CAF - WORKING PAPER #2021/01
First version: November 25, 2021
Current version: June 9, 2023

Intergenerational transmission of teen childbearing in Latin America

Matilde P. Machado[1] | **Ricardo Mora**[2] | **Karen Olivo**[3]

[1]Associate Professor, Department of Economics, Universidad Carlos III Madrid. matilde.machado@uc3m.es

[2]Associate Professor, Department of Economics, Universidad Carlos III Madrid. ricmora@eco.uc3m.es

[3]Karen.olivo@tec.Mx

Latin America has experienced high rates of teen childbearing for decades. Using DHS data for six Latin American countries, we estimate the relation between a mother's teenage childbearing status and that of her daughter. Restricting the estimating sample to mother-daughter matches in the data leads to a large negative selection bias in the estimated effect because missing matches are nonrandom and affected by the teen childbearing status of mothers and daughters. We deal with this selection bias by developing a maximum likelihood estimation using all available data, including incomplete mother-daughter pairs, and allowing missing observations to be endogenous. Our results show that being the daughter of a teen mother increases the chances of being a teen mother between 9.1 and 23.7 percentage points (75 and 123% relative to the mean incidence of teen childbearing). We conclude that the prevalence of such high intergenerational transmission is at the core of persistent high teenage childbearing rates in Latin America.

KEYWORDS

teen childbearing, teen pregnancy, teen motherhood, intergenerational transmission, nonignorable missingness, econometric methods for missing data, DHS data, reproductive health in Latin America.

CAF - DOCUMENTO DE TRABAJO #2021/01

Primera Versión: noviembre 25, 2021
Versión Actual: Junio 9, 2023

Transmisión intergeneracional de la maternidad adolescente en América Latina

Matilde P. Machado[1] | **Ricardo Mora**[2] | **Karen Olivo**[3]

[1]Associate Professor, Department of Economics, Universidad Carlos III Madrid. matilde.machado@uc3m.es

[2]Associate Professor, Department of Economics, Universidad Carlos III Madrid. ricmora@eco.uc3m.es

[3]Karen.olivo@tec.Mx

América Latina ha experimentado altas tasas de maternidad adolescente por décadas. Utilizando datos de las Encuestas Demográficas y de Salud (DHS) para seis países Latinoamericanos, estimamos la relación entre la condición de maternidad adolescente de una madre y de su hija. Restringir la muestra a los emparejamientos madre-hija en los datos conduce a un gran sesgo de selección negativo en el efecto estimado, dado que los emparejamientos faltantes no son aleatorios y están afectados por la condición de maternidad adolescente de madres e hijas. Abordamos este sesgo de selección desarrollando una estimación por máxima verosimilitud usando todos los datos disponibles, incluyendo pares incompletos madre-hija, y permitiendo a las observaciones faltantes ser endógenas. Nuestros resultados evidencian que ser hija de una madre adolescente aumenta la probabilidad de ser una madre adolescente entre 9.1 y 23.7 puntos porcentuales (entre 75 y 172% relativo a la incidencia promedio de maternidad adoslescente). Concluimos que la prevalencia de tan alta transmisión intergeneracional está en el centro de las elevadas tasas de persistencia de maternidad adolescente en América Latina.

KEYWORDS
Maternidad adolescente, embarazo adolescente, madre adolescente, transmisión intergeneracional, datos faltantes no ignorables, Encuestas Demográficas y de Salud, salud reproductiva en América Latina

1 | INTRODUCTION

In Latin America and the Caribbean (LAC), teen childbearing has been a significant issue for several decades. The region ranks third globally when it comes to the proportion of women initiating childbearing during adolescence, trailing only sub-Saharan Africa and Oceania (UNFPA 2022). This disparity with developed nations is striking, as evidenced by the statistics from 2020: LAC recorded a rate of 60.6 births per 1000 girls aged 15-19, in contrast to 15.9 in the United States and 7.3 in Canada (United Nations Population Division, World Population Prospects). The significance of this problem stems from its association with poverty and unfavorable socioeconomic outcomes (Fletcher and Wolfe 2009, Azevedo et al. 2012, Ashcraft et al. 2013, Lang and Weinstein 2015). In this paper, our focus lies specifically on examining the transmission of teen childbearing from mothers to daughters within the LAC context, a region known for its limited social mobility.[1]

