years of Education" indicates the Years of education completed. "Took Matric" is an indicator variable of whether the individual sat for Matric exam and "Post Secondary Education" is equal to one if the sampled woman kept studying after high school, and zero otherwise. The "First Stage F-stat" comes from the a first stage regression where "Pregnancy $\leq$ 18" is instrumented with the number of teenage fertile years. The Mean Dependent Var indicates the average of the Outcome if "Pregnancy $\leq$ 18" is equal to zero. Cohort and sampling-location fixed effects are included. Standard errors in parentheses, clustered at the sampling location level. *** p $\leq$ 0.01, ** p $\leq$ 0.05, * p$\leq$0.1

Table A12: Estimation of the Effects of Teenage Pregnancy - Schooling Attainment
Inverse probability Weights

	Years of Education (1)	Took Matric Exam (2)	Post Secondary Schooling (3)
Panel A: IV Estimation			
Pregnancy$\leq 18_i$	-1.275**	-0.096	-0.122
	(0.629)	(0.168)	(0.106)
Observations	1,735	1,735	1,735
R-squared	0.162	0.169	0.048
First Stage F-stat	22.85	22.85	22.85
Panel B: Reduced Form Estimation			
Fertile Years$_i$	-0.075*	-0.006	-0.007
	(0.041)	(0.011)	(0.007)
Observations	1,735	1,735	1,735
R-squared	0.429	0.395	0.396
Comparison Mean Var	10.98	0.524	0.157

Notes: Pregnancy $\leq$ 18 is defined one for those who got pregnant before 18 and zero for those who did not. The Number of Children born $>$ the age 19 is equal to the number of pregnancies each women reported between the age of 19 and 2009. Additionally, the Total Number of Children indicates the number of children each women in the sample has has before 2009, regardless of the age she gave birth. Controls include the race, an asset index, height, number of siblings, and the indicator variables for religion, language spoken by her family. The "First Stage F-stat" comes from the a first stage regression where "Pregnancy $\leq$ 18" is instrumented with the number of teenage fertile years. The Mean Dependent Var indicates the average of the Outcome if "Pregnancy $\leq$ 18" is equal to zero. Cohort and sampling-location fixed effects are included. Standard errors in parentheses, clustered at the sampling location level. *** p $\leq$ 0.01, ** p $\leq$ 0.05, * p$\leq$0.1

Table A13: Estimation of the Effects of Teenage Pregnancy - Labor Outcomes 2006

	Employed	Hours Worked	Earnings	Accept a Dom. Worker Position	Accept Security Guard Position
	(1)	(2)	(3)	(4)	(5)
Panel A: IV Estimation					
Pregnancy $\leq 18_i$	0.088	3.150	1,355.52	0.078	0.083
	(0.287)	(3.336)	(1,132.56)	(0.219)	(0.237)
Observations	1,741	1,741	1,741	1,741	1,741
R-squared	0.106	0.034	0.010	0.042	0.046
First Stage F-stat	17.61	17.61	17.61	17.61	17.61
Fertile Years_i	0.003	0.099	42.72	0.002	0.003
	(0.009)	(0.103)	(33.81)	(0.007)	(0.007)
Observations	1,741	1,741	1,741	1,741	1,741
R-squared	0.115	0.102	0.112	0.038	0.044
Mean Dependent Var	0.431	5.259	1263	0.234	0.329

Notes: PP for those who did not.Controls include the race, an asset index, height, number of siblings, and the indicator variables for religion, language spoken by her family. The Mean Dependent Var indicates the average of the Outcome if Pregnancy ≤ 18 is equal to zero.Cohort and sampling-location fixed effects are included. Standard errors in parentheses, clustered at the sampling location. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$

Table A14: Interactions between teenage pregnancy and grandfather alive

	Years of Education (1)	Took Matric Exam (2)	Post Secondary Schooling (3)
Pregnant $\leq 18_i$	-1.754	0.105	-0.152
	(1.083)	(0.272)	(0.195)
Grand Father alive	0.051	0.008	0.015
	(0.077)	(0.024)	(0.020)
Pregnant $\leq 18_i$ x Grand father alive	-0.104	-0.057	-0.054
	(0.185)	(0.047)	(0.034)
Observations	1,741	1,741	1,741
R-squared	0.166	0.173	0.059
First stage F	17.58	17.58	17.58
Mean Dependent Var	10.94	0.524	0.147

Notes: Pregnancy $\leq$ 18 is defined one for those who got pregnant before 18 and zero for those who did not. "Grandfather alive-Teens" is an indicator variable for whether the teenager's father was alive in her teen years, and zero otherwise. The variable was constructed using the CAPS data set. "Years of Education" indicates the Years of education completed. "Took Matric" is an indicator variable of whether the individual sat for Matric exam and "Post Secondary Education" is equal to one if the sampled woman kept studying after high school, and zero otherwise. The "First Stage F-stat" comes from the a first stage regression where "Pregnancy ¡ 18" is instrumented with the number of teenage fertile years. The Mean Dependent Var indicates the average of the Outcome if "Pregnancy ¡ 18" is equal to zero. Controls include the race, an asset index, height, number of siblings, and the indicator variables for religion, language spoken by her family. Cohort and sampling-location fixed effects are included. Standard errors in parentheses, clustered at the sampling location level. *** p $\leq$ 0.01, ** p $\leq$ 0.05, * p$\leq$0.1

