The Precocious Period: Impact of Early Menarche on Schooling in India

Madhulika Khanna *

June 4, 2020

Abstract

This paper studies the impact of early onset of menstruation on school enrollment in India, where the onset of menstruation (menarche) has immense socio-cultural importance, and menarche marks a child's transition into womanhood. Estimates based on a difference-in-differences model show that starting menses before age twelve decreases school enrollment by 13%. While understanding the role of menarche—a universal biological event for all women—in determining education is a critical end in itself, it also has implications for the effect of early nutrition on girls' well-being. Better-nourished girls typically reach menarche earlier, and if menarche impedes schooling, their gains due to better nutrition may be undercut. The effect of menarche on school enrollment is stronger if girls live in communities where the perceived safety among children is low or if they belong to social groups with restrictive gender norms. Conversely, dropout rates are lower in communities with higher expected wages for female-dominated professions. Jointly, these findings suggest that the interaction between cultural norms and the local economic opportunities shapes female education decisions in response to menses.

JEL Classification: J16; I24; O12

Keywords: School Enrollment; Education; Gender; Social Norms; India

*Madhulika Khanna: Georgetown University, 3700 O St NW, Washington, DC 20057, mk1469@georgetown.edu.
Acknowledgements: I am very grateful to Garance Genicot and Martin Ravallion for their guidance over the course of this project. This paper also benefited from the helpful comments from Ashish Aggarwal, Sarah Baird, Brad Barham, Anja Benshaul-Tolonen, Becka Brolinson, Caitlin Brown, David Eil, Deepti Goel, Maria Hernandez, Pamela Jakiela, Shareen Joshi, Nishtha Kochhar, Catherine Maclean, Juan Margitic, Umberto Muratori, Deniz Sanin, Dario Sansone, and Andrew Zeitlin as well as seminar participants at Georgetown University, Towson University, Colgate University, and conference participants at the Sustainability and Development Conference, the Midwest Economic Association Conference, the Washington Area Development Economics Symposium, the Midwest International Economic Development Conference, the Georgetown Center for Economic Research Conference, and the Southern Economic Association Conference. I have also benefited from fruitful discussions with several seminar speakers and co-workers at Georgetown University. The usual caveats apply.

"*It is a vicious circle. Women are deprived of rights because of their lack of education, and their lack of education results from their lack of rights.*"

-Leo Tolstoy (in Anna Karenina)

1 Introduction

Despite universal primary school enrollment in India, many adolescent girls drop out after completing eight years of compulsory schooling (ASER 2018). Consequently, a notable gender gap in school enrollment emerges and grows during teenage years (Muralidharan and Prakash 2017). The existing literature has posited many explanations for the emergence of a gender gap in schooling in this setting, including the burden of housework (Sundaram and Vanneman 2008), lower returns to female education (Chamarbagwala 2008), and the social barriers that impede access to schooling due to the allocation of schooling infrastructure across communities that ethnically are organized communities (Jacoby and Mansuri 2011). In this paper, I identify a biological driver of the decrease in school enrollment during adolescence: the onset of menstruation (menarche). To the best of my knowledge, this is the first paper that seeks to establish a causal link between menarche and schooling.

This paper is set in India, where menarche completely alters the norms of a girl's engagement with her family and her community (Seymour 1999). Parents mark their daughter's menarche with rituals and celebrations, and in some communities, they seclude their daughter within the house as part of these rituals (Dharmalingam 1994). Menstruating women are excluded from participating in religious practices or even entering the kitchen (Mahon and Fernandes 2010; Garg, Sharma, and Sahay 2001). Menarche also signifies the beginning of a critical period in her life when "she has acquired the capability to reproduce but she has no authority to do so" and her "vulnerability is at its peak" (Dube 1988). Once the girl reaches menarche, her family's honor is associated with its capacity to maintain their daughter's virginity. To ensure their daughter's safety, parents often restrict her mobility.

I study if the familial and societal responses to menarche in this context affect girls' school enrollment and characterize the relationship between early menarche and schooling in a conceptual framework where the parents decide on the optimal years of education and age at marriage for their daughter. The empirical evidence is consistent with the predictions of this framework and shows that an early onset of menstruation decreases school enrollment.

The primary empirical challenge of estimating the impact of early menarche is that girls experience many transitions, such as the transition to secondary schooling and increased burden of housework, during this time period. Any association between early menarche and enrollment also reflects these confounding changes. To address these concerns, I use a difference-in-differences design that exploits the variation in the timing of menarche within the same age cohort. I use data from a representative sample of 1,000 children in the state of Andhra Pradesh (now Andhra Pradesh and Telangana). These children were tracked over the course of their childhood, adolescence, and early adulthood. Importantly, the second round of the survey, during which the children were about twelve years old, asked girls if they had started menstruating yet. I use the enrollment status of those girls who had not reached menarche by age twelve as the counterfactual for those who experienced an earlier menarche.[1]

The main result of this paper is that reaching menarche before age twelve causes a 13.4% decrease in school enrollment rate (equivalent to a 12.2% point decrease over the counterfactual enrollment rate of 90.6%). Since the identification strategy is based on a comparison of children within the same age cohort, concerns regarding confounding changes in the quality of schooling are not relevant. The difference-in-differences design within a cohort is valid if the factors that affect the timing of menarche (such as nutrition in the early years) do not themselves affect the enrollment decision. To address this concern, I show that in addition to being associated with higher height and weight at age eight, early menarche is also associated with a more rapid increase in body mass index (BMI) and height between ages eight and twelve. Insofar as better nutrition leads to higher educational attainment, these patterns suggest that the effect of early menarche on school enrollment estimated using a difference-in-differences design, if anything, is underestimated. I address additional concerns regarding the endogenous timing of menarche by showing that the results are robust to the inclusion of individual fixed effects, which account for anything that may be different about the girls who reach menarche before age twelve.[2]

The result that early menarche decreases school enrollment implies that the aphorism that gains in early nutrition enhance schooling and later life outcomes may be weakened for girls. The predominant framework of conceptualizing this link from childhood to later life outcomes is a *virtuous cycle* of human development that builds in a complementarity between nutrition and learning across ages (Cunha and Heckman 2007). Today, this association between nutrition, education, and well-being is a core foundation for policy design (Currie and Almond 2011; Attanasio 2015). However, improvements in childhood nutrition are also linked to the timing of menarche in that better-nourished girls typically reach menarche earlier. The issue arises as to whether early menarche has adverse consequences for girls' schooling in these contexts, severing the symbiotic relationship between childhood nutrition, schooling, and economic returns in later life.

In this sample, the girls who reach menarche before age twelve are taller and heavier than other girls, and they outperform other girls at math tests and perform as well as boys in their cohort.[3] However, these girls are also more likely to drop out of school. By age fifteen, they lose any learning advantage they might have had on account of being healthier and fall behind boys in cognitive tests. Thus, if certain milieu foster gender-restrictive norms such that early menarche impedes schooling, the *virtuous cycle* of human development may be undermined for girls. Early menarche becomes another channel that fosters and sustains gender gaps in human development.

This paper also demonstrates that the pathway from nutrition to learning and returns in later

1. In this sample, a little more than a quarter of the surveyed girls had reached menarche before age twelve.

2. This is not the preferred specification because some of the analysis in compares outcomes across two groups in a given period, and individual fixed effects cannot be used in that setup.

3. These tests were administered to all the surveyed children, and not just those who were in school.

life is complex for girls. Generally, early childhood is a critical period, and investments in human capital during early childhood have the highest returns (see Cunha and Heckman (2007) for a review). Among girls, however, investments in childhood also influence the timing of menarche, and adolescence is critical for their human capital formation. Consequently, the timing of exposure to social programs may have additional effects due to the induced changes in their biology. In other words, providing transfers in early adolescence can have second-order effects through its impact on the age of menarche. For instance, results from a long-term assessment of a conditional cash transfer program in Nicaragua (which had nutrition, education, and reproductive health education components) corroborate this argument. Girls who joined the program between eleven and fourteen had a lower age at menarche, higher fertility, and lower labor force participation than those who joined the program earlier (Barham, Macours, and Maluccio 2018).

The main result of this paper—early menarche decreases school enrollment—also implies that the timing of menarche is not a valid instrument for the age at which a woman gets married if the outcome of interest is also affected by her education. This strategy was first proposed by Field and Ambrus (2008), who examine cohorts born between 1951 and 1970 in the Matlab region of Bangladesh. In their sample, 70% of the interviewed women had married within two years after their menarche. Therefore, menarche was the effective socio-biological constraint on marriage. Since then, the gap between the age at menarche and age at marriage has, if anything, increased over time. However, this strategy has been used to study the effect of early marriage on a variety of outcomes for younger cohorts of women. Using the methodology proposed by Conley, Hansen, and Rossi (2012), I show that the two-stage least squares estimate that use the timing of menarche to instrument for the timing of marriage is not stable. The estimated confidence intervals around the two-stage least squares estimates under plausible violations in the exclusion restriction assumption are wide and contain zero.

The existing evidence on the effect of menstruation in general and menarche in particular is far from conclusive. Oster and Thornton (2011) find that menstruation accounts for only 0.4 missed school days in a 180-day school year in Nepal. On the other hand, studies that use self-reported data and case studies argue that menstruation is a critical factor that keeps girls out of classrooms. For instance, in a study conducted in three states in India, over half of the interviewed adolescent girls reported that menstruation was a significant barrier to attending school (Sivakami et al. 2019). Similar conclusions are evident from the studies based in other contexts, including Northern Tanzania (Sommer 2010) and Western Kenya (Mason et al. 2013). The strong positive response in pubescent girls' enrollment to the construction of sex-separate toilets suggests that privacy and safety concerns during this age are critical factors in a girl's enrollment decision (Adukia 2017). However, none of these studies establish an explicit causal link between menarche and school enrollment, as done in the paper. In line with the evidence on the limited association between menstrual days and school attendance in Oster and Thornton (2011), I show that if enrolled at twelve, girls' menarche status does not affect the time they spend at school. Taken together, these results suggest that while menarche affects enrollment, it may have limited effect on school attendance.

Using the children's self-reports of their aspirations, I also show that reaching menarche before twelve increases a girl's desire to become a full-time parent or housewife by 80.4%.[4] While the construction of identity is critical as an outcome by itself, it can have consequences beyond the primary outcome studied in this paper. For instance, a woman's sense of self, not only determines her labor market outcomes (Bertrand 2011) but also has spillover effects on female labor force participation across generations (Olivetti, Patacchini, and Zenou 2020). In fact, early menarche also has effects that persist well into early adulthood. Consistent with the existing literature (Desai and Andrist 2010), I show that girls who reach menarche before twelve are 11% more likely to be married at twenty-two, about a year younger than their peers.

Clearly, the consequences of the onset of menstruation are more than just biological in India. The solution to the increased dropout rate after early menarche in this setting cannot be limited to providing menstrual hygiene management products, but will have to acknowledge the social and cultural context of these transitions.[5] I show that girls are more likely to drop out of school after menarche if they reside in communities with a lower average perception of safety among children or if they belong to castes with restrictive gender norms. By locating an increase in dropout rates at menarche, these results suggest that gender-specific social norms and concerns for safety are amplified after this one event. As such, policy responses to higher dropout rates among adolescent girls should address their safety concerns and challenge gender-based cultural norms.

At the same time, I find that the negative impact of menarche on the enrollment rate is attenuated in communities with higher expected wages for female-dominated professions. By and large, the findings of the paper are consistent with the view of Jayachandran (2015) that it is the interaction between cultural norms and local economic opportunities that drives female educational attainment.

The rest of the paper is organized as follows. Section 2 describes the socio-cultural norms surrounding menarche in India, and Section 3 presents the underlying conceptual framework. Section 4 describes the data followed by a discussion on the empirical strategy in Section 5. Section 6 presents the results describing the impact of early menarche on enrollment and discusses robustness checks and heterogeneity. Section 7 examines the relationship between age at menarche and other critical outcomes. Section 8 discusses the results and conclusion.

2 Context

Despite being a ubiquitous and regularly occurring experience for women, menstruation is surrounded by various myths and taboos. This is especially true in South Asia where these myths and taboos continue to define women's day-to-day activities. The origin of mythical narratives around menstruation can be traced back to *Rig Veda* which was composed between 1500 and 1200

4. This may be the direct effect of menarche or the effect of altered norms of interacting with her family and community after menarche.

5. Evidence exploring the effectiveness of menstrual hygiene management products has not been very promising so far. Section 8 discusses this literature and the explanation for what might be driving the limited impact.

BC (Chawla 1994). When Lord *Indra* slayed a high caste demon, *Vritra*, part of *Indra's* guilt was assumed by womankind. His guilt appears every month in the form of their menstrual flow, and menstruation is considered ritually impure.

The notion of menstruation's ritual impurity is most evident in taboos and cultural practices that confine the scope of women's daily lives. For instance, women are not supposed to enter the kitchen or the prayer room during their menses. Some dietary restrictions, such as the prohibition of sour food, are also followed during menstruation (Garg, Sharma, and Sahay 2001; Garg and Anand 2015). In some communities, young girls are instructed to stay away from water sources during their menses since they will pollute it (Mahon and Fernandes 2010).

Menarche marks a girl's transition into womanhood in South Asia. Now that she is a woman, her parents must plan her marriage. Delayed marriage signifies the failure of a girl's parents to discharge their duties (ICRW and Plan Asia 2013). According to the religious scriptures, by performing the act of *kanyadan*, or handing over his daughter to her husband, the bride's father earns good virtues. However, if the girl is perceived to be flamboyant or unrestrained, it would not be easy to arrange her marriage (Dube 1988). Parents have to safeguard her virginity until the day of her wedding. Now that she is a woman, she is more vulnerable to sexual harassment and is observed more carefully by the community members (Singh and Vennam 2016). Her mobility and the extent and the scope of her interactions are restricted (Seymour 1999).

To further unpack the relevance of the socio-cultural role of menarche in explaining the key results of this paper, the rest of this section discusses evidence from Vennam and Komanduri (2009), a report based on qualitative data collected from the cohort under consideration. Vennam and Komanduri (2009) use extensive qualitative data collected from children from every sampling site to discuss the critical transitions experienced by children between the ages of eight and fifteen.[6]

Families celebrate their daughters' menarche by hosting a meal for their relatives and friends within the community. Girls are kept at home for five to seven days after menarche. They are not allowed to touch any person during this period. Some families continue secluding their daughters during her menses period, even after menarche. After menarche, girls in Muslim communities may wear a *burqa*, and in Hindu communities, long skirts and *saris* instead of their usual clothes. They are no longer allowed to go out alone, even for short distances. Mothers explain that if their daughters do not adhere to these restrictions, the community might disapprove, and this will affect their marriage prospects. Importantly, parents are preoccupied with preparing their daughters for marriage and with approval from the local community.

Restrictions on mobility as a result of menarche are entwined with the schooling decision. One girl explains that now she wears a *burqa* while going to school to avoid unwanted attention. Girls report being *eve-teased*, a colloquial term for public sexual harassment, on their way to school. In sum, menarche characterizes tremendous changes in a girl's life that go well beyond the biological changes.

6. See Section 4 for a detailed discussion of the sampling strategy.

3 Conceptual Framework

Drawing on the discussion on the socio-cultural role of menarche in a girl's life in Section 2, this section describes a simple framework of how the onset of menstruation can affect her education and age at marriage. This framework has two goals. First, it describes a setup under which menarche affects schooling. Second, it derives testable implications for the heterogeneous impact of menarche on schooling. This framework has to be qualified in several respects. The choice problem is static and conceptualized in a unitary household setup.

Formally, I model parents' choice of length of schooling, and consequently, the age at dropout (e) and age at marriage (a). The age at menarche is a component of this decision-making process because menarche increases the cost of attending school for the girl. This formulation draws from the discussion in Section 2 and captures the safety and the social cost of having an unmarried daughter who she goes to school. For instance, an interview with an eighteen-year-old girl, which was conducted as part of the qualitative data collection exercise for the cohort studied in this paper, highlights how schooling experience modifies after menarche (Singh and Vennam 2016). The interviewee's parents worried about her safety and the risk of sexual harassment as she traveled to school. They pulled her out of school soon after menarche, and she was married at fourteen. Teenage girls also face greater restrictions on their mobility and their access to school for parents fear that socialization with boys brings them *bad name* and adversely affects their prospects in the marriage markets (Hardgrove et al. 2014).