Persistently high rates of teen childbearing may be due to two different scenarios that have opposite implications for social policy. First, persistence may result from external factors that are independent of the social and family environment and can be considered random and unavoidable. In this case, there would be no transmission of the problem from mothers to daughters. Alternatively, if persistence results from factors closely related to the social and family environment of the woman, then intergenerational transmission from mother to daughter would occur. The phenomenon of teen childbearing would then be regarded as a social problem of inequality, necessitating appropriate policy solutions. For example, preventive efforts should not miss those teenagers at greater risk, i.e., the daughters of teen mothers. Therefore, obtaining a measure of inter-generational persistence in teen-childbearing carries considerable policy relevance.[2]

The study of the mother-daughter link in teen-childbearing, however, poses a significant challenge in terms of data availability. In many countries, including most from the Latin American region, the only available databases to measure the teen childbearing status (TCS) of both mothers and daughters are household surveys or census data. Matching all mothers and daughters in such data, however, can be complicated as mothers and daughters do not always live in the same household. Consider, for example, a typical household survey where information is gathered only for individuals living in the household. Teenage daughters who no longer live with their parents usually appear as household heads or spouses of the household head and information on their mothers is missing. Similarly, the daughter information is missing in households of interviewed mothers whose daughters no longer live with them. These unmatched women are not present when data are restricted to matched mother-daughter pairs, i.e., the pairs of mothers and daughters living in the same household. Because teen childbearing increases the probability that the daughter leaves the parental home to live with her partner, unmatched mother-daughter pairs are not missing at random, i.e. they are nonignorable. As a consequence, the estimation of the relation between a mother's TCS and that of her daughter with only matched pairs results in coresidential bias. Standard selection methods à la Heckman are not directly applicable to our model

[1]The persistence of social inequality may be influenced by personal traits, family characteristics, and behaviors that are passed down from parents to children (Black and Devereux 2011, Heckman and Mosso 2014). For studies on mother-daughter transmission of teenage childbearing in other countries and regions see, among others, Card (1981), Kahn and Anderson (1992), Manlove (1997), Francesconi (2008), Haveman et al. (2008), and Aizer et al. (2020). For studies on the low social mobility in Latin America, see, for example, Berniell et al. (2021), Neidhöfer et al. (2018), Daude and Robano (2015), Torche (2014), and Azevedo and Bouillon (2010).

[2]Our focus on accurately quantifying the mother-daughter association in teen childbearing is akin to the importance placed on securing unbiased estimates of the intergenerational transmission of permanent income (for results and the methodological challenges in this literature, see reviews by Solon, 1999, Black and Devereux, 2011, Jäntti and Jenkins, 2015, and Stuhler et al., 2018).

due to the presence of missing data in both the mother's and daughter's TCSs, as well as their characteristics. We deal with this problem by the development of an empirical approach that delivers unbiased estimation when unmatched mother-daughter pairs are nonignorable. Using our methodology with a sample of six Latin American countries, we find that having a teen mother increases the probability of teen childbearing, between 9.1 and 23.7 percentage points (or, relative to their mean teen childbearing incidence, between 75 and 123%). These country-specific estimates are four to fifteen times larger than those obtained when estimation is restricted to matched mother-daughter pairs. By uncovering the true magnitude of the intergenerational transmission, our results show that inertia within the family is a sizable component in the persistence of teen childbearing in our sample of Latin American countries.

We use Demographic and Health Survey data (DHS) for the following Latin American countries: Bolivia, Colombia, Dominican Republic, Guatemala, Haiti and Peru. There are two essential features of the DHS that makes it suitable for our analysis. First, the DHS uses a common questionnaire in all countries and therefore allows for international comparisons. Second, it provides detailed information on birth histories, which we can use to establish whether women were teenagers at first birth. There are no good alternatives to the DHS for the six Latin American countries that we consider, an issue we further discuss in section 2.