Table A15: IV Estimation of the Effects of Teenage Pregnancy - Marriage and Fertility

	Married Ever	Subsequent Fertility	Total Fertility
	(1)	(2)	(3)
Panel A: IV Estimation			
Pregnancy$\leq 18_i$	0.480	0.517	-0.744**
	(0.303)	(0.360)	(0.363)
Observations	1,741	1,741	1,741
R-squared	-0.014	0.252	0.197
First Stage F-stat	17.59	17.59	17.59
Panel B: Reduced Form Estimation			
Fertile Years$_i$	0.015	0.016	-0.023*
	(0.009)	(0.012)	(0.013)
Observations	1,741	1,741	1,741
R-squared	0.082	0.127	0.150
Mean Dependent Var	0.354	0.536	0.536

Notes: Pregnancy $\leq$ 18 is defined one for those who got pregnant before 18 and zero for those who did not. Subsequent Fertility is equal to the number of pregnancies each women reported between the age of 19 and 2009. Additionally, Total Fertility indicates the number of children each women in the sample has has before 2009. Controls include the race, an asset index, height, number of siblings, and the indicator variables for religion, language spoken by her family. Married ever is defined as one for those who got married at some point before 2009, and 0 otherwise. The "First Stage F-stat" comes from the a first stage regression where "Pregnancy $\leq$ 18" is instrumented with the number of teenage fertile years. The Mean Dependent Var indicates the average of the Outcome if "Pregnancy $\leq$ 18" is equal to zero. Cohort and sampling-location fixed effects are included. Standard errors in parentheses, clustered at the sampling location level. *** p $\leq$ 0.01, ** p $\leq$ 0.05, * p$\leq$0.1

Table A16: IV Individual Level Estimation of the Effects of Teenage Pregnancy

Panel A: Education outcomes

	Yrs of Education	Write Matric Exam	Post Secondary Schooling
	(1)	(2)	(3)
Pregnancy ≤ 18	-2.075**	-0.021	-0.256*
	(1.052)	(0.327)	(0.136)
Observations	1,717	1,717	1,717
R-squared	0.179	0.203	-0.271
F-stat	12.82	12.82	12.82
Comparison Mean	10.97	0.436	0.0655

Panel B: Labor market outcomes

	Working at	Working at 20	Working at 21	Working at 22
	(1)	(2)	(3)	(4)
Pregnancy ≤ 18	0.112	3.089	-0.316	-0.096
	(0.294)	(3.938)	(0.289)	(0.292)
Observations	1,717	1,714	1,717	1,717
R-squared	0.027	-0.010	0.120	0.132
F-stat	12.82	12.91	12.82	12.82
Comparison Mean	0.587	5.940	0.686	0.453

Notes: Pregnancy ≤ 18 is defined 1 for those who got pregnant before 18 and zero for those who did not. The most important control is a PCA index of child trauma questions. Other controls include the race, an asset index, height and hip sizes, number of siblings, the log of the household level and the indicator variables for religion, language spoken by her family, place of birth and residence. Cohort and sampling-location fixed effects are included. The Comparison Mean indicates the average of the Outcome if Pregnancy ≤ 18 is equal to zero. Standard errors in parentheses, clustered at the sampling location. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.1$

Table B1: Variable Construction by Wave

	2002 Wave 1 (1)	2003-2004 Wave 2 (2)	2005 Wave 3 (3)	2006 Wave 4 (4)	2009-2010 Wave 5 (5)
Health, Fertility & Marriage					
Age at Menarche	✓	-	✓ for those w/o data in wave 1	-	-
Pregnancy	Retrospective yearly for 1979-2002	-	✓	✓	✓
Births	Retrospective yearly for 1979-2002	-	✓	✓	✓
Marriage	Retrospective yearly for 1979-2002	-	Retrospective yearly for 2003-2005	✓	Retrospective yearly for 2007-2009
Adult Height	-	-	-	✓	-
Education					
Literacy Exam	✓	-	-	-	-
Years of Education	Retrospective yearly for 1979-2002	✓	✓	✓	Retrospective yearly for 2007-2009
Grade Progress	Retrospective yearly for 1979-2002	✓	✓	✓	Retrospective yearly for 2007-2009
Matriculation	Retrospective yearly for 1979-2002	✓	✓	✓	Retrospective yearly for 2007-2009
Employment					
Employment	✓	✓	✓	✓	✓
Employment Charact.	✓	✓	✓	✓	✓
Control Variables					
Background Charac.	✓	-	-	-	-
Childhood Info	✓	-	-	-	-
Parents Demographics	✓	-	-	Health	-
Parents Death	Retrospective yearly for 1979-2002	-	Retrospective yearly for 2003-2005	✓	Retrospective yearly for 2007-2009
Household Charac.	✓	-	-	-	-

Source: CAPS Data waves 1-5

The ✓ indicates that the information is available in that specific wave. “Retrospective yearly for 1979-2002” indicates that questions about the specif category were asked for each year since the year of birth until 2002.