Parents' utility maximization problem is as follows:

$$\max_{\{e,a\}} \quad \omega b(e) + v(a) - c(e, max\{a-m, 0\})$$

$$\text{subject to}$$

$$(1) \quad a \geq m$$

$$(2) \quad a \geq e$$

where $\omega b(e)$ is the direct benefit to the parents from their daughter's education and her time at her natal home before marriage, which is strictly increasing ($b_1 > 0$) and concave ($b_{11} < 0$). ω captures the future monetary returns the benefits from education. Next, $v(a)$ denotes the gain from delaying the marriage, which is increasing ($v_1 > 0$) up to some age, $\bar{a}$, and then starts decreasing; delaying marriage beyond some age would entail negative returns.[7] Further, $v(a)$ is concave: ($v_{11} < 0$). Education is costly, so is delaying marriage after attaining menarche (m): $c(e, max\{a-m, 0\})$. The cost function is increasing ($c_1 > 0$ and $c_2 > 0$) and convex ($c_{11} > 0$ and $c_{22} > 0$). One can think of $max\{a-m, 0\}$ as the *reputation*, or the *bad reputation*, that the girl may build once she crosses the threshold of womanhood and still lives in her natal home. As discussed above, traveling to school and socializing with boys (arguably at school) amplifies the reputation cost after menarche: $c_{12} > 0$.

Menarche is a lower bound on the age at marriage as described by the constraint $a \geq m$. In

7. Rao (1993) discusses that there is a pressure on women to get married within an acceptable age range, and it becomes increasingly difficult to find a match as they get older.

addition, a girl cannot continue her education after marriage; this feature of the choice problem is described by the second constraint, that is, $a \geq e$.

Consider the interior solution where $a > m$ and $a > e$.[8] [9] The relationship between years of education and age at menarche is:[10]

$$\frac{\partial e(m)}{\partial m} = \frac{-v_{11}c_{12}}{(\omega b_{11} - c_{11})(v_{11} - c_{12}) - c_{12}^2} > 0 \tag{3.1}$$

Prediction 1: Completed years of schooling is increasing in the age at menarche. Therefore, girls are more likely to drop out of school after menarche.

Now consider choice problems for girls who reach menarche before their peers. Since $c_{12} > 0$, safety concerns, and other associated costs are higher for the former group. The primary contribution of the paper is to estimate the effect of early menarche on school enrollment. In addition to these results, I discuss suggestive evidence on the effect of later menarche.

Prediction 2: The effect of the onset of menstruation on schooling, if any, is smaller for girls who reach menarche later than their peers.

This framework can also be used to derive implications for heterogeneity in the impact of early menarche on school enrollment along some policy-relevant dimensions. Consider two communities A and B such that the complementarity between education and delay in marriage beyond menarche is stronger in community A. In other words, the safety and social costs of attending school after menarche are higher in community A. Note that $\partial e(m)/\partial m|_A > \partial e(m)\partial m|_B$; the relationship between menarche and education is stronger in community A than in community B.

Prediction 3: The decrease in enrollment due to early menarche is more pronounced if the safety or reputation costs are more pertinent.

Next, consider the returns to education realized by parents. In this setup, ω captures the monetary returns to daughter's education. Consider ω_A and ω_B such that $\omega_A > \omega_B$; higher ω denotes a higher returns to education, including the returns realized in the labor market. Note that $\partial e(m;\omega_A)/\partial m < \partial e(m;\omega_B)/\partial m$, that is, the relationship between the age at menarche and educa-

8. Consider the case wherein the first constraint ($a \geq m$) is binding. In this case, as the age at menarche increases, the binding constraint on the age of marriage is tightened, increasing the age at marriage. If the girl is already out of school, her education is not affected. However, if she is currently in school, she will drop out since $a \geq e$. Indeed, this is the case discussed in Field and Ambrus (2008). However, for the state of Andhra Pradesh, the median length of time between when the girl reaches menarche and when she gets married is five years, suggesting that the constraint is not binding (see Section 4 for details of data used in the study.).

9. Consider when the second constraint ($a \geq e$) is binding. In this case, the optimal level of education is more than the age of marriage. Higher age at menarche will increase the age at marriage, and therefore education. Predictions on the effect of menarche that are described below will still hold.

10. Second-order conditions associated with the interior solution are: (1) $\omega b_{11} - c_{11} < 0$, (2) $v_{11} - c_{22} < 0$, and (3) $(\omega b_{11} - c_{11})(v_{11} - c_{22}) - (c_{12})^2 > 0$. These conditions hold under the assumptions of the framework described above.

tion is weaker when the returns to education are higher. Thus, the prediction for the heterogeneity in the impact of early menarche along returns to education is as follows:

Prediction 4: The decrease in enrollment due to early menarche is abated if the potential returns to education are higher.

Lastly, the relationship between age at marriage and age at menarche is:

$$\frac{\partial a(m)}{\partial m} = \frac{-c_{22}(\omega b_{11} - c_{11}) - c_{12}^2}{(\omega b_{11} - c_{11})(v_{11} - c_{12}) - c_{12}^2} > 0 \tag{3.2}$$

In this setup, the relationship between ages at marriage and menarche is not unambiguous. The condition that ensures that the age at marriage is increasing in the age at menarche ($-c_{22}(\omega b_{11} - c_{11}) > c_{12}^2$) is a stricter requirement on the curvature of the maximand than the second-order condition for an interior solution requires. In addition to the evidence on the effect of early menarche on schooling, the paper also discusses the evidence on the relationship between early menarche and marital outcomes.

To summarize, this framework shows that as long as a girl experiences substantial safety and reputation costs of attending school after menarche, earlier menarche decreases school enrollment. The effect of a later onset of menstruation, if any, will be smaller. Therefore, this framework provides an intuitive and parsimonious setup to interpret the main results of this paper.

4 Data

This paper uses data collected by the Young Lives longitudinal study that follows two cohorts of children in Andhra Pradesh over the course of their childhood, and focuses on the older cohort that was born in 1994-95.[11] Within each sampling site, households were screened to compile a list of eligible children (aged eight in 2001). From this sampling site-level list, a sample of 50 children born between January 1994 and June 1995 was randomly selected.

The Young Lives study tracks children in the same cohort over their childhood, adolescence, and early adulthood. Therefore, it is ideal for implementing the identification strategy used in this paper given that it exploits the variation in the timing of menarche.[12] The Young Lives study also collected detailed socioeconomic and demographic data from household, child and community surveys. In particular, primary care-givers were asked about consumption expenditure, economic shocks and their perceptions of children's health status. Children took standardized tests that

11. Andhra Pradesh was split into Andhra Pradesh and Telangana in 2014. Appendix A describes the administrative and the socio-cultural context of the state of Andhra Pradesh.

12. The other longitudinal surveys in India (such as the Rural Economic and Demographic Survey and the India Human Development Survey) do not collect information on the timing of the menarche of the school-going girls in the households, and I cannot use them to study the effect of early menarche on schooling.

measure their cognitive ability, and were also interviewed about important aspects of their experiences like schooling, time use, aspirations, social networks and their perceptions of the social milieu. Children were also asked about their marital outcomes in later survey rounds. Community surveys were conducted through focus-group discussions and collected information on topics like local wages and public services. The Young Lives study also collected data on children's physical development through on-site anthropometric measurements. Importantly, the second survey round, during which children were about twelve years old, asked girls if they had started menstruating yet, and, if so, at what age.

The first round of the study was conducted in 2002 when these children were eight years old. The next four rounds of data were collected in 2006, 2009, 2013, and 2016. Attrition was low: the first four rounds tracked 94.4% of the children, and all the five rounds tracked 92.4% of the children. Importantly for this paper, the rate of attrition did not differ across girls and boys, or across girls by their menarche status.

There are about 1,000 children in the cohort studied in the paper. Male and female are nearly equally represented in the sample: 49% of the children are boys, and 51% are girls. About one-fourth of the data are collected from urban areas, and the remaining from rural areas. Backward castes constitute about 47% of the sample, while scheduled castes are 21%, scheduled tribes, 11%, and Muslims, 6% of the sample. The remaining sample represents the upper caste households in the state. About 35% of the children are first-borns.

Although the primary variable of interest is the enrollment status of the children, I also use their test scores for supplementary analysis. Peabody Picture Vocabulary Tests (PPVT) and Mathematics tests were administered in the second and third rounds of the survey.[13] [14] For the analysis purpose, I normalize children's raw test scores to the distribution of scores in each survey round.

5 Empirical Specification

To identify the effect of early menarche on enrollment, I exploit the variation in the timing of menarche within a cohort. The information on the timing of menarche is taken from the second survey round when the typical cohort age was twelve. A little more than a quarter of girls had started menstruating by this time.[15] For the rest of the paper, these girls are referred as the "the early menarche group", and the remaining girls are referred as "the late menarche group".[16]

I use a difference-in-differences strategy that compares two groups of children across the first and

13. The PPVT test is designed to measure receptive vocabulary. In a PPVT test, an examiner shows a set of four images simultaneously and asks the children to select the image that best represents the word spoken by the examiner.

14. These tests were administered to all the children at their homes during the interview, and not only to those children who were in school at the time of the survey.

15. Recall data for age at menarche is missing for a small sub-sample of girls. Further, for some girls, the recall age of when the started menstruating is higher than the accurately measured current age. Therefore, girls are categorized into two groups.

16. I use the second round of the nationally-representative India Human Development Survey to confirmed that starting menses before twelve indeed denotes early menarche: only 25% of the interviewed women in Andhra Pradesh reached menarche by this age.

the second survey rounds, that is, when they were eight and twelve, respectively.[17] The specification for this comparison is:

$$y_{ict} = \beta_0 + \beta_1 EarlyMenarche_{ic} + \beta_2 Round2_{ict} + \beta_3 EarlyMenarche_{ic} Round2_{ict} + \gamma X_{ict} + \delta_c + \epsilon_{ict} \quad (5.1)$$

where y_{ict} is the enrollment status of child i in sampling site c in survey round t. $EarlyMenarche_{ic}$ is an indicator equal to one if child i reached menarche before twelve, and $Round2_{ict}$ is an indicator equal to one when we consider data from the second survey round. X_{ict} is a vector of child-specific characteristics, and δ_c is a set of sampling site dummies. ϵ_{ict} is a conditionally-mean-zero error term. Since a sampling site is also the primary sampling unit, standard errors are clustered at the sampling site-level. To estimate the impact of early menarche, I compare girls in the early menarche groups to girls in the late menarche group and β_3 is the difference-in-differences estimator.

Sampling-site fixed effects capture unobserved characteristics of the local region, including cultural practices and the quality of schooling. Sampling-site fixed effects also control for the administrative capacity of the local bureaucracy because the sampling-sites are the sub-district level administrative units. Finally, including sampling-site fixed effects alleviates the concerns of the effect of climate on age at menarche (Sohn 2016).

Causal identification based on difference-in-differences strategy requires that the timing of menarche is conditionally independent of those factors that also influence enrollment. If the timing of menarche solely depends on random genetic variation, the identification based on a difference-in-differences strategy is valid. Kaprio et al. (1995) show that the correlation between the ages at menarche is 0.75 for monozygotic (identical) twin pairs, but only 0.31 for dizygotic (fraternal) twin pairs. Using a seventy-four years-long panel study from the United States, Towne et al. (2005) show that almost half of the variation in the timing of menarche is due to genetic factors. However, the pre-eminence of genetic factors in explaining the variation in age at menarche is an established fact only for developed countries. In South Asia, skeletal height, weight, physical activity, and socioeconomic status are correlated with age at menarche (Bagga, Kulkarni, et al. 2000; Chowdhury et al. 2000; Dambhare, Wagh, and Dudhe 2012; Rah et al. 2009). Therefore, I include household size, mother's literacy status, wealth index, sex of the head of the household, caste, an indicator of whether the first language is the most spoken in the state (Telugu), number of older siblings, number of younger siblings, number of older brothers, number of older sisters and BMI and height at age eight as child-level controls. To remove any confounding effect of children's latent abilities, I also include their score in Raven's test at eight.

Table 1 summarizes these control variables. There are some important differences across girls in the early and late menarche groups. First, girls in the early menarche group are taller and have higher BMI than girls in the late menarche group and boys. This difference corroborates the evidence on the relationship between age at menarche and nutritional status. Note that better

17. There are three relevant groups of children: girls in the early menarche group, girls in the late menarche group and boys.

nutrition is typically associated with higher school enrollment (Nandi et al. 2015). Therefore, if anything, estimates from a difference-in-differences model will underestimate the true effect of early menarche. Second, in line with the son-biased fertility stopping rule documented in Jayachandran and Pande (2017), boys have fewer younger siblings and more older sisters. Girls in the late menarche group also have more older sisters than girls in the early menarche group.

To explore the heterogeneity in the impact of early menarche on enrollment, I use a triple-difference specification. Let the dimension of heterogeneity be H_{ic}. The specification for the comparison when girls in the early and the late menarche groups are compared across the first and the second survey rounds, therefore, is:

$$\begin{aligned} y_{ict} =& \beta_0 + \beta_1 EarlyMenarche_{ic} + \beta_2 Round2_{ict} + \beta_3 H_{ic} \\ &+ \beta_4 EarlyMenarche_{ic} H_{ic} + \beta_5 Round2_{ict} H_{ic} + \beta_6 EarlyMenarche_{ic} Round2_{ict} \\ &+ \beta_7 EarlyMenarche_{ic} Round2_{ict} H_{ic} + \gamma X_{ict} + \delta_c + \epsilon_{ict} \end{aligned} \quad (5.2)$$

where all variables are defined as in Equation 5.1, β_7 is the main parameter of interest (the triple-difference estimate), and β_1 through β_6 are the estimates of the double interaction and the linear terms. As before, standard errors are clustered at the sampling site-level.

The first survey round did not collect data on some important aspects of children's experiences. For these outcomes (such as safety perceptions, time use, etc.), I estimate a contemporaneous regression. For instance, when time use of girls in the early and the late menarche groups are compared at age twelve, the specification is:

$$y_{ic} = \beta_0 + \beta_1 EarlyMenarche_{ic} + \gamma X_{ic} + \delta_c + \epsilon_{ic} \quad (5.3)$$

and β_1 is the intent-to-treat estimator of the impact of menarche if the timing of menarche is as good as random once observables are controlled for. In case this assumption does not hold, β_1 is still a useful description that complements the understanding of menarche's effects on children's experiences. As before, X_{ic} is a vector of child specific characteristics, δ_c is a set of sampling site dummies and ϵ_{ic} is a conditionally-mean-zero error term.

6 Results on Enrollment

6.1 Descriptive Evidence

Figure 1 plots school enrollment by gender. A gap in school enrollment across genders emerges when children are twelve and widens as they grow older. At nineteen, the gender gap in the enrollment rate is 25%. Figure 2 plots these trends again, but children are now classified into three groups: boys, girls in the early menarche group, and girls in the late menarche group. There is no difference in enrollment rates across the three groups when children are about eight. Girls in the early menarche group and boys were still equally likely to be in school at twelve. The enrollment rate for girls in the early menarche group, however, falls visibly between eight and twelve; while the enrollment rate for girls in the late menarche group shows a sharp drop between twelve and fifteen.[18]

The modal reason for dropping out of school between the first and the second survey rounds for boys is “truancy” followed by “needed for work at home”. For girls in the late menarche group, the top reasons for dropping out of school are “needed for work at home” and “truancy”. For girls in the early menarche group, the modal reason is “other reasons”, that is, not one of the reasons listed among the extensive prespecified options in the survey. The second most often stated reason is “needed for work at home”. Clearly, girls in the early menarche group drop out of school due to reasons that are very different from other children in their cohort.

There are two key patterns in these data that suggest that early menarche induces girls to drop out of school. First, there is a substantial drop in school enrollment rate for girls after they start menstruating. This sharp decrease is not experienced by other children in their cohort. Second, after girls start menstruating, the reasons that they cite for dropping out of school are different from the reasons cited by other children in their cohort.

6.2 Key Results

Results from various specifications of Equation 5.1 are shown in Table 2. Columns (1)-(3) compare girls in the early menarche group to girls in the late menarche group, while columns (4)-(6) compare girls in the early menarche group to boys.