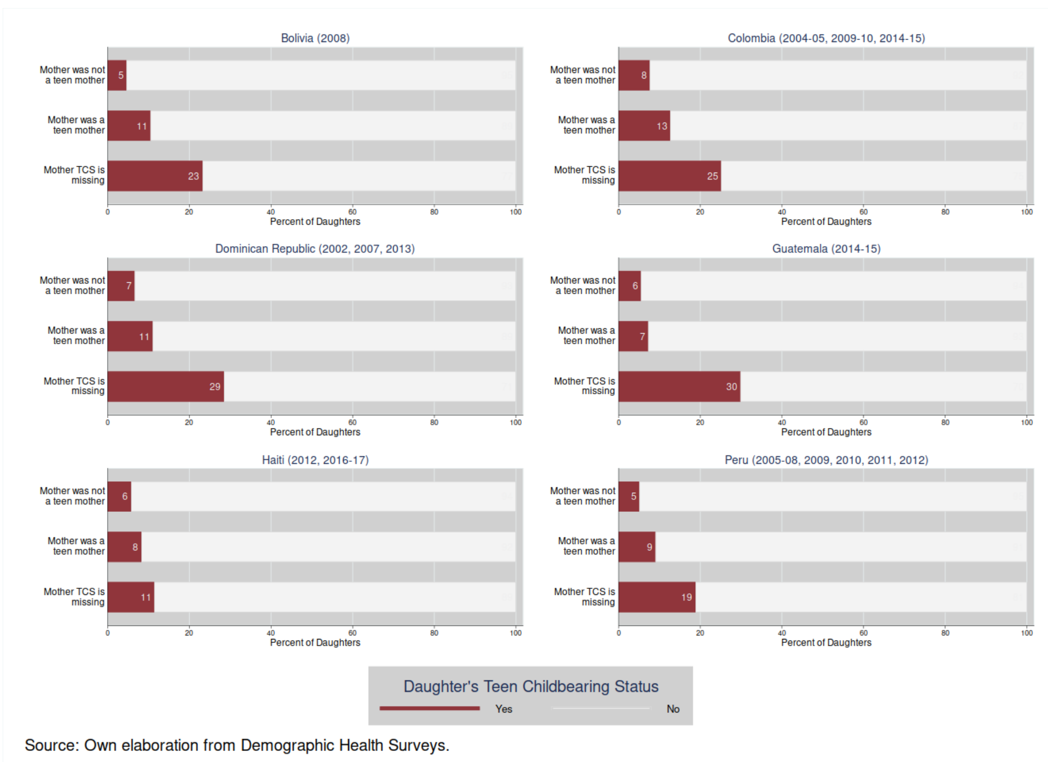

FIGURE 1 Daughter's Teen Chidbearing Status by her Mother's. *Notes:* Own elaboration using birth histories from sample *Standard Demographic Health Surveys* (DHS). These are: (i) Bolivia: 2008; (ii) Colombia: 2004-05, 2009-10 and 2014-15; (iii) Dominican Republic: 2002, 2007, and 2013; (iv) Guatemala: 2014-15; (v) Haiti: 2012, 2016-17; and Peru: 2005-08, 2009, 2010, 2011, and 2012.

A first look at 15-18 year-old teens and their mothers in the DHS data, as shown in Figure 1, reveals that daughters of teen mothers have a two- to six-percentage points higher probability of teen childbearing compared to daughters whose mothers were not teen mothers.[3] The figure also shows that teens whose mothers' TCS is missing—that is, daughters who are unmatched to their mothers—are close to three times more likely to be teen mothers. This difference strongly suggests that missing mother-daughter matches are endogenous to the process of teen childbearing. Excluding them from the estimation sample may not be a cause of concern when they represent a small percentage of observations. In our sample, however, half of the observations are unmatched, i.e., they have either information on the daughter or on her mother, but not both.

We address these concerns by means of a maximum likelihood (ML) procedure that uses all available data, including unmatched daughters and unmatched mothers, and allows the missing process to depend on the TCS of both daughters and mothers. Essentially, we start with a probit model where the daughter's TCS depends on her mother's TCS (and other characteristics). We add a nonparametric model of the conditional missing process to this probit. There are three possible situations: We either observe only the daughter, only her mother or we observe both. We ensure the identification of the parameters of the model by placing exclusion restrictions in the conditional missing process as in Ramalho and Smith (2013) so that the probability of a match only depends on the daughter and the mother's TCS. In our case, the exclusion restrictions imply that any correlation between the events that the mother or daughter are missing and the controls in the probit specification occurs only through the direct effect of the latter on the daughter's and mother's TCS. These restrictions are weaker than those imposed when estimation uses only the matched mother-daughter pairs, allowing us to test the null that we can exclude the unmatched pairs in the model's estimation. Using this test, we strongly reject it.

We employ several waves of the DHS to estimate the model for each country. After presenting the main findings in section 4, we then use indirect evidence to inform a discussion on the causal interpretation of our estimates. Specifically, we expand our main variable specification to investigate if the inclusion of potentially endogenous characteristics of the daughter influences the estimated mother-daughter inertia in TCS. By comparing the results, we can assess for which countries the interpretations of the estimates lean more toward a causal explanation and for which ones we find less support for such an interpretation.

We perform two extensions in section 5. We first estimate our model on different subsamples defined according to the daughter's age and the household's wealth. The main conclusions from section 4 remain valid: the role of the mother's TCS on the daughter's probability of teen motherhood is large and very statistically significant, and discarding the unmatched pairs always results in large negative biases in all countries and subsamples.