Table B2: Literature Review - High income Countries

	Country (1)	Identification Strategy (2)	Outcomes (3)	Results (5)
Heiland, Korenmana, Smith (2019)	US	HH FE	Yrs of education	≈ zero in outcomes
Zito (2018)	US	PSM	Attitudes & norms	↑ risk aversion. No self-worth or relationship effects
Duncan, Lee, Rosales-Rueda, Kalil (2018)	US	OLS, HH FE	Yrs of education & behavior problems	1 yr delay in birth ↑ 0.02 to 0.04 SD in school achievement & ↓ problems
Diaz & Fiel (2016)	US	Smoothing-diff. & IPW	Educational attainment and earnings	↓ college completion, early earnings
Yakusheva (2011)	US	PSM	Yrs of education	≈ 0 for high-risk teens & low effects for teens at low risk of TP
Ashcraft, Fernández-Val & Lang (2006, 2013)	US	IV (miscarriages)	Yrs Education, GED Score, employment & marriage	GED ↓ by about 5 pp & ↓ 0.15 yrs educ. Employment: ↓ 5 pp. Marriage ↓ 3 pp.
Kane, Morgan, Harris, Guilkey (2013)	US	OLS, PSM & ML	Yrs Education	↓ 0.7 and 1.9 yrs. of education
Lee (2010)	US	PSM	Education, labor force	↓ early socioeconomic outcomes
Fletcher & Wolfe (2009)	US	OLS & IV (miscarriages)	Graduation, earnings	↓ 5-10 pp high school graduation, ↓ $1,000 to $2,400 annual income
Francesconi (2008)	UK	OLS, HH FE	Yrs education, bmi	↓ yrs education, employment. ↓ Child health in single parent
Hotz, McElroy & Sanders (2005)	US	IV (miscarriages)	Yrs of education, earnings	No education effects, ↑ earnings at older ages
Kaplan, Goodman, Walker (2004)	UK	OLS, PSM & IV (miscarriages)	Education attainment, employment	↓ large educ. attainment, no labour effects
Levine & Painter (2003)	US	PSM, HH FE	Yrs Education	↓ yrs education & bigger for teenagers at risk
Bitler (2001)	US	IV(Abortion laws)	Timinig of abortions	≈ zero in outcomes
Klepinger, Lundberg, Plotniek (1995, 1999)	US	IV (teenage fertility) & HH FE (1999)	Yrs of education & wages	↓ -2.14 yrs of education, ↓ 2 yrs work experience
Ribar (1994)	US	IV (age at menarche)	High school completion	↓ labor force participation, hours of work
Geronimus & Korenman (1992)	US	HH FE	High school completion	↓ small effects in school completion

Table B3: Literature Review- Low and middle income Countries

	Country (1)	Identification Strategy (2)	Outcomes (3)	Results (4)
Branson & Byker (2019)	South Africa	Diff-in-Dif - Policy	Number of births, yrs. of education	↓ 6.3% birth rate , ⇑ 30% monthly earnings
Ranchhod, Lam, Leibbrandt, & Marteleto (2011)	South Africa	PSM	High school graduation	↓ 5.9 & 2.7 pp at ages 20-22. Later catch-up
Ardington, Menendez, Mutevedzi (2015)	Rural South Africa	OLS, PSM, HH FE	Yrs of education, child mortality	↓ 0.67 & 0.79 years. High mortality by 30
Branson, Ardington, Leibbrandt (2015)	South Africa	PS reweighting	Birth weight, height & stunting	6.5 pp low bw, 18.5 pp of stunting
Berthelon & Kruger (2017)	Chile	HH FE	Graduation, enrollment, employment	↓ high school grad. & higher educ., and ↓ 0.45 yrs. No labor effects
Urdinola & Ospino (2015)	Colombia	TM (18–19) vs. older mothers (20–21)	Job type & domestic violence	↓ 0.091 job quality, ↑ severe DV 0.051 pp, ↑ 1.2% child mort.
Arceo-Gomez & Campos-Vasquez (2014)	Mexico	PSM	Enrollment, yrs of education, employment	↓ 27-33 pp in enrollment, 1-1.2 yrs. educ., ↓ 13-15 employment
Azevedo, Lopez-Calva, Perova (2012)	Mexico	Miscarriages vs teen births	Yrs of education & income	↑ 0.34 yrs of education, ↑ 21 pp more likely to be employed, but ↑ assistance income by 36 %
Narita & Montoya Diaz (2016)	Brazil	PSM	High school completion, employment	↓ TP 1 SD explains ↑ 9.2% in high school comp. & ↑ 5.4% part.
Herrera, Almanza & Sahn (2018)	Madagascar	IV - access & exposure to condoms	Yrs of education & test scores	↑ drop out by 42 pp. ↓ 1.1-1.5 sd test scores in math & French

Notes: TM is short for Teenage Pregnancy, PSM for Propensity Score Matching, and FE Fixed effects