Column (1) considers Equation 5.1 while not including any controls and fixed effects: the estimated impact is -12.3% point. This estimate changes little when child specific controls and sampling site fixed effects are included. The most conservative specification yields an estimate of -12.2% point, still a substantial drop in enrollment rate (Table 2). This drop represents 13.4% of the average enrollment rate for girls in the late menarche group at twelve. The estimated effect of the onset of menstruation when girls in the early menarche group is compared to boys is -10.7% point in the specification without any controls and fixed effects. Progressively adding controls and fixed effects does not alter the estimates. Estimated β_3 in the most conservative specification is -11.1% point, that is, 12.2% of the enrollment rate for boys at twelve.

If the gender gap in enrollment at twelve is due to early menarche, enrollment rates for boys and girls in the late menarche group should be similar until the latter group start menstruating. Indeed, in a comparison of the enrollment rates between boys and girls in the late menarche group, the difference-in-differences estimate is small (1.2% point), and not significantly different from zero (Table 3).

In order to understand that the impact of early menarche on school enrollment is economically significant, it is worth noting that the magnitude of the estimated effect of early menarche is comparable to the effect of most education related policy reforms and schemes. Adukia (2017) found that latrine construction led to an 8% increase in enrollment among children enrolled in

18. This pattern is also replicated with exact age data (Figure B.1, Appendix B).

middle school, and a 12% increase in enrollment among children enrolled in primary school. The effect was moderately larger for girls, and the difference between the effect for girls was larger for middle school than for primary school; these differences, however, were not statistically significant in all specifications. Kaur (2017) shows that providing free school lunch to primary school students increased the net primary enrollment rate by 7.3%, while the increase for girls was higher at 10.5%. Azam and Saing (2017) find that a large scale school construction program in low literacy districts increased the likelihood of ever attending primary school by 3.4% among girls.[19] Providing bicycles to school going girls aged fourteen or fifteen increased their school enrollment by 32% (Muralidharan and Prakash 2017).

6.3 Parallel Trends

The validity of the difference-in-differences identification strategy hinges on the assumption of parallel trends: the average change in the control group reflects the counterfactual change in the treatment group if there were no treatment. In this context, the assumption of parallel trends implies that the trends in enrollment rate for any two groups will be indistinguishable if their menstruation status does not change across two consecutive survey rounds. While the nature of the Young Lives study does not allow a test of pre-treatment parallel trends, note that three groups of children are equally likely and universally to be in school at eight (Table 1). Further, the children across the three groups start school at the same age, at around five years (Table 1). The parallel trends assumption holds when I compare enrollment rates for girls in the early and late menarche groups across the third and fourth survey rounds (Table 4). Trends in enrollment are also parallel for comparisons between girls in the early menarche group and boys across (1) the second and third survey rounds and (2) the third and fourth survey rounds.

Kahn-Lang and Lang (2019) discuss the need to carefully consider assumptions inherent in difference-in-differences analysis by (1) addressing the differences in the initial levels, (2) exploring if the parallel trends assumption requires a justification of the chosen functional form, and (3) acknowledging that comparing pre-treatment trends does not establish parallel trends during the treatment period.

I addressed concerns (1) and (3) in the discussion on parallel trends earlier in this subsection. Average enrollment rates for children across groups do not differ when the children are eight, that is, in the pre-treatment stage. The comparisons across the third and the fourth survey rounds demonstrate that post-treatment parallel trends hold. Thus, it can be claimed with some credibility that parallel trends would have been the counterfactual trends.

If the initial distributions of the outcome (enrollment rate in this case) are different across comparison groups, it is difficult to justify the parallel trends assumption using pretesting with linear trends. This is the motivation for the Kahn-Lang and Lang (2019)'s second concern. They agree that it is hard to make a case for one model over the other in the absence of a theory, and

19. This increase represents a 2.8% point increase in school enrollment for girls over a base enrollment rate of 82% in non-program districts

recommend testing parallel trends for different functional form assumptions. In addition to the tests using a linear probability model described above, tests using logit functional form also indicate that the counterfactual trends in enrollment are parallel across comparison groups (Table 5).

6.4 Robustness Checks

Robustness to Alternative Hypotheses Spuriously Linking Menarche to School Enrollment

As discussed in Section 4, age at menarche is associated with local environmental conditions, especially in developing countries. Menarche is correlated with better nutritional status in South Asia. However, better nutritional status is associated with higher educational attainments in the same population (Nandi et al. 2015). Since the girls who reach menarche before twelve are healthier, it is unlikely that they dropped out of school on account of poor nutritional status. Further, including children's BMI and height when they were eight years old mitigates the concerns of bias induced by the correlation between better nutritional status and earlier menarche. Not only do girls in the early menarche group have better nutritional status at age eight, their height and BMI also increase at a faster rate between ages eight and twelve (Table 6).

The fact that girls in the early menarche group had better nutritional status at age eight, however, raises another potential concern. If equity considerations dictate parents' decision to allocate resources to children within the household, they may reduce complementary investments in their healthier daughter, which may affect her schooling. While I do not observe child-specific investments between ages eight and twelve, non-food expenditure and dietary inputs do not differ across girls in the early and the late menarche groups at twelve (Table 7).

One may also be worried that parents pull their children out of school once they reach a certain height to have them undertake wage labor. Since girls in the early menarche group were taller than their counterparts at age eight, the decrease in enrollment at age twelve might be such a phenomenon. Three pieces of evidence point that this is unlikely to be the case. First, enrollment levels never converge across the two groups of girls (Figure 2). Second, while girls in the early menarche group do spend more time on paid work, this effect entirely driven by those dropped out of school (Table 8). Finally, for both the younger siblings and the elder siblings, their time use patterns do not differ by the timing of their sisters' menarche (Table B.1; Table B.2, Appendix B), suggesting that it is unlikely that girls were pulled out of school to help the family with housework.[20]

Since high levels of cortisol due to stress leads to earlier menarche, unexpected negative economic shocks that induce mental stress can also lead to early menarche (Karapanou and Papadimitriou 2010). Any association between earlier menarche and school enrollment might be capturing the effect of external stressors on enrollment. To test for the robustness to this possible source of bias, I use health emergencies and deaths in the family as proxies for external stressors.[21] The key result

20. The Young Lives study collects information on the time use of all the children between the ages of five and eighteen in the household for a typical day in the last week. An important caveat is that the time use data for siblings is missing for 491 siblings out of the total 1,847 siblings of the Young Lives children.

21. Health emergencies and deaths in the family are the most significant drivers of a family's descent into poverty in developing countries (Krishna 2010).

is robust to the inclusion of negative health shocks and deaths in the children's families, or an economic shock due to droughts (Table B.3; Table B.4, Appendix B). Moreover, the incidence of health emergencies, deaths, and economic shock due to droughts are similar for girls in the early and late menarche groups (Table 1). Another test to confirm that this alternative hypothesis is not driving the key result of this paper is to check if co-resident siblings of girls in the early menarche group drop out of school at a higher rate than siblings of other surveyed children. Table B.5 in Appendix B presents these results and confirms that siblings of girls in the early menarche group are not more likely to drop out of school between the first and the second survey rounds.

It is also plausible that the enrollment patterns reflect these children's latent ability and performance in school. Low ability and poor performing children drop out of school at an earlier age. This explanation is unlikely to be driving the drop in enrollment among girls in the early menarche group as their test scores outperform girls in the late menarche group in tests designed to measure their mathematical ability (Table 9).

Accuracy of the Early Menarche Indicator

Another potential threat is due to the concern that the main *treatment* variable, that is, the indicator that the child belongs to the early treatment group is not measured accurately. This concern is especially valid in India, where menstruation is a taboo. There are three things to note here. First, the Young Lives study extensively piloted the survey and amended those questions that fieldworkers thought were culturally inappropriate (Young Lives 2017). Second the correlates of the early menarche indicator line up with what we know from the literature in biology: girls in the early menarche group are taller than their counterparts in the late menarche group at age eight (Table 1). Their height exhibits a growth spurt between the ages of eight and twelve and slows down after that. The height for the girls in the late menarche group, on the other hand, shows a spurt between the ages of twelve and fifteen (Figure 3). These patterns, jointly, demonstrate a pre-menarche growth spurt and a deceleration in growth after menarche and are in line with the existing evidence on changes during puberty (Villamor and Jansen 2016). Finally, the proportion of girls who reach menarche before age twelve in these data is very similar to the number when I use a larger data set that is representative of the state of Andhra Pradesh.[22] I show that about 25% of the interviewed women reached menarche before age twelve, whereas 26% of the sampled girls are in the early menarche group (Figure 4).

Individual Fixed Effects

To assuage any other concerns regarding endogeneity of early menarche, I compute difference-in-differences estimates while including individual fixed-effects. Including individual fixed effects is a useful robustness check as it de facto includes the fixed effects for the household, the community, and anything about the child that does not change over time. The estimates of β_3 are remarkably stable when we consider the comparison across the first two survey rounds (Table B.6, Appendix

22. I use the second round of the India Human Development Survey that was conducted in 2011-12 and is representative at the state level (Desai and Vanneman 2015).

B).[23]

Randomization Inference

Randomization inference is a non-parametric method of hypothesis testing that does not rely on asymptotic properties of the error terms. Randomization inference is relevant for the paper as the sample was drawn from twenty sampling sites where the error terms are likely to be correlated within sampling sites. I cluster the standard errors at the level of a sampling site in the empirical estimates described above. However, the robustness of these estimates may still be of concern as the sample has a relatively small number of sampling sites (twenty). Randomization inference does not rely on a theoretical distribution of the error term. Instead, it compares the test statistic to the distribution derived under different possible allocations of the treatment status, thereby addressing the likely correlation between error terms within a sampling site. As the number of draws increases, this procedure performs better. Following Young (2019), I use 2000 replications.

Figure B.2 in Appendix B shows the distribution of β_3, the placebo impact of early menarche on school enrollment when girls in the early and the late menarche groups are compared. Reassuringly, the placebo estimates are centered around zero, and only three of the 2000 replications estimates reach the estimated result of 12.2% point. The distribution of the associated t-statistics should approximately be normal and the t-statistic associated with the difference-in-differences estimate should be in tails of the resulting distribution. Indeed, the t-statistics associated with the difference-in-differences estimate fall within the tails of the distribution (Figure B.3, Appendix B). These results are consistent with the finding of a p-value equal to 0.001 for the estimated β_3. In a comparison between girls in the early menarche group and boys, none of the 2,000 replications reach the estimated effect of 11.1% point, confirming that results described in Section 6.2 are not spurious (Figure B.4; Figure B.5, Appendix B).

6.5 The Impact of Attaining Menarche Between Twelve and Fifteen

The third survey round of the Young Lives study did not collect information about menarche status. To explore if reaching menarche after twelve affects school enrollment, I compare girls in the late menarche group with boys across the second and third survey rounds. The underlying assumption for this analysis is that most of the girls in the late menarche group start menstruating between the second and third survey rounds. In the the second round of the India Human Development Survey (IHDS), which was conducted in 2011-12, 99% of the interviewed women in Andhra Pradesh had reached menarche before they turned fifteen, and 90% of the entire sample of interviewed women had reached menarche before age fifteen. Since age at menarche among Indian women is steadily declining (Pathak, Tripathi, and Subramanian 2014), the assumption that most girls in the late menarche group reached menarche by the time they were fifteen, patterns from the IHDS support the assumption that girls in the late menarche group reach menarche before fifteen.

23. This is not the preferred specification because some of the analysis in the paper compares outcomes across two groups in a given period. Child fixed effects cannot be used in that setup. Same empirical framework is used to discuss all the results in the paper.

When girls in the late menarche group are compared with boys, the impact estimator is -4.7% point, that is, 5.8% of the average enrollment rate for boys at fifteen (Table 10). When I compare girls in the early and the late menarche group, β_3 is not significantly different from zero.

In line with Prediction 2 in Section 3, the estimated effect of late menarche on school enrollment is smaller than the estimated effect of an early menarche. Further, enrollment rates do not converge, and girls in the early menarche group have persistently lower enrollment rates than other girls in their cohort.

6.6 Heterogeneity in the Impact on Enrollment

Overall, the results show that the decrease in enrollment caused by menarche drives the gender gap in enrollment that emerges between the ages of eight and twelve. This subsection tests Predictions 3 and 4 and explores how the strength of the relationship between early menarche and enrollment varies with safety and reputation costs, and with local labor market conditions.[24] This section focuses on the comparison between girls in the early and the late menarche groups across the first and the second survey rounds.

Reputation and Safety Costs

In South Asia, the notion of "family honor" is inextricably linked to the behavior of women in the family. As discussed in Section 2, the fear of being labeled as a girl of bad character who may bring dishonor upon the family intensifies as a girl transitions into womanhood after menarche. Further, these concerns are more pertinent among upper castes than lower castes (Eswaran, Ramaswami, and Wadhwa 2013). In Andhra Pradesh, the evidence from the second round of the IHDS shows that the female mobility is least restricted among women from the scheduled castes.[25] Further, women from the scheduled castes are less like to practice *purdah*, a social practice wherein women are physically segregated from men or veil their faces (Table 11). The decrease in school enrollment due to menarche is entirely driven by the girls from non-scheduled caste families, suggesting that the cultural norms that restrict female mobility could be critical drivers of higher dropout rate after menarche (Table 12; p-value: 0.12).

The discussion in Section 2 suggests that the onset of menstruation might alter a girl's perceived needs and experiences of safety in her community. Indeed, a comparison of girls' responses to "whether they feel safe in the community" at twelve shows that girls in the early menarche group were 6.8% point less likely to strongly agree that they feel safe in their community. The corresponding average for girls in the late menarche group was 81%. An immediate implication of this change is that the effect of early menarche on enrollment could be pronounced in relatively unsafe communities. I use the community-level average of boys' safety perceptions to encapsulate children's concerns and experiences of the safety of their neighborhood.[26] Indeed, the decrease in enrollment is larger in communities with a lower average perceived safety (Table 12). Although not statistically significant

24. Results on other dimensions of heterogeneity are described in Appendix C.

25. Scheduled caste is the official designation for the historically disadvantaged caste group in India.

26. I only use data for boys because girls' perceptions of safety are affected by menarche.

at the conventional levels of significance, the triple-difference estimate is substantial at -12.4% point (p-value: 0.18).

Local Labor Market Conditions

If potential gains from education are substantial, the negative impact of menarche of enrollment may be mitigated. Since parents base their expectations of their daughter's earning potential on the female wage profile in their neighborhood (Chamarbagwala 2008; Jensen 2012), average salaries for female-dominated professions, teaching and nursing, are used as proxies for local wages for educated women. The Young Lives study collected this information through community-level focus-group discussions. Triple-difference estimates are computed for the indicator that the average wages for these professions are less than the median for the whole sample. Results based on both these proxies suggest that the negative impact of menarche on enrollment is mitigated in communities with higher wages for women (Table 13). The triple-difference estimate when teachers' wages are considered is -8.9% (p-value: 0.16) and when nurses' wages are considered is -10.7% (p-value: 0.06).

7 Other Related Outcomes

The evidence on parallel trends and a battery of robustness checks described in Section 6 confirm that early menarche reduces the school enrollment rate. In this section, I explore if early menarche affects outcomes other than school enrollment.

7.1 Test Scores

Girls in the early menarche group outperform girls in the late menarche group when children are twelve, although differences in PPVT test scores are not statistically significant (Table 9). While boys outperform girls in the late menarche group, there is no difference in the average test scores for girls in the early menarche group and boys at age twelve. However, if I consider the trends in these test scores across the second and the third survey (when the children were twelve and fifteen, respectively) girls in the early menarche group lose the advantage they had at age twelve. The difference-in-differences estimator for a comparison between girls in the early menarche group and boys is negative. Moreover, the gender gaps in the test scores for both PPVT and Mathematics tests widen (Table 14). Taken together with the fact that girls in the early menarche group are healthier at age eight, these patterns provide additional evidence that specific social contexts can attenuate the gains from the *virtuous cycle* of human development, and drive the gender gaps in not only school enrollment but also learning.