Our second extension posits that the intergenerational pattern of teen childbearing is one of many interrelated events, all shaped by the mother's TCS. Specifically, we estimate models connecting a mother's TCS with her daughter's likelihood of initiating sexual activity at an early age, marrying during adolescence, and only completing primary education. The consistency in direction, magnitude, and statistical significance of the estimates across all models strengthens the hypothesis that all these outcomes share a common underlying process.

Based on our analysis, we may conclude that the enduring prevalence of high rates of teen childbearing in the region can be attributed to factors deeply rooted within the family.

The remainder of this paper is structured as follows. We present the data in section

[3]Throughout the paper, we define a mother as a teen mother if the eldest of her live births was born when she was a teenager. Some studies use data on siblings to assess the differential effect of being the child of a teen pregnancy on teenage motherhood status (such as Francesconi, 2008). Note that our estimates, based on our definition of teen motherhood, encompass not only the intergenerational link with the first daughter but also the link with other siblings.

2, the econometric model and the empirical strategy in section 3, and the main results in section 4. Section discusses several possible extensions, and section 6 presents a discussion of the results and concludes. Appendix A derives the Likelihood function and Appendix B presents additional tables.

2 | DATA

We use comparable individual-level data from the standard Demographic and Health Survey data (DHS) for the following Latin American countries: Bolivia, Colombia, Dominican Republic, Guatemala, Haiti and Peru. DHSs are nationally representative household surveys comprising independent cross-sections conducted approximately every five years. DHSs are widely used by researchers because, among other reasons, they provide comparable information across countries. In addition, DHSs collect live birth histories from all women in the household aged between 15 and 49. Therefore, it is possible to compute TCSs for both mothers and daughters participating in the survey.

Here we argue that when studying the mother-to-daughter transmission of teen-childbearing, there are no good alternatives to the DHS for the six Latin American countries in our sample. One possibility would be to use administrative records with date of birth or age of individuals and link, using personal identifiers, all members of the same family regardless of their residence. To the best of our knowledge, this option is not available for Latin American countries. The second option would be to use data that extends its time coverage over a full generation so that the daughter's TCS is revealed. (Examples are longitudinal data or birth registries with information that allows the linkage of the daughter's and the mother's records.) Such datasets do not exist yet for the six countries in our sample: The ELCO data for Colombia (from the Spanish acronym "Encuesta Longitudinal de Colombia") and the Young Lives survey from Peru are the only longitudinal datasets available that may be suitable for this analysis. However, neither of these datasets covers information on a full generation. The third option would be to use surveys that include birth histories, as is the case of the DHS.[4] One drawback of these surveys, however, is that they only interview household members. As a result, it becomes impossible to match mothers and daughters who live in different locations.[5] Hence, this last alternative is only acceptable if the problem of co-residential bias is addressed.

Our population of interest is teenage women aged 15 to 18 and their mothers.[6] Unique household and individual identifiers permit the matching of teenage daughters to their mothers aged 49 or younger living in the same household. However, some interviewed daughters cannot be matched with their mothers because either (i) their mothers are older

[4]Birth histories are not part of most census questionnaires or of household and labor force surveys where information about children is restricted to those living in the household. (An example is the Harmonized Household Surveys of Latin America and the Caribbean.) This is also a problem with longitudinal data that do not cover a full generation. For example, information related to birth history in ELCO is missing for most observations in all currently available waves and there is no birth history information in Young Lives. Without the mother's birth history, mothers who were teens when they became pregnant with a child who is no longer in the household would be wrongly classified as non-teenage mothers.

[5]Certain questionnaires don't require the construction of mother-daughter pairs as they simply ask respondents about their mother's characteristics. However, the accuracy and completeness of the retrospective information provided by the child are essential for obtaining reliable data. We are not aware of the existence of these data for Latin American countries.

[6]If we extend the population of daughters to include women older than 18, the proportion of unmatched pairs increases significantly. On the other hand, the presence in the sample of daughters who have not yet completed their adolescence may generate a right-censoring problem. In section 5, we address the issue of right censoring by examining results across various age groups.

than 49 or deceased[7] or (ii) they live in a different household. Similarly, some interviewed mothers cannot be matched to their teenage daughters because (i) the daughter was not interviewed (exceptional) or (ii) she does not live with her mother.[8]

2.1 | Early and sample DHS

Our empirical strategy does not require the imputation of daughter's or mother's characteristics for the unmatched pairs. Nonetheless, as we show in Appendix A, the likelihood function includes as arguments parameters that characterize the joint distribution of the mothers' TCS, their cohort, and education. For simplicity, we refer to each of these parameters as a 'teen childbearing rate' (TCR).