7.2 Time Use

When a girl drops out of school, the most immediate change is through the reallocation of the time she used to spend on education-related activities. The decrease in time spent at school and on studying at home translates into an additional twenty-nine minutes on paid work and seventeen minutes on household chores. Changes in time use patterns of other activities are not statistically significant. In the sub-sample of those girls still enrolled in school, time spent at school and on studying does not vary with the girl's menarche status, suggesting that early menarche does not affect the intensive margin of schooling (Table 8). This is in line with Oster and Thornton (2011) that menstruation explains only 0.4 missed days in a 180 day school-year.

7.3 Aspirations

Besides changes in time use, by restricting mobility, menarche also redefines the scope of a girl's social interactions (Seymour, 1999). Do changes in parents' and society's interactions with girls in the early menarche group affect how they see themselves? Early menarche increases a girl's desire to become a full-time parent/housewife by 8.6% point. Notably, at twelve, only 10.7% of girls in the late menarche group report that they want to become full-time parents or housewives when they grow up (Table 15). However, if I restrict the analysis to the sample of girls who are still in school, the effect size is smaller (-1.7% point) and statistically insignificant. While girls in the early and the late menarche groups are equally likely to aspire to receive post-high school education, girls in the early menarche group are less likely to believe that they would reach their desired level of education. It is an open question whether it is the change in aspirations after menarche that induces a girl to drop out or her withdrawal from formal education that alters how she sees herself.

7.4 Marital Outcomes

At nineteen, 45% of the girls in the early menarche group are married, while 31% of girls in the late menarche group are married. The marriage rate increases for both groups between nineteen and twenty-two. At twenty-two, 65% of the girls in the early menarche group are married, while 54% of girls in the late menarche group are married (Figure 5).[27] Once child-specific characteristics and sampling site fixed effects are included, girls in the early menarche group are 11% point (p-value: 0.12) more likely to be married at nineteen and also 11% point (p-value: 0.16) more likely to be married at twenty-two.[28] Rates at which girls in the early and the late menarche groups get married between nineteen and twenty-two are statistically indistinguishable (Table 16).

Figure 6 shows the distribution of age at marriage for girls in the early and late menarche groups. The distribution of age at marriage for girls in the late menarche group is shifted to the right of the distribution for girls in the early menarche group: girls in the late menarche group marry later, on average, and at all quantiles of age at marriage. For the restricted sample of ever-married girls,

27. The Young Lives study started collecting data on marital outcomes only from the fourth survey round, when the typical cohort age was nineteen.

28. There are two key differences in the empirical strategy adopted in this subsection. First, the child-specific controls correspond to her natal household instead of the household she currently resides in because marital outcomes would have depended on the characteristics of her natal home since, after marriage, she must live with her husband's family. Second, instead of an asset index, land-holding of the natal household when children were fifteen is used to capture their natal household's wealth. Since the turnover rates in land markets are low (Basu Roy and Ghosh Dastidar 2018), land-holdings capture the natal household's wealth at the time of the wedding.

the average age at which girls in the early menarche group get married is 0.81 years lower than the average age at which girls in the late menarche group get married (Table 16

The Young Lives study also collects information on other characteristics of a girl's marriage. 42.2% of the girls who were married by twenty-two met their husband on the day of the wedding, and 39.4% married a man chosen by their parents. These features of marriages do not differ across girls in the early and late menarche groups. Although not statistically significant, the average expenditure on the wedding by the bride's family is lower for girls in the early menarche group by about 41,000 rupees (in 2013 prices). There is no difference in the average education of husbands, the likelihood that girls' families took a loan for their wedding, and the likelihood that they are related to their husbands by blood across two groups (Table 17). It is worth emphasizing that only 65% of the girls in the sample are yet married, and these comparisons do not describe what the differences in these characteristics in marriages would look like once all girls are married.

7.4.1 Menarche as an Instrument for Early Marriage

In South Asia, girls typically don't marry until they reach menarche because menarche marks the biological readiness for childbearing, and therefore marriage (Sheela and Audinarayana 2003). Indeed girls in the early menarche group marry younger. The difference in the likelihood of being married at nineteen (14% point) is in the order of the difference in the enrollment rates (10% point) at nineteen. Besides, once all girls are past menarche, trends in their enrollment rate are indistinguishable (Table 4). Menarche could increase the demand for marriage because now the girl is out of school, or because of the socio-biological relevance of menarche, or because of a combination of the two. Recent evidence point towards the former explanation: using preference elicitation experiments in India, Adams and Andrew (2019) show that girls' marriage prospects start deteriorating as soon as they leave school, and early dropout accelerates early marriage. However, parents have a strong preference for delaying their daughter's marriage until eighteen, and they value her education until high school.

There is an important strand of literature that explores the impact of the timing of marriage on various outcomes of well-being for married women, and their offspring. Often, to consistently estimate the effect of the age of marriage, the variation in the timing of menarche is used to isolate the exogenous variation in the age at marriage. This strategy was first proposed in Field and Ambrus (2008), who use it to study the effect of women's age at marriage on their educational attainment. Their sample includes women born between 1951-1970 in the Matlab region of Bangladesh. At the time, the legal minimum age at marriage was fifteen, and often not binding. For their sample, over 70% of marriages took place within two years of menarche. In such a setting, the reaching menarche would have been the only constraint to enter the marriage market.

However, margins of the education and the marriage decision that are affected by menarche have changed with time. The age at marriage consistently increased between 1950-54 and 1965-70 in South Asia (Jensen and Thornton 2003). The average age at marriage was just over sixteen for cohorts born in 1965-1970, but over eighteen for cohorts born between 1986-1996.[29] However, the age at menarche declined by three months between the cohorts who were born around 1955-64 and the cohorts that were born around 1985-90 (Pathak, Tripathi, and Subramanian 2014). Therefore, the length of time between the ages at menarche and marriage increased over this period.

To formally examine the validity of the timing of menarche as an instrument for the age at marriage, consider the following setup: the outcome of interest is years of schooling (Y), the endogenous independent variable is an indicator that the girl did not marry until she was eighteen (X), and the indicator for early menarche (menarche before twelve) is the instrument for the endogenous variable (Z).[30] Consider the probability limit of the two-stage least squares estimator:

$$plim\beta^{2sls} = \underbrace{\beta}_{population parameter} + \underbrace{\frac{Cov(Z,U)}{Cov(Z,X)}}_{bias}$$

where U is the error term. If the incidence of early menarche does not independently affect educational attainment, the two-stage least squares estimator is consistent. However, if the relationship between early menarche and educational attainment is negative (that is, $Cov(Z,U) < 0$, as demonstrated in Section 6) and $Cov(Z,X) < 0$, then the two-stage least squares estimator overestimates the true parameter. However, if the size of the positive bias is not large, the estimated effect is still qualitatively useful.

Conley, Hansen, and Rossi (2012) propose a method to undertake inference and generate bounds around the two-stage least squares estimate when the instrument is only plausibly exogenous, that is when the exclusion restriction is suspect. This method requires the researcher to have a prior belief about the extent of the violation of the exclusion restriction. To implement this method, I take the extent of the violation of the exclusion restriction from the Young Lives study, a dataset for which I have established that menarche affects schooling. I use this prior to estimate bounds around the two-stage least squares estimator for the described application using the much larger IHDS dataset, which covers the cohort of women born between 1957 and 1991.[31]

Using early menarche as an instrument for being unmarried until eighteen, the estimated effect of the latter on completed education is 4.5 years (Table D.2, Appendix). Following Conley, Hansen, and Rossi (2012), I allow plausible violations in the exclusion restriction, and assume that the potential direct effect of early menarche is either drawn from (1) known support with its upper bound set at zero, or (2) from a known distribution. Section D in the Appendix describes the associated methodology in detail. The more conservative bounds around the two-stage least squares estimate

29. I calculated the average age at marriage for cohorts born in 1986-1996 using the IHDS.

30. These variables are different from the ones considered in Field and Ambrus (2008) to fully utilize the information available in the Young Lives study. Instead of using age at marriage, an indicator that the girl is unmarried until eighteen is used as the endogenous variable as all girls were not married by the fifth survey round.

31. Since the Young Lives study has a relatively small sample size. Consequently, the first stage is weak, it is not appropriate to implement the procedure described in Conley, Hansen, and Rossi (2012). Moreover, the IHDS has been used for many applications where age at menarche is used as an instrument for age at marriage to study the effect of the latter on outcomes that may be affected by a woman's education (for instance, Sekhri and Debnath 2014; Chari et al. 2017; Dhamija and Roychowdhury 2018).

are [-9.3, 4.5] and still contain zero.

The extent of violation of the exclusion restriction estimated from the Young Lives data is more than three times the maximum allowed violation that would ensure that the bounds do not contain zero (Figure D.2; Figure D.3, Appendix). Therefore, the two-stage least squares estimate is not stable to plausible violations in the exclusion restriction, which are still smaller than the estimated violation using the Young Lives study.

To summarize, two results discussed in this paper suggest that the timing of menarche may not satisfy the exclusion restriction when used as an instrument for age at marriage in the Indian setting. First, this paper shows that reaching menarche before age twelve leads to a 13.4% decline in school enrollment rate. Second, the two-stage least squares estimates where early menarche is used to instrument for the timing of marriage are not stable when strict exogeneity of the instrument is replaced by plausible exogeneity.

8 Conclusion

In this paper, I examine the effect of menarche on schooling in India, where menarche is associated with substantial changes in a girl's life. Using a dataset that follows the same cohort of children during childhood, adolescence, and early adulthood, I show that reaching menarche before age twelve causes a 13.4% decrease in school enrollment. This result is robust to a variety of alternate specifications and explanations spuriously linking early menarche with enrollment.

There is a broad consensus that improving childhood nutrition is a long-term investment that catalyzes the *virtuous cycle* of human and economic development. Among girls, however, improvements in health and nutrition have a close biological association with the timing of their pubertal development: a better nutritional environment leads to earlier menarche. In socio-cultural contexts where menarche initiates a series of transitions in a young girl's life, the link between better nutrition in childhood and higher economic returns in later life may be undercut.

Healthier and taller girls start menstruating at a younger age. At age twelve, girls in the early menarche group perform as well as boys in tests designed to measure their mathematical ability and their vocabulary. However, boys outperform girls in the early menarche group when the children are fifteen. Therefore, in settings where menstruation is associated with ritual impurity and job prospects for women are limited, the onset of menstruation can trigger transitions that can minimize the long-term gains through the link between better nutrition and schooling. This result also contributes to our understanding of the complex process of human capital formation by showing adolescence is another sensitive period for girls. In this light, the long-term effects of a transfer program (with a nutritional component) will also depend on when it was provided (Barham, Macours, and Maluccio 2018).

Given the sizable impact of menarche on education, improving access to menses management technology may seem the most appropriate policy response. While menstrual hygiene management is a growing sector within international development, the existing evidence on these interventions is not very promising. Despite high adoption rates, providing menstrual cups to school girls in Nepal does not improve school attendance (Oster and Thornton 2011, 2012). Stopford (2011) finds similar patterns in rural Kenya, where commercial sanitary napkins usage does not translate into higher school attendance. A more recent evaluation set in Kenya finds a small positive effect of providing sanitary napkins, but not of giving menstrual cups (Benshaul-Tolonen et al. 2019). An evaluation of two supplementary interventions in Ghana—distribution of sanitary pads and puberty-related education—finds limited additional impact of sanitary pads on school attendance (Montgomery et al. 2012). It is worth noting that these evaluations study the efficacy of menstrual hygiene management technology among the selected group of girls who are still in school after menarche. If menarche affects school enrollment, as I show in this paper for India, an absence of any impact of menstrual hygiene management on attendance is not a confirmation that menstruation does not affect schooling.

Another relevant policy response could focus on conditional-cash transfers that incentivize schooling. These programs have been effective in improving school enrollment across different contexts, including Mexico (Schultz 2004), Brazil (Glewwe and Kassouf 2012), and Nicaragua (Millán et al. 2020). The conditional-cash transfer programs in India tend to be long-term as opposed to the interventions listed here where parents receive regular payments conditional on meeting the transfer requirements. The programs in India typically give the parents a savings bond when they have a daughter, which is redeemable if she is unmarried at eighteen with additional bonuses for education. The existing studies that evaluate these programs primarily focus on fertility outcomes (Anukriti 2018; Balakrishnan 2018; Sinha and Yoong 2009. Sinha and Yoong (2009) find positive but weak education effects of such a program in the state of Haryana. Exploring the impact of these programs on adolescent girls' schooling and understanding their role in mitigating the negative effect of early menarche is a promising area for future research.

Early menarche has implications for a girl's life beyond those implied by the change in her physiology. A better understanding of these changes has to be at the center of any policy response. The first step in that direction is an analysis of the interaction between menarche and the decision around school enrollment. Results from exploring the heterogeneity in the impacts show that higher dropout rates due to early menarche are located in relatively unsafe communities. Indeed, interventions that address the safety concerns of young girls, such as providing bicycles (Muralidharan and Prakash 2017) and separate-sex toilets (Adukia 2017), have been successful in encouraging higher female enrollment. Furthermore, in urban India, the fear of street harassment affects women's college choices (Borker 2017).

The impact of menarche is more pronounced among girls from non-scheduled caste households, where restrictive gender norms are more pertinent. In such conservative communities, while families personally reject restrictive norms, they incorrectly estimate that other people accept them and continue to follow them (Bursztyn, González, and Yanagizawa-Drott 2018). These perceived cultural norms are not immutable. Bursztyn, González, and Yanagizawa-Drott (2018) show that correcting false beliefs does alter behavior. Persuasion and discussion can also change gender attitudes: a social campaign that engaged children in discussions about gender equality shifted gender attitudes to be more progressive in India (Dhar, Jain, and Jayachandran 2018), and an intervention that fosters grit closes the gender gap in competitiveness among school children in Turkey (Alan and Ertac 2019). The appropriate policy responses to higher dropout rates among adolescent girls should address girls' safety concerns in particular and gender-based cultural norms in general.

The factors that counter the effect of early menarche on enrollment are equally important. Girls are less likely to drop out of school after menarche if they live in neighborhoods with higher wages for female-dominated professions. Socio-cultural norms restrict girls' mobility and society's progress towards gender parity, but increased economic opportunities act as an important counterweight to the impact of these norms on girls' education.

References

Adams, Abi, and Alison Andrew. 2019. *Preferences and beliefs in the marriage market for young brides.* Technical report. IFS Working Papers.

Adukia, Anjali. 2017. "Sanitation and education." *American Economic Journal: Applied Economics* 9 (2): 23–59.

Alan, Sule, and Seda Ertac. 2019. "Mitigating the gender gap in the willingness to compete: Evidence from a randomized field experiment." *Journal of the European Economic Association* 17 (4): 1147–1185.

Anukriti, S. 2018. "Financial incentives and the fertility-sex ratio trade-off." *American Economic Journal: Applied Economics* 10 (2): 27–57.

ASER. 2018. *'Beyond Basics' 2017.* Technical report.

Attanasio, Orazio P. 2015. "The determinants of human capital formation during the early years of life: Theory, measurement, and policies." *Journal of the European Economic Association* 13 (6): 949–997.

Azam, Mehtabul, and Chan Hang Saing. 2017. "Assessing the impact of district primary education program in India." *Review of Development Economics* 21 (4): 1113–1131.

Bagga, Amrita, S Kulkarni, et al. 2000. "Age at menarche and secular trend in Maharashtrian (Indian) girls." *Acta Biologica Szegediensis* 44 (1-4): 53–57.

Balakrishnan, Uttara. 2018. "The little push: Role of incentives in determining household behaviour in India." PhD diss.

Barham, Tania, Karen Macours, and John A Maluccio. 2018. "Experimental evidence of exposure to a conditional cash transfer during early teenage years: Young women's fertility and labor market outcomes."

Basu Roy, Sharanya, and Sayantan Ghosh Dastidar. 2018. "Why do men rape? Understanding the determinants of rapes in India." *Third World Quarterly:* 1–23.

Benshaul-Tolonen, Anja, Garazi Zulaika, Elizabeth Nyothach, Clifford Oduor, Linda Mason, David Obor, Kelly T Alexander, Kayla F Laserson, and Penelope A Phillips-Howard. 2019. "Pupil absenteeism, measurement, and menstruation: Evidence from western Kenya." *CDEP-CGEG WP,* no. 74.

Bertrand, Marianne. 2011. "New perspectives on gender." In *Handbook of labor economics,* 4:1543–1590. Elsevier.