We avoid the estimation of TCRs (whose potentially large number may risk sample identification) by replacing them with their values obtained from precise external data. Because we lack a long and homogeneous series from a different source, we compute TCRs using DHS data from other periods. Specifically, we classify all available DHSs between those that we use to estimate the model—the 'sample surveys'—and those that can only be used in the computation of TCRs—the 'early surveys'.

To illustrate how we distinguish early from sample surveys, consider the case of Colombia, a country that participated in 1986 (wave I), in 1990, and subsequently every five years until 2015. The oldest women in Colombia for whom we have birth history information are those interviewed in 1986 who were, at the time, 49 years old (i.e., born in 1937). If we assume that mothers of teen daughters are between 27 and 64 years, then those born in 1937 will be 64 years old in 2001. Therefore, the first sample survey for Colombia is 2005 because it is the first survey for which we have TCRs by age for all women aged 27 to 64. The 2000 Colombian DHS cannot be a sample survey because we would need TCRs for women aged 64, i.e., for those born in 1936.[9]

The six countries that we use are those with enough early surveys to compute TCRs for sample surveys. The resulting set of countries provides a significant representation of the socioeconomic and cultural differences between countries in the region (see Table B.1 in Appendix B). In Table 1, we summarize all available DHSs for each of the six countries, distinguishing sample surveys that form the estimation sample from early surveys that are not included in the estimation sample. To ensure that TCRs are external information in each specific sample survey, we calculated the TCRs for each sample survey using all the other DHSs available. For example, for the Colombia 2005 sample, we compute the TCRs with all other DHSs, including not only Colombia's early surveys but also Colombia 2010 and 2015.

2.2 | Estimation sample

In Table 2, DHS sample sizes include all matched and unmatched pairs of mothers and their teenage daughters. The number of observations refers to the total number of mother-daughter pairs by country and wave.

[7] Girls up to age 14 are asked whether their mother is alive. Only 2.37% of girls aged 14 answered that their mother was not alive.

[8] The proportion of women aged 25-49 who had a daughter who died in adolescence is only 0.32%. We observe neither the daughter's cause of death nor her TCS. Given the Maternal Mortality rate (Schwarcz and Fescina 2000) and observed teen pregnancy rates in these countries, we could expect at most 0.03% of observations in our sample where the daughter is missing because she died during pregnancy. Therefore, taking into account the negligible potential for bias, we include in our study only women whose teenage daughters are alive.

[9] If we were to use external information from other sources (such as official statistics on age-specific fertility rates), the estimation sample would be even more constrained as these statistics start in the 1970s at the earliest and, in most cases, they are not available by education.

TABLE 1 Available Standard Demographic Health Surveys

Country	Early DHS	Sample DHS
Bolivia	1989, 1993-94 1998, 2003-04	2008
Colombia	1986, 1990 1995, 2000	2004-05, 2009-10, 2014-15
Dominican Republic	1986, 1991, 1996, 1999	2002, 2007, 2013
Guatemala	1987, 1995 1998-99	2014-15
Haiti	1994-95, 2000 2005-06	2012, 2016-17
Peru	1986, 1991-92, 1996 2000, 2003-04(*)	2005-08, 2009, 2010 2011, 2012

Notes: All available DHSs for the selected countries. Notation "YYY1-Y2" indicates that the survey was conducted in all years from YYY1 to YYY2. For each country, we distinguish early Demographic Health Surveys (DHSs) (under the heading "Early DHS") from later DHSs (under the heading "Sample DHS"). Sample surveys make up our estimation sample. Early surveys comprise all available DHSs prior to the first sample DHS. For each sample DHS we compute cohort-specific TCRs using all available DHSs but that one.
(*) Peru 2003-04 cannot be included in the Sample DHS because it is not possible to match mothers with their daughters.

These totals vary mostly by country and are fairly stable by wave. Overall, our sample consists of 65,008 observations. Mother-daughter matches (reported as percentage under the heading 'Matched pairs (%)') represent on average 50.28% of the observations and, with the exceptions of Haiti and Dominican Republic 2002, they vary little across countries (ranging from 46.02% to 56.57%). Pairs where the mother is not interviewed are also common, on average 45.69% of all pairs, with four DHSs where it is the most common situation. Last, although the proportion of pairs in which the daughter is not interviewed is residual, ranging from 1.11% in Haiti 2016-17 to 6.87% in Dominican Republic 2002, we include them in our estimation sample.

Each DHS provides demographic characteristics for every member of the household—such as age, gender, marital status, and education—and basic information on the characteristics of the household—such as relation to the household head, number of members of the household ('household size'), a wealth index, and whether the household is located in a rural area.