Borker, Girija. 2017. "Safety first: Perceived risk of street harassment and educational choices of women." *Job Market Paper, Department of Economics, Brown University.*

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott. 2018. *Misperceived social norms: Female labor force participation in Saudi Arabia.* Technical report. National Bureau of Economic Research.

Chamarbagwala, Rubiana. 2008. "Regional returns to education, child labour and schooling in India." *The Journal of Development Studies* 44 (2): 233–257.

Chari, AV, Rachel Heath, Annemie Maertens, and Freeha Fatima. 2017. "The causal effect of maternal age at marriage on child wellbeing: Evidence from India." *Journal of Development Economics* 127:42–55.

Chawla, Janet. 1994. "Mythic origins of menstrual taboo in Rig Veda." *Economic and Political Weekly:* 2817–2827.

Chowdhury, Sameena, AK Mohammed Shahabuddin, Anil J Seal, Kurshid K Talukder, Quamrul Hassan, Rowshan A Begum, Q Rahman, A Tomkins, A Costello, and MQK Talukder. 2000. "Nutritional status and age at menarche in a rural area of Bangladesh." *Annals of Human Biology* 27 (3): 249–256.

Conley, Timothy G, Christian B Hansen, and Peter E Rossi. 2012. "Plausibly exogenous." *Review of Economics and Statistics* 94 (1): 260–272.

Cunha, Flavio, and James Heckman. 2007. "The technology of skill formation." *American Economic Review* 97 (2): 31–47.

Currie, Janet, and Douglas Almond. 2011. "Human capital development before age five." In *Handbook of labor economics,* 4:1315–1486. Elsevier.

Dambhare, Dharampal G, Sanjay V Wagh, and Jayesh Y Dudhe. 2012. "Age at menarche and menstrual cycle pattern among school adolescent girls in Central India." *Global Journal of Health Science* 4 (1): 105.

Desai, Sonalde, and Lester Andrist. 2010. "Gender scripts and age at marriage in India." *Demography* 47 (3): 667–687.

Desai, Sonalde, and Reeve Vanneman. 2015. "India human development survey-ii (ihds-ii), 2011-12." *ICPSR36151-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]* 31.

Dhamija, Gaurav, and Punarjit Roychowdhury. 2018. "The impact of women's age at marriage on own and spousal labor market outcomes in India: causation or selection?" *Available at SSRN 3173636.*

Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2018. *Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India.* Technical report. National Bureau of Economic Research.

Dharmalingam, A. 1994. "The implications of menarche and wedding ceremonies for the status of women in a South Indian village." *Indian Anthropologist* 24 (1): 31–43.

Dube, Leela. 1988. "On the construction of gender: Hindu girls in patrilineal India." *Economic and Political Weekly:* WS11–WS19.

Eswaran, Mukesh, Bharat Ramaswami, and Wilima Wadhwa. 2013. "Status, caste, and the time allocation of women in rural India." *Economic Development and Cultural Change* 61 (2): 311–333.

Field, Erica, and Attila Ambrus. 2008. "Early marriage, age of menarche, and female schooling attainment in Bangladesh." *Journal of Political Economy* 116 (5): 881–930.

Garg, Suneela, and Tanu Anand. 2015. "Menstruation related myths in India: strategies for combating it." *Journal of Family Medicine and Primary Care* 4 (2): 184.

Garg, Suneela, Nandini Sharma, and Ragini Sahay. 2001. "Socio-cultural aspects of menstruation in an urban slum in Delhi, India." *Reproductive Health Matters* 9 (17): 16–25.

Glewwe, Paul, and Ana Lucia Kassouf. 2012. "The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil." *Journal of Development Economics* 97 (2): 505–517.

Hardgrove, Abby, Jo Boyden, Paul Dornan, and Kirrily Pells. 2014. "Youth vulnerabilities in life course transitions."

ICRW and Plan Asia. 2013. *Plan Asia Child Marriage Initiative: Summary of Research in Bangladesh, India and Nepal.* Technical report. Plan Asia Regional Office.

Jacoby, Hanan G, and Ghazala Mansuri. 2011. *Crossing boundaries: gender, caste and schooling in rural Pakistan.* The World Bank.

Jayachandran, Seema. 2015. "The roots of gender inequality in developing countries." *Annual Review of Economics* 7 (1): 63–88.

Jayachandran, Seema, and Rohini Pande. 2017. "Why are Indian children so short? The role of birth order and son preference." *American Economic Review* 107 (9): 2600–2629.

Jensen, Robert. 2012. "Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India." *The Quarterly Journal of Economics* 127 (2): 753–792.

Jensen, Robert, and Rebecca Thornton. 2003. "Early female marriage in the developing world." *Gender & Development* 11 (2): 9–19.

Kahn-Lang, Ariella, and Kevin Lang. 2019. "The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications." *Journal of Business & Economic Statistics:* 1–14.

Kaprio, Jaakko, Arja Rimpelä, Torsten Winter, Richard J Viken, Matti Rimpelä, and Richard J Rose. 1995. "Common genetic influences on BMI and age at menarche." *Human Biology:* 739–753.

Karapanou, Olga, and Anastasios Papadimitriou. 2010. "Determinants of menarche." *Reproductive Biology and Endocrinology* 8 (1): 115.

Kaur, Randeep. 2017. "Essays in human capital development." PhD diss.

Krishna, Anirudh. 2010. "Who became poor, who escaped poverty, and why? Developing and using a retrospective methodology in five countries." *Journal of Policy Analysis and Management* 29 (2): 351–372.

Mahon, Thérèse, and Maria Fernandes. 2010. "Menstrual hygiene in South Asia: a neglected issue for WASH (water, sanitation and hygiene) programmes." *Gender & Development* 18 (1): 99–113.

Mason, Linda, Elizabeth Nyothach, Kelly Alexander, Frank O Odhiambo, Alie Eleveld, John Vulule, Richard Rheingans, Kayla F Laserson, Aisha Mohammed, and Penelope A Phillips-Howard. 2013. "'We keep it secret so no one should know'–A qualitative study to explore young schoolgirls attitudes and experiences with menstruation in rural Western Kenya." *PloS one* 8 (11).

Millán, Teresa Molina, Karen Macours, John A Maluccio, and Luis Tejerina. 2020. "Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program." *Journal of Development Economics* 143:102385.

Montgomery, Paul, Caitlin R Ryus, Catherine S Dolan, Sue Dopson, and Linda M Scott. 2012. "Sanitary pad interventions for girls' education in Ghana: a pilot study." *PloS one* 7 (10): e48274.

Muralidharan, Karthik, and Nishith Prakash. 2017. "Cycling to school: increasing secondary school enrollment for girls in India." *American Economic Journal: Applied Economics* 9 (3): 321–50.

Nandi, Arindam, Ashvin Ashok, Sanjay Kinra, Jere R Behrman, and Ramanan Laxminarayan. 2015. "Early Childhood Nutrition Is Positively Associated with Adolescent Educational Outcomes: Evidence from the Andhra Pradesh Child and Parents Study (APCAPS)–3." *The Journal of Nutrition* 146 (4): 806–813.

Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou. 2020. "Mothers, peers, and gender-role identity." *Journal of the European Economic Association* 18 (1): 266–301.

Oster, Emily, and Rebecca Thornton. 2012. "Determinants of technology adoption: Peer effects in menstrual cup take-up." *Journal of the European Economic Association* 10 (6): 1263–1293.

———. 2011. "Menstruation, sanitary products, and school attendance: Evidence from a randomized evaluation." *American Economic Journal: Applied Economics* 3 (1): 91–100.

Pathak, Praveen Kumar, Niharika Tripathi, and SV Subramanian. 2014. "Secular trends in menarcheal age in India-evidence from the Indian human development survey." *PLoS One* 9 (11): e111027.

Rah, Jee H, Abu Ahmed Shamim, Ummeh T Arju, Alain B Labrique, Mahbubur Rashid, and Parul Christian. 2009. "Age of onset, nutritional determinants, and seasonal variations in menarche in rural Bangladesh." *Journal of Health, Population, and Nutrition* 27 (6): 802.

Rao, Vijayendra. 1993. "Dowry 'inflation'in rural India: A statistical investigation." *Population Studies* 47 (2): 283–293.

Schultz, T Paul. 2004. "School subsidies for the poor: evaluating the Mexican Progresa poverty program." *Journal of Development Economics* 74 (1): 199–250.

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

Seymour, Susan C. 1999. *Women, family, and child care in India: A world in transition.* Cambridge University Press.

Sheela, J, and N Audinarayana. 2003. "Mate selection and female age at marriage: A micro level investigation in Tamil Nadu, India." *Journal of Comparative Family Studies* 34 (4): 497–508.

Singh, Renu, and Uma Vennam. 2016. "Factors shaping trajectories to child and early marriage: Evidence from Young Lives in India."

Sinha, Nistha, and Joanne Yoong. 2009. *Long-term financial incentives and investment in daughters: Evidence from conditional cash transfers in North India.* The World Bank.

Sivakami, Muthusamy, Anna Maria van Eijk, Harshad Thakur, Narendra Kakade, Chetan Patil, Sharayu Shinde, Nikita Surani, Ashley Bauman, Garazi Zulaika, Yusuf Kabir, et al. 2019. "Effect of menstruation on girls and their schooling, and facilitators of menstrual hygiene management in schools: surveys in government schools in three states in India, 2015." *Journal of Global Health* 9 (1).

Sohn, Kitae. 2016. "The influence of climate on age at menarche: Augmented with the influence of ancestry." *HOMO-Journal of Comparative Human Biology* 67 (4): 328–336.

Sommer, Marni. 2010. "Where the education system and women's bodies collide: The social and health impact of girls' experiences of menstruation and schooling in Tanzania." *Journal of Adolescence* 33 (4): 521–529.

Stopford, Amy Louise. 2011. "The Effect of Commercial Sanitary Pad Use on School Attendance and Health of Adolescents in Western Kenya." PhD diss.

Sundaram, Aparna, and Reeve Vanneman. 2008. "Gender differentials in literacy in India: The intriguing relationship with women's labor force participation." *World Development* 36 (1): 128–143.

Towne, Bradford, Stefan A Czerwinski, Ellen W Demerath, John Blangero, Alex F Roche, and Roger M Siervogel. 2005. "Heritability of age at menarche in girls from the Fels Longitudinal Study." *American Journal of Physical Anthropology* 128 (1): 210–219.

Vennam, Uma, and Anuradha Komanduri. 2009. "Young Lives qualitative research: round 1–India."

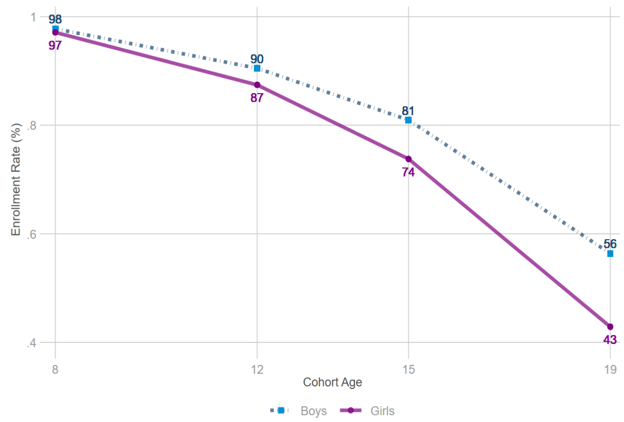
Villamor, Eduardo, and Erica C Jansen. 2016. "Nutritional determinants of the timing of puberty." *Annual Review of Public Health* 37:33–46.

Young, Alwyn. 2019. "Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results." *The Quarterly Journal of Economics* 134 (2): 557–598.

Young Lives. 2017. *A Guide to Young Lives Research.* Technical report. Young Lives.

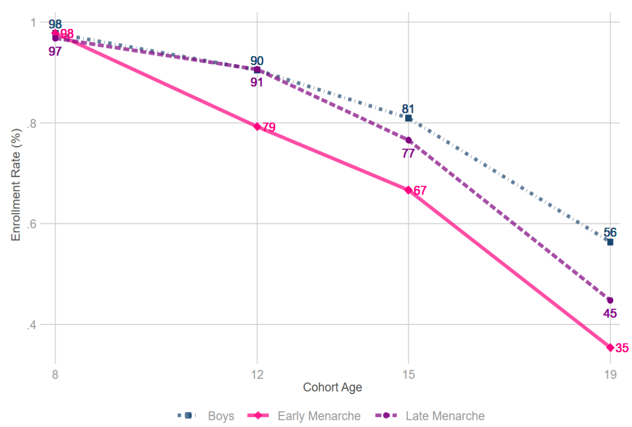
Figures

Figure 1. Enrollment Rate by Gender



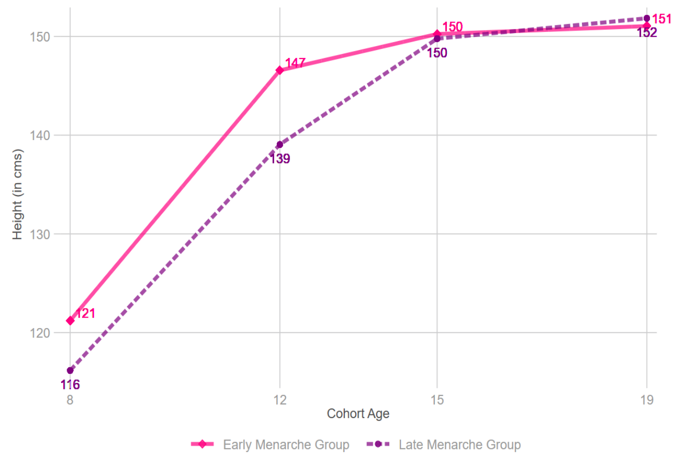
Note: Data are taken from the Young Lives study. The solid (blue) line represents the trends in enrollment rate for boys and the dashed (purple) line represents the trends in enrollment rate for girls.

Figure 2. Enrollment Rate by Gender and Menarche Status



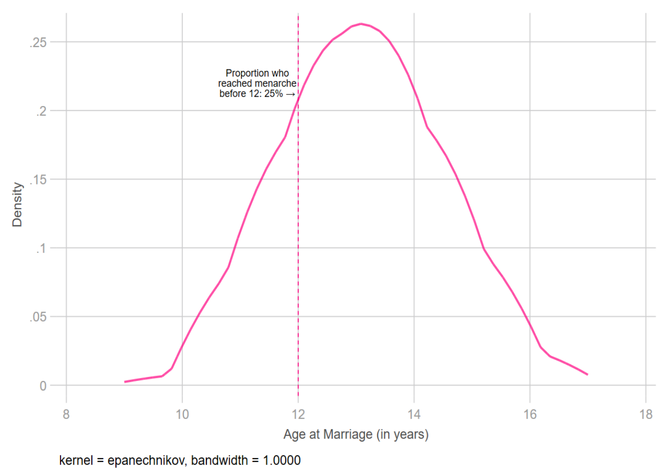
Note: Data are taken from the Young Lives study. The dash-dotted (blue) line represents the trends in enrollment rate for boys, the dashed (purple) line represents the trends in enrollment rate for girls in the late menarche group and the solid (pink) line represents the trends in enrollment rate for girls in early menarche group.

Figure 3. Height by Menarche Status



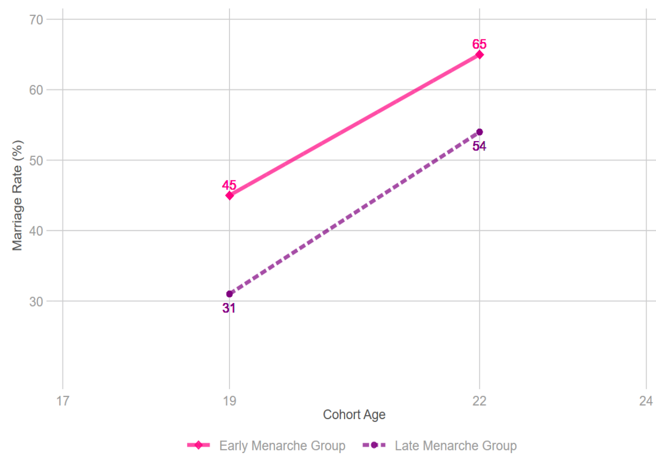
Note: Data are taken from the Young Lives study. The solid (pink) line represents the trends in height (in centimeters) for girls in the early menarche group and the dashed (purple) line represents the trends in height (in centimeters) for girls in the late menarche group.