TABLE 2 Sample DHSs: Sizes and missing observations

	Number of observations	Matched pairs (%)	Mother missing (%)	Daughter missing (%)
Bolivia 2008	2,982	46.38	50.34	3.29
Colombia 2004-05	6,101	51.60	43.62	4.79
Colombia 2009-10	7,906	54.91	41.17	3.92
Colombia 2014-15	5,716	50.59	44.16	5.25
Dominican Republic 2002	4,149	41.31	51.82	6.87
Dominican Republic 2007	5,112	46.19	48.92	4.89
Dominican Republic 2013	1,532	46.02	48.96	5.03
Guatemala 2014-15	4,845	56.57	40.72	2.70
Haiti 2012	2,886	36.56	61.92	1.52
Haiti 2016-17	2,789	39.30	59.59	1.11
Peru 2005-08	5,771	53.13	42.87	4.00
Peru 2009	3,992	55.69	41.31	3.01
Peru 2010	3,776	52.60	43.67	3.73
Peru 2011	3,654	53.75	42.36	3.89
Peru 2012	3,797	52.88	42.56	4.56
All sample DHSs	65,008	50.28	45.69	4.04

Notes: DHS samples include all matched and unmatched pairs of mothers and their teenage daughters aged 15 to 18. "Number of observations" refers to the total number of pairs by country and wave. The proportion of pairs of mothers, aged 19 to 49, and their teenage daughters who are both interviewed is shown under the heading "Matched pairs (%)". Column "Mothers missing (%)" shows the percentage of mother-daughter pairs where the mother is not interviewed and, hence, cannot be matched to her daughter. Column "Daughters missing (%)" shows the percentage of mother-daughter pairs where the daughter is not interviewed.

In addition, the DHS contains information to compute teen childbearing status of interviewed mothers and daughters and detailed information on i) sexual behavior, such as age at first sexual intercourse; ii) personal beliefs regarding gender roles in sexual relations; iii) knowledge of contraceptive health, such as fertility status throughout the menstrual cycle; and iv) fertility preferences such as the ideal number of children.

Our main dependent variable is a dummy for the TCS of the daughter, which takes a value of one if the teenager is pregnant or has given birth to a live child. (As an extension, in section 5, we also consider as dependent variables dummies for sex before 16 years old, for marriage before the age of 19, and for school dropout.) Our main covariate of interest is a dummy for the TCS of the mother, that is whether the mother was a teen when her first child was born. Other controls include dummies for year of interview, mother's and daughter's age, mother's household is located in a rural area, mother's household size is larger than 6, parents' household belongs to the two poorest quintiles of the country's population based on a continuous wealth measure produced by the DHS, daughter is the firstborn, and mother's education is at most primary level. In an extended specification we add daughter's characteristics that are arguably endogenous, such as whether the daughter has no more than primary education, her knowledge of fertility throughout the menstrual cycle, her ideal number of children, and whether the daughter shows a submissive role in sexual relationships.

| Information on terminations

Abortion is either very restricted or banned altogether in all six countries during the sample periods.[10] Research with U.S. data has shown that women who undertake an abortion are of higher socioeconomic status than other pregnant women (Fletcher and Wolfe 2009; Ashcraft et al. 2013). Hence, high abortion rates could compromise the representativeness of our data. In the DHS, women are only asked whether they ever had a pregnancy that terminated in a miscarriage, abortion, or still birth. If they answer affirmatively, the DHS further asks whether the last termination occurred within the previous five years. Due to these limitations, we cannot differentiate between abortion and other reasons for terminations, quantify the number of terminations per woman, or identify women who had a termination during adolescence and are older than 23 years at the time of the interview.

We have reviewed the available information on terminations and believe that underreporting of illegal abortions is not a problem in our sample. There are several reasons for this. First, women in our sample are only asked whether they have ever had a termination, which means they could conceal an illegal abortion without repercussions. Second, termination rates for teenage daughters in our sample (1.97%) and those reported during adolescence by women aged 23 (see PANEL A in Table B.2) are similar to miscarriage rates reported in other studies (Lang and Weinstein 2015, Lang and Nuevo-Chiquero 2012). Third, for sexually active women termination rates are lower during adolescence (compare PANEL A, sixth row with PANEL B in Table B.2), which suggests that most terminations during this period are likely to be miscarriages. Last, the presence of adults during the interview did not affect the rate of reported terminations.