Figure 4. Age at Menarche in Andhra Pradesh



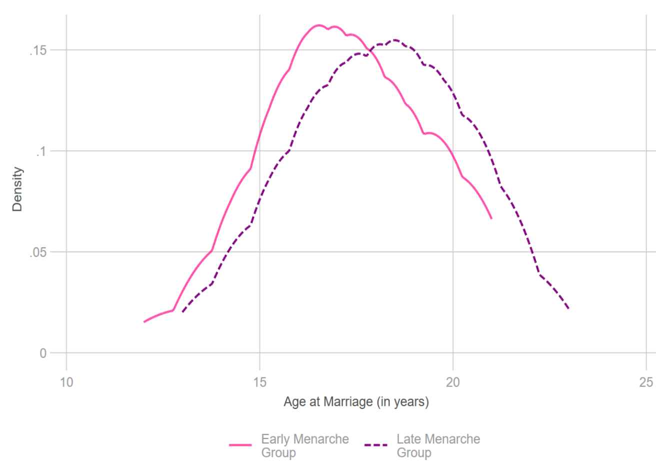
Note: Data are taken from the IHDS. The solid (pink) line represents the distribution of the age at menarche for the state of Andhra Pradesh.

Figure 5. Marriage Rate by Menarche Status



Note: Data are taken from the Young Lives study. The solid (pink) line represents the trends in marriage rate for girls in the early menarche group and the dashed (purple) line represents the trends in marriage rate for girls in the late menarche group. Both the lines describe trends the ages of nineteen and twenty-two.

Figure 6. Age at Marriage by Menarche Status



Note: Data are taken from the Young Lives study. The solid (pink) line represents the distribution of age at marriage for girls in the early menarche group and the dashed (purple) line represents the trends in age at marriage for girls in the late menarche group.

Tables

Table 1. Sample Characteristics in Round 1

	Early Menarche	Late Menarche	Boys	Difference	
	(1)	(2)	(3)	(1) - (2)	(1) - (3)
Household size	5.32	5.62	5.56	0.30	-0.24
Wealth index	0.43	0.40	0.41	0.03	0.02
Access to drinking water	0.80	0.82	0.84	-0.02	-0.04
Electricity connection	0.84	0.82	0.82	0.02	0.03
Area of land owned (in ha.)	1.24	0.68	0.89	0.56	0.35
Literate mother	0.27	0.21	0.21	0.06	0.06*
Age of the household head	39.84	40.51	40.00	-0.67	-0.15
Schedule Caste	0.22	0.195	0.22	-0.67	0.00
Experience of health shocks	0.27	0.30	0.34	-0.03	-0.07
Experience of drought	0.23	0.28	0.30	-0.05	-0.07*
Speaks Telugu	0.86	0.85	0.83	0.01	0.03
Number of older siblings	0.91	1.07	1.16	-0.16	-0.25**
Number of younger siblings	0.69	0.80	0.67	-0.11	0.02
Number of older brothers	0.66	0.61	0.72	0.05	-0.06
Number of older sisters	0.42	0.67	0.72	-0.24**	-0.30**
Raven's test score at age 8	23.37	22.83	22.97	0.54	0.40
BMI at age 8 (kg/m^2)	14.21	13.74	13.99	0.48***	0.22
Height at age 8 (cm)	121.21	116.16	118.56	5.05***	2.65***
Weight at age 8 (kg)	20.96	18.57	19.70	2.39***	1.26***
Age in first year at school (in years)	4.98	5.05	5.04	-0.07	-0.06
Enrollment rate at age eight	97.78	96.79	98.14	0.99	-0.36

Note: This table reports simple differences in means across the three groups of children.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 2. Enrolment Across Rounds 1 and 2 for Girls in the Early Menarche Group

	vs Late menarche Girls			vs Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
Early menarche	0.010	0.018	0.026	-0.004	-0.003	-0.005
	(0.011)	(0.019)	(0.022)	(0.014)	(0.014)	(0.014)
Round 2	-0.062**	-0.035	-0.034	-0.078***	-0.043**	-0.048**
	(0.022)	(0.031)	(0.030)	(0.014)	(0.017)	(0.018)
Early menarche X Round 2	-0.123**	-0.121***	-0.122***	-0.107**	-0.111**	-0.111**
	(0.032)	(0.031)	(0.031)	(0.038)	(0.037)	(0.037)
Observations	1017	1013	1013	1240	1234	1234
R^2	0.0494	0.106	0.162	0.056	0.100	0.128
Control average	0.906	0.906	0.906	0.903	0.903	0.903
Sampling site FE	No	No	Yes	No	No	Yes
Controls	No	Yes	Yes	No	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1) for comparisons that span the first and the second survey rounds. Controls include household size, caste, sex of the household head, wealth index, mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older sisters, number of older brothers, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 3. Enrolment Across Rounds 1 and 2 for Girls in the Late Menarche Group

	vs Boys		
	(1)	(2)	(3)
Late menarche	-0.014	-0.014	-0.020
	(0.009)	(0.010)	(0.012)
Round 2	-0.078***	-0.061**	-0.063**
	(0.014)	(0.017)	(0.018)
Late menarche X Round 2	0.017	0.012	0.012
	(0.023)	(0.024)	(0.024)
Observations	1717	1707	1707
R^2	0.0228	0.0611	0.0857
Control average	0.903	0.903	0.903
Sampling site FE	No	No	Yes
Controls	No	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1) for comparisons that span the first and the second survey rounds. Controls include household size, caste, sex of the household head, wealth index, mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older sisters, number of older brothers, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 4. Tests for Parallel Trends: OLS

	Girls in the Early Menarche Group vs Boys		Girls in the Early vs Late Menarche Groups
	(1)	(2)	(3)
Early menarche X Round 3	-0.037		
	(0.039)		
Early menarche X Round 4		-0.051	0.017
		(0.050)	(0.042)
Control average	0.808	0.561	0.448
Sampling site FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1) for comparisons that span the second and the third survey rounds (column 2), and the third and the fourth survey rounds (columns 3 and 4). Controls include household size, caste, sex of the household head, wealth index, mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older sisters, number of older brothers, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 5. Tests for Parallel Trends: Logit

	Girls in the Early Menarche Group vs Boys		Girls in the Early vs Late Menarche Groups
	(1)	(2)	(3)
Early menarche X Round 3	0.121		
	(0.219)		
Early menarche X Round 4		-0.111	0.040
		(0.276)	(0.236)
Control average	0.808	0.561	0.448
Sampling site FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1) for comparisons that span the second and the third survey rounds (column 2), and the third and the fourth survey rounds (columns 3 and 4). Controls include household size, caste, sex of the household head, wealth index, mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older sisters, number of older brothers, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 6. Girls in the Early and the Late Menarche Group- BMI, Height and Caretaker's Perception of Child's Health

	BMI	Height (in cms.)	Caretaker's Perception of Child's Health Relative to Peers
	(1)	(2)	(3)
Early menarche	0.406**	4.477***	-0.063
	(0.131)	(0.543)	(0.042)
Round 2	1.628***	22.483***	-0.032
	(0.134)	(1.187)	(0.082)
Early menarche X Round 2	1.612***	2.427**	0.153**
	(0.248)	(0.928)	(0.072)
Observations	1003	1007	1014
R^2	0.403	0.667	0.102
Control average	15.56	139.1	0.425
Sampling site FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1). Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older sisters, number of older brothers, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 7. Other Investments in Girls in the Early Menarche Group in Round 2

	Food Intake (last seven days)						
	Exp. (in rupees)	Cereal	Roots	Legumes	Milk	Eggs	Meat
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Early menarche	969.268	0.029	-0.037	-0.037	0.068*	0.007	0.037
	(1785.812)	(0.035)	(0.046)	(0.054)	(0.038)	(0.033)	(0.031)
Observations	501	507	507	507	507	486	478
R^2	0.128	0.255	0.208	0.224	0.0845	0.117	0.136
Control average	7559.8	0.251	0.380	0.647	0.171	0.0922	0.0627
Sampling site FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results from a single-difference specification (see Equation 5.3) using data from the second survey round. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older sisters, number of older brothers, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 8. Time Use Patterns, Proportion of Time in the Day

	(1) Work	(2) Work (if enrolled)	(3) Household Chores	(4) Care	(5) Sleep	(6) Play	(7) School	(8) Study	(9) School (if enrolled)
Early menarche	0.019**	0.002	0.012	0.001	0.006	0.000	-0.028*	-0.014**	-0.002
	(0.009)	(0.002)	(0.008)	(0.003)	(0.006)	(0.011)	(0.015)	(0.005)	(0.009)
Observations	483	432	483	483	483	483	496	483	432
R^2	0.125	0.110	0.202	0.159	0.342	0.192	0.257	0.298	0.440
Control average	0.0111	0	0.0488	0.0114	0.396	0.166	0.269	0.0844	0.0844
Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results from a single-difference specification (see Equation 5.3) for data from the second survey round when children were twelve. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 9. Normalized Test Scores in Round 2

	PPVT		Mathematics	
	(1)	(2)	(3)	(4)
	vs Late Menarche Group	vs Boys	vs Late Menarche Group	vs Boys
Early menarche	0.072	-0.022	0.121*	0.020
	(0.068)	(0.070)	(0.072)	(0.087)
Observations	497	601	498	610
R^2	0.396	0.387	0.449	0.410
Control average	-0.0983	0.0585	-0.0607	0.0470
Sampling site FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Late menarche		-0.148**		-0.117
		(0.051)		(0.069)
Observations		832		842
R^2		0.364		0.378
Control average		0.06		0.04
Girls		-0.130**		-0.079
		(0.046)		(0.068)
Observations		965		975
R^2		0.371		0.390
Control average		0.06		0.05
Sampling site FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Note: This table reports the results from a single-difference specification (see Equation 5.3). Controls include child's enrollment status, household size, caste, sex of the household head, wealth index, mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older sisters, number of older brothers, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 10. Enrolment Across Rounds 2 and 3 for Girls in the Late Menarche Group

	vs Early menarche Girls			vs Boys		
	(1)	(2)	(3)	(4)	(5)	(6)
Late menarche	0.114**	0.088**	0.073*	0.003	-0.009	-0.021
	(0.032)	(0.032)	(0.039)	(0.027)	(0.028)	(0.032)
Round 3	-0.126**	-0.137**	-0.140**	-0.095***	-0.108***	-0.110***
	(0.039)	(0.042)	(0.043)	(0.016)	(0.019)	(0.019)
Late menarche X Round 3	-0.014	-0.016	-0.014	-0.045*	-0.047**	-0.047**
	(0.043)	(0.046)	(0.046)	(0.022)	(0.022)	(0.022)
Observations	1003	995	995	1700	1681	1681
R^2	0.0445	0.154	0.252	0.0271	0.102	0.151
Control average	0.667	0.667	0.667	0.808	0.808	0.808
Sampling site FE	No	No	Yes	No	No	Yes
Controls	No	Yes	Yes	No	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1) for comparisons that span the second and the third survey round. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 11. Markers of Gender Performane, by Caste

	(1) Practice purdah	(2) Can go to: Health center	(3) Can go to: Grocery shop	(4) Can go to: Friend's house
Scheduled caste	-0.046*	0.039	0.051**	0.038
	(0.025)	(0.035)	(0.022)	(0.036)
Observations	1774	1774	1774	1774
R^2	0.288	0.216	0.209	0.229
Control average	0.145	0.588	0.802	0.651
Sampling site FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Note: This table reports the results on the association between *markers of gender performance* and caste for the state of Andhra Pradesh using the second round of the nationally representative IHDS Survey. Controls include an indicator for land ownership, logged per capita consumption expenditure, age, age squared, an indicator for whether the woman married before eighteen, number of household members and number of married women between the ages of fifteen and forty-nine. Primary Sampling-site fixed effects are included. Robust standard errors clustered at the level of the primary sampling unit are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 12. Impact of Early Menarche of Enrollment, by Caste and Local Safety Environment

	(1) Scheduled Castes	(2) Average Perception of Safety
Early menarche X Round 2	-0.014	-0.065
	(0.063)	(0.051)
Early Menarche X Round 2 X Non-SC	-0.136†	
	(0.085)	
Early Menarche X Round 2 X Lower Perceived Safety		-0.124‡
		(0.090)
Observations	1013	1013
R^2	0.167	0.168
Control average	0.906	0.906
Sampling site FE	Yes	Yes
Controls	Yes	Yes

Note: This table reports the results from a triple-difference specification (see Equation 5.2) for comparisons that span the first and the second survey rounds. Column (1) reports results when the difference-in-differences specification is interacted with the caste of the family and Column (2) reports results when the difference-in-differences specification is interacted with an indicator that the average community level perceived safety among boys is less than the median. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.
† p-value: 0.12; ‡ p-value: 0.18.

Table 13. Impact of Early Menarche of Enrollment, by Average Local Wages for Female Dominated Professions

	(1) Teachers' Wages	(2) Nurses' Wages
Early menarche X Round 2	-0.079***	-0.054*
	(0.020)	(0.026)
Early Menarche X Round 2 X Lower Teacher Wages	-0.089	
	(0.062)	
Early Menarche X Round 2 X Lower Nurse Wages		-0.107*
		(0.055)
Observations	1013	1013
R^2	0.168	0.177
Control average	0.906	0.906
Sampling site FE	Yes	Yes
Controls	Yes	Yes

Note: This table reports the results from a triple-difference specification (see Equation 5.2) for comparisons that span the first and the second survey rounds. Column (1) reports results when the difference-in-differences specification is interacted with with an indicator that the typical wages for nurses in the community is less than that for the median community and Column (2) reports results when the difference-in-differences specification is interacted with an indicator that the typical wages for nurses in the community is less than that for the median community. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.
† p-value: 0.16.