As we do not have data on the mothers' teen pregnancies that ended in terminations, we focus on teen childbearing. Given the low incidence of abortions in our sample, we think that the representativeness of our findings is not compromised.[11]

3 | THE ECONOMETRIC MODEL

3.1 | Nonignorable missing observations

In Figure 1, we provided graphical evidence of nonignorable missingness. It shows that daughters who are not matched to their mothers are close to three times more likely to be teen mothers. Additional evidence of nonignorable missingness is provided in Table 3. Pooling all countries together, we use all pairs for which we have the daughter's information to show differences in means in some of the daughter's characteristics (normalized between -1 and 1) between matched and unmatched pairs. The differences are statistically significant in

[10] Abortion is prohibited altogether (no explicit legal exception) in Dominican Republic and Haiti (all years), and Colombia in 2005. It is allowed: (*i*) only to save the life of the mother in Guatemala; (*ii*) to save the life of the mother and preserve physical health in Bolivia and Peru; (*iii*) to save the life of the mother and preserve physical and mental health in Colombia 2010 and 2015; and (*iv*) in exceptional circumstances such as in case of rape or incest in Bolivia and Colombia 2010 and 2015.

[11] We have run a model of the daughters' teen pregnancy rates (including both live births and terminations) as the dependent variable, with the mother's TCS as the control. The results of this model, available upon request, are virtually identical to those using the daughter's TCS as the dependent variable.

most cases, which is hardly surprising given the number of observations. In some variables, they are revealingly sizable. The size of the difference in TCS, reported in the first line, confirms what we have already shown, at the country level, in Figure 1: the probability of TCS in unmatched pairs is almost three times higher than in matched pairs. We also find sizable differences in whether the household head is one of the daughter's parents and in the daughter's marital status. Another variable with a sizable difference is whether the daughter has ever had sex. Arguably, all these differences are expressions of the same process behind the differences in TCS. In contrast, the mean differences observed in all other variables are much smaller.

Estimation using only matched pairs when the missing process is not random results in sample selection bias. Applying selection equation methods à la Heckman, as in Francesconi and Nicoletti (2006), is not directly applicable to our model due to the occurrence of missing data not only in the mother's and daughter's TCSs but also in their characteristics. We develop a ML procedure in the spirit of the GMM approach first proposed by Ramalho and Smith (2013).[12] Our ML procedure uses all observations and permits the missing process to be endogenous. The procedure also allows us to test whether we can ignore the unmatched pairs when estimating the model, which we refer to as a test of ignorability.

3.2 | The empirical strategy

We define the dependent variable to be the TCS of the daughter, that is $y_i = 1$ if daughter i is either pregnant or a teen mother and $y_i = 0$ otherwise. We model y_i as dependent on her mother's TCS, y_i^m. We assume the following probit specification:

$$\Pr\{y_i = 1 \,|y_i^m, x_i, x_i^m, z_i\} = \Phi\left(\alpha y_i^m + x_i\beta + x_i^m\beta^m + z_i\gamma\right) \tag{1}$$

Control vectors x_i, x_i^m, and z_i are discrete. Vector x_i includes variables that are missing when the daughter is not interviewed. Similarly, x_i^m includes variables that are missing when the mother is not interviewed. Some controls may always be observable. We denote these controls by vector z_i. The aim is to estimate parameter vector $\theta \equiv \{\alpha, \beta, \beta^m, \gamma\}$.

Let $F\left(y_i, y_i^m, x_i, x_i^m, z_i; \theta\right)$ represent $\Pr\left\{y_i \left|y_i^m, x_i, x_i^m, z_i\right.\right\}$. Define I_i as a binary indicator which takes a value of one if the daughter is interviewed and zero otherwise. Similarly, let I_i^m take a value of one if the mother is interviewed and zero otherwise. For an observation with nonmissing information, the joint probability of nonmissingness $\left(I_i = I_i^m = 1\right)$ and the vector variables $\left\{y_i, y_i^m, x_i, x_i^m, z_i\right\}$ is:

$$\begin{aligned}\Pr\left\{I_i = I_i^m = 1, y_i, y_i^m, x_i, x_i^m, z_i\right\} = \; & \Pr\left\{I_i = I_i^m = 1 \left|y_i, y_i^m, x_i, x_i^m, z_i\right.\right\} \times \\ & F\left(y_i, y_i^m, x_i, x_i^m, z_i; \theta\right) \times \Pr\left\{y_i^m, x_i, x_i^m, z_i\right\}.\end{aligned} \tag{2}$$

There are two situations in which a given observation may have missing information: when the mother's information is missing but the daughter's is not and when the daughter's information is missing but the mother's is not. We derive the joint probability of each of these two cases in Appendix A.