Table 14. Normalized Test Scores of Girls in the Early Menarche Group versus Other Children Across Rounds 2 and 3

	PPVT			Mathematics		
	(1)	(2)	(3)	(4)	(5)	(6)
	vs Late Menarche Group	vs Boys		vs Late Menarche Group	vs Boys	
Early menarche X Round 3	-0.033	-0.304**		0.021	-0.229*	
	(0.086)	(0.088)		(0.105)	(0.133)	
Girls X Round 3			-0.284**			-0.256**
			(0.079)			(0.098)
Observations	934	1133	1821	987	1209	1933
R^2	0.327	0.344	0.334	0.381	0.402	0.386
Control average	-0.253	0.230	0.230	-0.220	0.208	0.208
Sampling site FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1). Controls include enrollment status of the child, household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 15. Aspirations and Beliefs

	(1) Aspiration: Fulltime parent or housewife	(2) Aspiration: Fulltime parent or housewife (if enrolled)	(3) Aspiration: Attend college	(4) Aspirtation likely to be fulfilled
Early menarche	-0.009	0.011	-0.043	-0.064
	(0.026)	(0.020)	(0.045)	(0.038)
Round 2	0.060**	0.020		
	(0.025)	(0.020)		
Early menarche X Round 2	0.088**	-0.017		
	(0.039)	(0.032)		
Observations	1014	935	507	507
R^2	0.161	0.0793	0.115	0.230
Control average	0.107	0.107	0.738	0.861
Sampling site FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Note: Column (1) reports the results from a difference-in-differences specification (see Equation 5.1) for comparisons that span the first and the second survey rounds. Columns (2) and (3) reports the results from a single-difference specification (see Equation 5.3). Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 16. Marital Timing

	(1) Ever Married at 19	(2) Ever Married at 22	(3) Age	(4) Gap Between Ages at Dropout and Marriage	(5) Ever Married
Early menarche	0.114	0.112	-0.813**	-0.795*	0.128*
	(0.072)	(0.076)	(0.347)	(0.383)	(0.065)
Round 5					0.201***
					(0.020)
Early menarche X Round 5					-0.029
					(0.042)
Observations	495	495	272	264	953
R^2	0.203	0.218	0.246	0.272	0.251
Control average	0.325	0.539	18.18	2.461	0.461
Sampling site FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes

Note: Columns (1)-(4) of this table reports the results from a single-difference specification (see Equation 5.3) and column (5) reports results from a difference-in-differences specification (see Equation 5.2). Controls include girl's natal home's household size, caste, sex of the household head, landholding in round 3, mom's literacy status, age of the head of the household, indicator of whether the girl speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 17. Other Features of Marriage

	(1) Met on the Day of the Wedding	(2) Parents Chose the Spouse	(3) Value on Gifts Given (in rupees)	(4) Spouse's Education (in years)	(5) Family took a loan for wedding	(6) Related to husband by blood
Early menarche	-0.065	0.023	-0.411	-0.686	0.076	0.087
	(0.082)	(0.075)	(1.095)	(0.525)	(0.059)	(0.073)
Observations	272	272	267	231	271	272
R^2	0.201	0.272	0.221	0.211	0.231	0.149
Control average	0.422	0.395	3.258	3.258	0.741	0.232
Sampling site FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

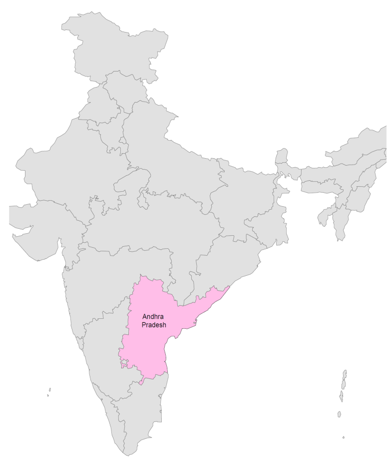
Note: This table reports the results from a single-difference specification (see Equation 5.3). Controls include girl's natal home's household size, caste, sex of the household head, landholding in round 3, mom's literacy status, age of the head of the household, indicator of whether the girl speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Appendices (not for publication)

A Context: State of Andhra Pradesh and the Sampling Strategy

Andhra Pradesh (including the state of Telangana) is the fifth-largest state in India. In 2002, when the Young Lives study started in Andhra Pradesh, only 27% of the population was living in urban areas. While agriculture was important and contributed to 30% of the state domestic product, economic activity was increasingly shifting away from agriculture. In 2004-05, the rural poverty rate was 11%, while the urban poverty rate was 28%. Andhra Pradesh was close to all India averages on various measures of human development. In 2004-05, the Andhra Pradesh's infant mortality rate was 53 as opposed to 57 for the whole country. The average life expectancy in Andhra Pradesh was 63.9, while it was 65.4 in India. Adult literacy and enrollment rates were 50.9 and 87.6 in Andhra Pradesh, respectively, and 61.8 and 82.1 in India (Centre for Economic and Social Studies, 2007).

Figure A.1. Andhra Pradesh



Andhra Pradesh has three historically distinct socio-cultural regions, Coastal Andhra, Rayalseema, and Telangana (Figure A.2). These divisions are divided into twenty-three districts (Figure A.3). Each district composed of three to five divisions, and each division has ten-fifteen *mandals*. A *mandal* is the lowest administrative tier of the government of Andhra Pradesh.

The sample for the Young Lives study was drawn such that the three regions the region were well represented. In addition to the metropolitan city of Hyderabad, two districts from each region, one poor and one non-poor were drawn. Next, from each district *mandals* were chosen to be representative of key development indicators. These chosen mandals, along with three urban slums chosen from Hyderabad, were the primary sampling units (sampling sites). From a list of villages that would have had at least fifty eight-year-olds and at least a hundred one-year-olds, sample villages were randomly chosen from the selected sampling sites.

Households covered by the Young Lives study are wealthier than the average household in Andhra Pradesh. A comparison with the households surveyed by the Demographic and Health

Survey, 1998-99 reveals that the surveyed households also had better access to services. Despite these biases, the sample covered by the Young Lives study captures the diversity of children in poor households in Andhra Pradesh.

Figure A.2. Socio-cultural Regions in Andhra Pradesh

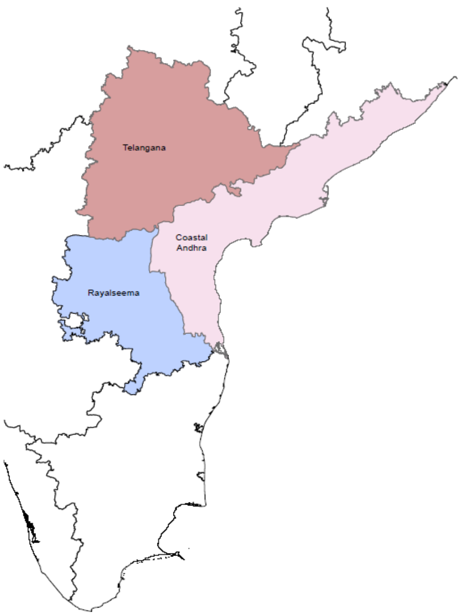
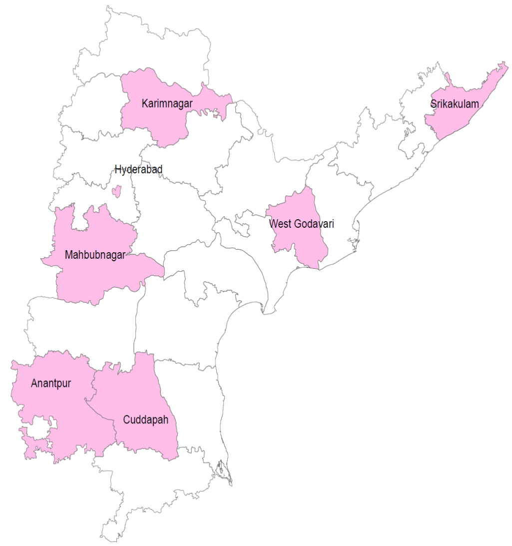


Figure A.3. Selected Districts in Andhra Pradesh



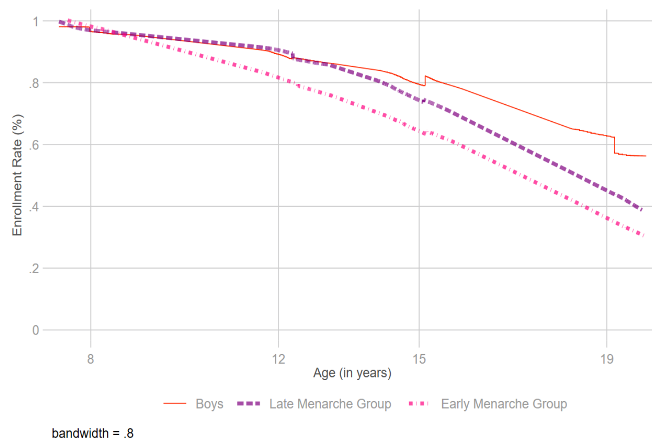
References

Centre for Economic and Social Studies. (2007). *Human Development Report 2007*. Retrieved from https://www.undp.org/content/dam/india/docs/human_revelop_report_andhra_pradesh_2007_full_report.pdf

B Robustness Checks

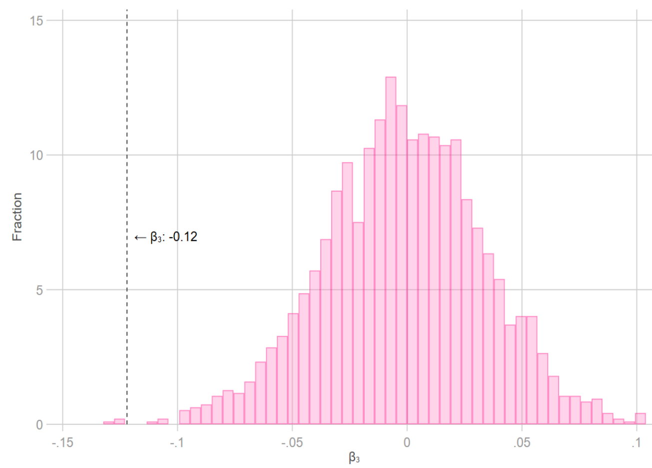
Figures

Figure B.1. Enrollment Rate by Gender and Menarche Status - locally weighted regression



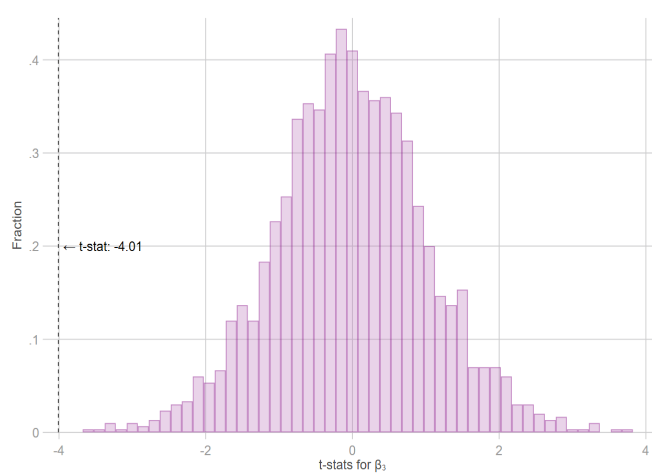
Note: Data are taken from the Young Lives study. The solid (red) line represents the trends in enrollment rate for boys, the dashed (purple) line represents the trends in enrollment rate for girls in the late menarche group and the dash-dotted (pink) line represents the trends in enrollment rate for girls in early menarche group.

Figure B.2. Permutation Tests for Estimates of β_3



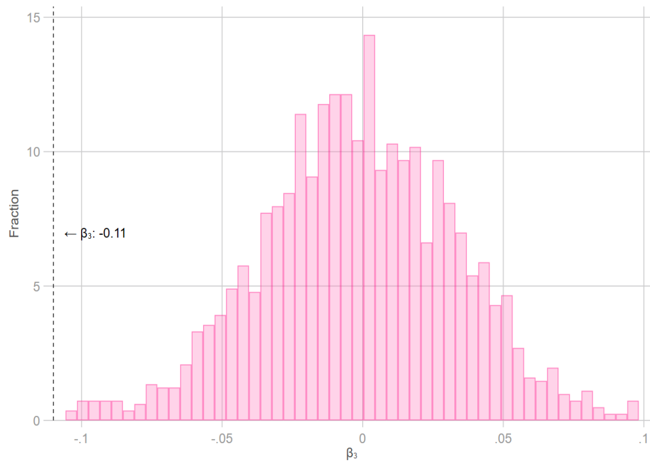
Note: This figure shows the distribution of estimates of β_3 from a permutation test of for equation 5.1 when girls in the early menarche group are compared to girls in the late menarche group. For each child in the comparison sample, menarche status was randomly generated and these assignments were used to compute β_3 2000 times. This figure shows the distribution of the estimates of β_3

Figure B.3. Permutation Tests for Estimates of t-statistics for β_3



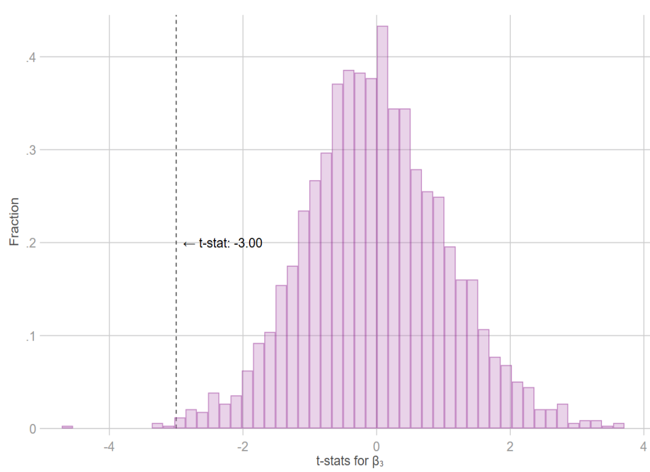
Note: This figure shows the distribution of t-statistics from a permutation test of for equation 5.1 when girls in the early menarche group are compared to girls in the late menarche group. For each child in the comparison sample, menarche status was randomly generated and these assignments were used to compute β_3 2000 times. This figure shows the distribution of the estimates of the associated t-statistics

Figure B.4. Permutation Tests for Estimates of β_3



Note: This figure shows the distribution of estimates of β_3 from a permutation test of for equation 5.1 when girls in the early menarche group are compared to boys. For each child in the comparison sample, menarche status was randomly generated and these assignments were used to compute β_3 2000 times. This figure shows the distribution of the estimates of β_3

Figure B.5. Permutation Tests for Estimates of t-statistics



Note: This figure shows the distribution of estimates of t-statistics from a permutation test of for equation 5.1 when girls in the early menarche group are compared to boys. For each child in the comparison sample, menarche status was randomly generated and these assignments were used to compute β_3 2000 times. This figure shows the distribution of the estimates of the associated t-statistics

Tables

Table B.1. Time Use Patterns of Older Siblings, Proportion of Time in the Day

	(1) Work	(2) Household Chores	(3) Care	(4) Sleep	(5) Play	(6) School	(7) Study
Early menarche	0.002	-0.004	-0.000	-0.007	-0.013	-0.014	-0.016
	(0.012)	(0.004)	(0.002)	(0.020)	(0.020)	(0.016)	(0.012)
Observations	496	496	496	496	496	496	496
R^2	0.0807	0.141	0.198	0.174	0.0794	0.170	0.0678
Control average	0.0155	0.0211	0.00561	0.304	0.158	0.196	0.0577
Sampling site FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results from a single-difference specification (see Equation 5.3) for data from the second survey round when children were twelve. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table B.2. Time Use Patterns of Younger Siblings, Proportion of Time in the Day

	(1) Work	(2) Household Chores	(3) Care	(4) Sleep	(5) Play	(6) School	(7) Study
Early menarche	0.034	0.037	0.034	0.048	0.024	0.045	0.032
	(0.034)	(0.034)	(0.034)	(0.028)	(0.042)	(0.030)	(0.033)
Observations	403	403	403	403	403	403	403
R^2	0.0866	0.0817	0.0810	0.128	0.0760	0.180	0.0995
Control average	0.0155	0.0211	0.00561	0.304	0.158	0.196	0.0577
Sampling site FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results from a single-difference specification (see Equation 5.3) for data from the second survey round. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table B.3. Robustness Check- Including Health Shocks as a Control

	(1) vs Late Menarche Group	(2) vs Boys
Early menarche	0.026	-0.005
	(0.022)	(0.014)
Round 2	-0.034	-0.048**
	(0.030)	(0.018)
Early menarche X Round 2	-0.122***	-0.111**
	(0.031)	(0.037)
Observations	1013	1234
R^2	0.162	0.129
Control average	0.906	0.903
Sampling site FE	Yes	Yes
Controls	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1) for comparisons that span the first and the second and the second and the third survey round. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Additionally, experience of health shocks in the past four years is included. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table B.4. Robustness Check- Including Economic Shocks due to Droughts as a Control

	(1) vs Late Menarche Group	(2) vs Boys
Early menarche	0.026	-0.005
	(0.022)	(0.014)
Round 2	-0.034	-0.048**
	(0.030)	(0.018)
Early menarche X Round 2	-0.122***	-0.111**
	(0.031)	(0.037)
Observations	1013	1234
R^2	0.162	0.129
Control average	0.906	0.903
Sampling site FE	Yes	Yes
Controls	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1) for comparisons that span the first and the second and the second and the third survey round. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Additionally, experience of health shocks in the past four years is included. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table B.5. Robustness Check- Effect on Young Lives Children's Siblings' Enrollment

	Round 1 vs 2	
	(1)	(2)
	vs Late Menarche Group	vs Boys
Early menarche	-0.060	-0.039
	(0.050)	(0.043)
Round 2	-0.074**	-0.096**
	(0.023)	(0.031)
Early menarche X Round 2	0.052	0.070
	(0.052)	(0.047)
Observations	1423	1729
R^2	0.198	0.196
Control average	0.748	0.682
Sampling site FE	Yes	Yes
Controls	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1) for comparisons that span the first and the second and the second and the third survey round. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Additionally, experience of health shocks in the past four years is included. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table B.6. Robustness Check- Individual Fixed Effects

	(1)	(2)
	vs Late Menarche Group	vs Boys
Round 2	-0.009	-0.030
	(0.037)	(0.036)
Early menarche X Round 2	-0.130**	-0.109**
	(0.045)	(0.052)
Observations	1013	1234
R^2	0.604	0.599
Control average	0.906	0.903
Individual FE	Yes	Yes
Controls	Yes	Yes

Note: This table reports the results from a difference-in-differences specification (see Equation 5.1). Controls include number of younger siblings, number of older siblings, number of older brothers and number of sisters. Individual fixed effects are also included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

C Heterogeneity in Impact on Enrollment: Additional Results

After menarche, the consequences of failing to ensure the girl's safety are more severe. As a result, additional restrictions are imposed on the extent of her social interactions and mobility. These restrictions are binding if the girl has to travel a long distance to attend school. The decrease in enrollment rate due to menarche is muted if she resides in a community that has a secondary school (Table C.1). Perhaps due to small sample size, the triple-difference estimate is not statistically significant.

Along with the availability of schools in the neighborhood, the availability of sanitation facilities at these schools could also characterize the local safety environment when a girl goes to school after her menarche (Adukia, 2017). As discussed in Section 2, qualitative reports describe that girls miss school during their menses because of a lack of sanitation facilities at their schools. To explore if menarche makes the concerns regarding lack of sanitation facilities more salient, data from the District Information System for Education[a] for Andhra Pradesh for 2006 is used to classify the Young Lives study districts into two groups: one group where the availability of toilets for girls is above the state median (65% of schools in a district), and one where the availability of toilets for girls is below the state median.[b] As a result of menarche, the decrease in enrollment is larger in the latter group (Table C.1). As before, the difference between these estimates is not significant.

Next, safety concerns are likely to be more serious in communities with a higher level of crime. To quantify the local crime environment, I use the incidence of different types of crimes reported in the community focus-group discussion in the second survey round to compute a 'crime-index'.[c] The impact of menarche on enrollment is not different across the sub-samples defined by high or low crime-index (Table C.2). The decrease in enrollment due to menarche is not mitigated by the presence of a police station in the community (Table C.2). However, crime-index and presence of a police station may not represent the children's (and their parents') perceptions of the safety of their neighborhood. This is especially true since police stations' presence is ineffective in deterring actual incidents of crime against women (Basu Roy & Dastidar, 2018).

References

Adukia, A. (2017). Sanitation and education. *American Economic Journal: Applied Economics*, 9(2), 23-59.

Basu Roy, S., & Ghosh Dastidar, S. (2018). Why do men rape? Understanding the determinants of rapes in India. *Third World Quarterly*, 1-23.

a. District Information System for Education is a database of all the schools in India. It is updated annually.

b. To protect the privacy of the children followed in the Young Lives study; any information on the location of their residence is not shared with external researchers. The lowest level at which their location can be identified is a district.

c. The components of this crime index include robbery, cattle theft, gangs, rivalry among groups, prostitution, drug addiction, alcoholism, and violence and might not be particularly relevant to young girls' safety concerns.

Tables

Table C.1. Table D1 Impact of Early Menarche of Enrollment, by School Infrastructure

	(1) Secondary School Presence	(2) Female Toilets Availability
Early menarche X Round 2	-0.146**	-0.113**
	(0.059)	(0.051)
Early menarche X Round 2 X Secondary School Present	0.035	
	(0.071)	
Early menarche X Round 2 X Lower Toilet Availability		-0.015
		(0.064)
Observations	1013	1013
R^2	0.170	0.162
Control average	0.906	0.906
Sampling site FE	Yes	Yes
Controls	Yes	Yes

Note: This table reports the results from a triple-difference specification (see Equation 5.2) for comparisons that span the first and the second survey rounds. Column (1) reports results when the difference-in-differences specification is interacted with the presence of a secondary school in the community and Column (2) reports results when the difference-in-differences specification is interacted with if the girl resides in a district where the availability of toilets for girls in schools is greater than the median availability in the state. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table C.2. Impact of Early Menarche of Enrollment, by Local Crime Environment

	(1) Community Crime Index	(2) Police Station Present of Safety
Early menarche X Round 2	-0.118**	-0.116**
	(0.044)	(0.044)
Early menarche X Round 2 X Low crime community	-0.008	
	(0.076)	
Early menarche X Round 2 X Police Station Present		-0.009
		(0.075)
Observations	1013	991
R^2	0.165	0.164
Control average	0.906	0.906
Sampling site FE	Yes	Yes
Controls	Yes	Yes

Note: This table reports the results from difference-in-differences specification (see Equation 5.1) for comparisons that span the first and the second survey rounds. Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. Sampling-site fixed effects are included. Robust standard errors clustered at the level of the sampling site are reported.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

D Early Menarche as an Instrument for Early Marriage

This section will explore the validity of early menarche as an instrument for early marriage. Following Conley, Hansen & Rossi (2012), I seek to understand the validity of the instrument to the plausible violation in the exclusion restriction through the lens of the following population model:

$$Y = X\beta + Z\kappa + \epsilon \tag{E.1}$$

A violation of the exclusion restriction would imply that κ is not zero. Under certain assumptions on the point estimate of κ or the distribution it is drawn from, one can estimate bounds around the two-stage least squares estimate. More specifically, if the true value of κ, say κ_0, was known then we could estimate β by subtracting $Z\kappa_0$ from both sides of the equation E.1:

$$Y - Z\kappa_0 = X\beta + \epsilon \tag{E.2}$$

Next, we can use the usual asymptotic approximations to estimate 95% confidence intervals around β under the assumption that $\kappa = \kappa_0$. However, we do not know κ_0. Let's assume that κ is drawn from the support G, and the true value is $\kappa_0 \in G$. For all points in the support G, one can estimate the 95% confidence interval around β^{IV} using two methods. The union of confidence intervals method assumes that only the support of κ is known, and the union of confidence intervals corresponding around β^{IV} for different values of κ in G will contain the population parameter. A primary drawback of this methodology is that the confidence intervals are often too wide. A second methodology discussed by Conley, Hansen & Rossi (2012) assumes that κ is a random draw from a known distribution, and the knowledge of this distribution is used to compute the 95% bounds around the population parameter.

This paper uses these methodologies to discuss the validity of early menarche as an instrument for early marriage. First, I take κ from the Young Lives study. Next, I use the nationally representative second round of the India Human Development Survey (IHDS) to estimate the effect of early marriage on completed schooling using an instrumental variable strategy and explore the stability of this estimate for the estimate of κ derived in the first step. I consider two specifications in the first step: one with sampling site fixed effects (the preferred specification in the paper) and one with district fixed effects (the specification that can be estimated with the IHDS data).

Estimates of κ from the Young Lives study

As discussed in Section 6, reaching menarche before age twelve leads to a 13.4% decline in school enrollment rate, and a variety of robustness checks support this result. Table D.1 describes the relationship between early menarche and completed education.[d] The difference in the completed

d. For the girls who are still enrolled at twenty-two, the completed education is taken to be their current completed education. Since the children have not completed their schooling by the fifth survey round, this difference does not years of education for girls in the early and the late menarche groups, though not statistically significant, is 0.91 years, that is, girls in the early menarche group get 0.80 years less education, on average than other girls in their cohort. We can take the estimate of the relationship between early menarche and years of schooling to recover the bounds around the two-stage least squares estimates. I also replicate this specification with district fixed effects instead of sampling site fixed effects, and the difference in the completed education across the two groups of girls is 1.01 years. Finally, the last column presents the results of the effect of being unmarried by eighteen on completed schooling while using the incidence of early menarche to isolate the variation in the timing of marriage. The instrumental variable approach involves estimating a two-stage model which is specified as follows:

$$\text{Second Stage}: YearsSchooling_{id} = \beta \widehat{Unmarriedby18}_{id} + \gamma X_{id} + \eta_d + \epsilon_{id} \tag{E.3}$$

$$\text{First Stage}: Unmarriedby18_{id} = \rho EarlyMenarche_{id} + \gamma X_{id} + \eta_d + \mu_{id} \tag{E.4}$$

where $YearsSchooling_{id}$ is the completed years of schooling for woman i in district d, $Unmarriedby18_{id}$ is an indicator equal to one when the woman had not married by eighteen, and $EarlyMenarche_{id}$ is an indicator equal to one if the woman started menstruating before twelve. X_{id} denotes a vector of additional controls, and η_d denotes district fixed effects.[e] The two-stage least squares estimate is seven years. However, F-statistic from the first stage is only 8.3, and the problem of a weak instrument renders these results not suitable for estimating bounds around the two-stage least squares estimates (Conley, Hansen & Rossi, 2012).[f] Therefore, the following discussion explores the stability of these estimates using a larger nationally representative dataset.

Two-stage Least Squares Estimation Using the IHDS Data

The second round of the IHDS was conducted in 2011-12. The interviews with ever-married women aged 15-49 collected recall data on age at menarche, years of schooling, and age at first marriage. As is common for the literature that uses the timing of menarche as an instrument for the timing of marriage, I restrict the sample to women who were still in school at age nine (Field & Ambrus, 2008). In addition, I consider only those women who married after menarche.[g] The distribution of the difference between age at dropout and age at menarche is bimodal, where one of the two peaks is at zero: a substantial mass of girls drop out of school right after menarche (Figure D.1). The distribution of the difference between at age at dropout and age at menarche shows that most girls marry about five years after dropping out of school.

For estimating results comparable to the ones described in Table D.1, I included control variables that were similar to the set of controls included in the preferred specification estimated using data from the Young Lives study. The final set of controls includes current land ownership, current log

reflect the final difference in the completed education across the two groups of girls. Importantly, this difference is the lower bound on the difference in completed education as more girls in the late menarche group are likely to continue education because they are 34% more likely to be in college at twenty-two.

e. Since the evidence from the IHDS data uses district fixed effects, district fixed effects are included.

f. The 95% confidence interval using the union of confidence intervals method is [-6.8, 14.7] and using the local to zero method when it is assumed that κ is drawn from a uniform distribution on the support (-1.01,0) is [-0.8, 13.5].

g. Twenty-one girls were not enrolled in school at age ten in the Young Lives study. The second restriction is not binding.

per-capita consumption, age, age-squared, location of residence (urban or rural), and the number of members in the household. In addition, variables that capture information about the woman's natal home are included: an indicator for if the natal home is economically better off than the marital home, father's education, mother's education, number of brothers, and number of sisters. The IHDS does not contain information about the geographical location of the woman's natal home. To capture her childhood experiences that may be correlated to her schooling, I include district fixed effects.[h] Robust standard errors clustered at the level of a district are reported.

The two-stage least squared estimates show that being unmarried by eighteen increases schooling by 4.6 years (Table D.2). The 95% confidence interval estimated using the union of confidence intervals method is [-12.6, 6.1] when κ lies in the support [-1.01,0].[i] To implement the local-to-zero method, it is assumed that κ is drawn from a uniform distribution on the support $(\kappa^L, 0)$. The 95% confidence interval in this case when κ^L is -1.01 is [-9.3, 4.5]. Thus, the two-stage least squares estimates are not robust to departures from the exclusion restrictions. Figure D.2 and Figure D.3 describe the 95% confidence intervals for different levels of departure from the exclusion restriction.

References

Conley, T. G., Hansen, C. B., & Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1), 260-272.

Bloch, F., Rao, V., & Desai, S. (2004). Wedding celebrations as conspicuous consumption signaling social status in rural India. *Journal of Human Resources*, 39(3), 675-695.

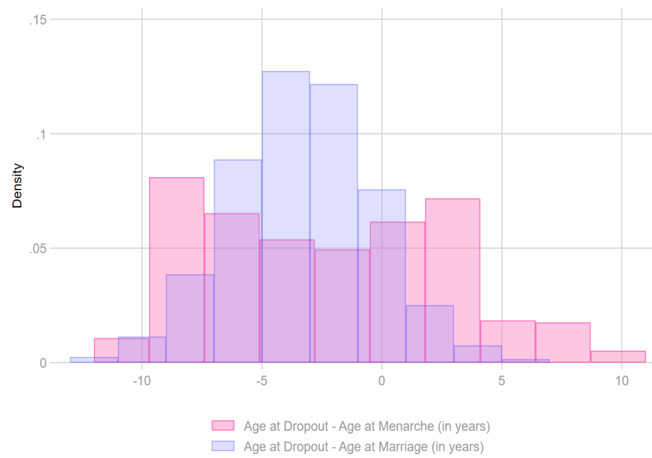
Field, Erica, and Attila Ambrus. 2008. Early marriage, age of menarche, and female schooling attainment in Bangladesh. *Journal of Political Economy*, 116(5): 881-930.

h. Bloch, Rao, & Desai (2002) estimate that the average distance from between martial and natal homes is twenty-one miles, and therefore her natal home is likely to fall within the district of current residence.

i. The estimated value of κ from the Young Lives Study is -1.01. Therefore, it is taken to be the upper bound of the violation of the exclusion restriction.

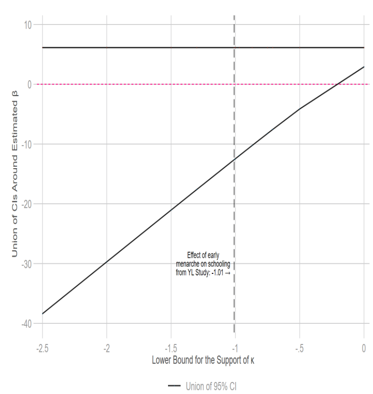
Figures

Figure D.1. Age at Marriage, School Dropout and Marriage



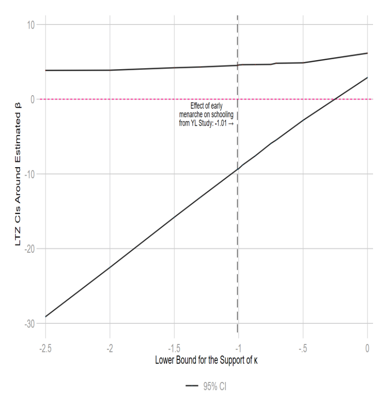
Note: Data are taken from the IHDS. The histogram in the lighter (pink) shade represents the difference between the age at which girls drop out of school and the age at which they reached menarche (in years) and the histogram in the darker (purple) shade represents the difference between the age at which girls drop out of school and the age at which they get married (in years).

Figure D.2. 95% CI Around IV Estimates, UCI Method



Note: Data are taken from the IHDS. Grey lines represent the 95% CI derived using the UCI method.

Figure D.3. 95% CI Around IV Estimates, LTZ Method



Note: Data are taken from the IHDS. Grey lines represent the 95% CI derived using the LTZ method.

Tables

Table D.1. Menarche and Years of Schooling, Evidence from the YL Study

	(1)	(2)	(3) (OLS)	(4) (2SLS)
Early menarche	-0.910	-1.010*		
	(0.560)	(0.451)		
Unmarried by 18			3.041***	7.041***
			(0.295)	(0.835)
Observations	474	474	474	474
R^2	0.331	0.290	0.407	0.185
Control average	12.75	12.75	12.75	12.75
Sampling site FE	Yes	Yes	Yes	Yes
District FE	No	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Fstat				8.377

Note: Columns (1)-(2) report results from single-difference specification (see Equation 5.3). Controls include household size, caste, sex of the household head, wealth index,mom's literacy status, age of the head of the household, indicator of whether the child speaks the majority language, number of younger siblings, number of older siblings, number of older brothers, number of older sisters, test scores at age eight, weight at age eight, and BMI at age eight. In column (1), sampling site fixed effects are included and standard errors are clustered at the sampling site level. In column (2), district fixed effects are included and standard errors are clustered at the district level. The last two columns present the results on the effect of being unmarried until eighteen on completed schooling using OLS and 2SLS specifications. Estimation is restricted to those who dropped out of school after ten.

* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table D.2. Menarche and Years of Schooling, Evidence from the IHDS Data

	(1)	(2) OLS	(3) 2SLS
Early Menarche	-0.334***		
	(0.071)		
Single by 19		1.600***	4.604***
		(0.049)	(0.975)
Observations	21591	21591	21591
R^2	0.391	0.440	0.265
Control average	3.907	3.907	3.907
District FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Fstat			42.03

Note: Column (1) report results from single-difference specification (see Equation 5.3). Controls include household size, caste, age, age-squared, log per capita consumption, indicator for land ownership, location (rural or urban), indicator if the natal home if economically better-off, mother's education, father's education, mother's education, number of brothers and number of sisters. District fixed effects are included and standard errors are clustered at the district level. The last two columns present the results on the effect of being unmarried until eighteen on completed schooling using OLS and 2SLS specifications. Estimation is restricted to those who dropped out of school after ten and married after menarche.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.