[12] Our methodological procedure is closest to Carro, Machado and Mora (2023), which estimates a mother-daughter transmission model of female labor force participation using historical data. On the one hand, in their model, only the labor participation variables for the daughter and the mother can be missing. In contrast, in our model when the TCS of the daughter (mother) is not observed, we allow all her (her mothers') characteristics to be missing. This feature makes our approach more relevant for other applications where missing information stems from noninterviewed subjects. On the other hand, they allow for observations in which both the daughter and the mother's labor force participation are missing.

TABLE 3 Mean tests of daughter characteristics by mother observability

	Mother observed	Mother missing	Normalized Difference	p-value
TCS	0.08	0.22	-0.15	0.0000
Sexual behavior:				
Ever had sex	0.23	0.44	-0.20	0.0000
Number of sex partners	0.37	0.69	-0.03	0.0000
Household formation:				
Ever married	0.05	0.26	-0.20	0.0000
Parent is household head	0.94	0.43	0.51	0.0000
Grandparent is household head	0.06	0.23	-0.18	0.0000
Contraception and gender roles:				
Uses contraception	0.08	0.15	-0.08	0.0000
Fertility knowledge	0.19	0.18	0.01	0.0000
Knows contraceptive method	0.93	0.94	-0.01	0.0000
No condom ok if STD	0.04	0.06	-0.01	0.0000
Sex ok if adultery	0.06	0.09	-0.02	0.0000
Ideal number of children	2.15	2.25	-0.01	0.0000
Demographics:				
Age	16.31	16.61	-0.10	0.0000
Years of education	8.74	8.03	0.05	0.0000
Firstborn	0.31	0.18	0.12	0.0000
Never menstruated	0.01	0.01	0.00	0.8417
Rural	0.37	0.39	-0.02	0.0000
Household size	0.31	0.27	0.04	0.0000
Wealth index	2.81	2.73	0.02	0.0000

Notes: Means tests are two-tail t-tests on the equality of means allowing for unequal variances. 'Normalized difference' refers to the difference of the group means of the variables normalized between -1 and 1. Variable 'TCS' indicates whether the daughter is pregnant or has already had a live birth. 'Ever had sex' indicates whether the daughter has already had sexual intercourse. 'Number of sex partners' is the total lifetime number of sexual partners. 'Ever married' indicates whether daughter has ever been married or lived with her partner. 'Parent is household head' indicates whether one of the daughter's parents is the household head. 'Grandparent is household head' indicates whether one of the daughter's grandparents is the household head. 'Uses contraception' indicates whether the daughter currently uses a modern contraceptive methods. 'Fertility knowledge' takes value one if the daughter correctly states the moment during the menstrual cycle in which a woman is likely most fertile. 'No condom ok if STD' takes value one if the daughter agrees with 'the partner having a sexually transmitted disease is no reason to use a condom'. 'Sex ok if adultery' takes value one if the daughter agrees with 'the partner having a sexual affair with another individual is no reason to refuse sex'. 'Ideal number of children' is the number of children the daughter would have liked to have in her whole life, irrespective of the number she already has. 'Age' is age in years. 'Years of education' is the number of years of complete education. 'Firstborn' indicates that the daughter is the mother's eldest child. 'Never menstruated' indicates that the daughter has never menstruated. 'Rural' takes value one if the daughter's household is located in a rural area. 'Household size' takes value one if the household has at least seven members. 'Wealth index' is the quintile to which the household belongs in the continuous household wealth measure produced by the DHS.

We ensure the identification of θ by restricting the conditional missing process (Ramalho and Smith, 2013). In our case, we make the following assumption:

Assumption 1 *Observability of the mother and the daughter information,* I *and* I^m*, is conditionally independent of* x*,* x^m*, and* z*; i.e.,*

$$\Pr\{I, I^m | y, y^m, x, x^m, z\} = \Pr\{I, I^m | y, y^m\}. \tag{3}$$

Assumption 1 allows for a flexible nonparametric specification of the missing process. Importantly, this assumption deals with non-ignorability by assuming that any correlations between $\{I, I^m\}$ and $\{x, x^m, z\}$ occur through the direct effect of the latter on y and y^m.[13] Crucially, several reasons justify Assumption 1 in our application.

First, as Figure 1 shows, unmatched daughters are much more likely to be teen mothers than matched daughters. This suggests that motherhood in adolescence decreases the probability of a mother-daughter match. Hence, it seems necessary to allow the missing process to depend on the daughter's TCS, i.e., y.

Second, it is likely that the mother's TCS also affects the probability of the mother-daughter match, although the direction of the effect is unclear. On the one hand, children of teenage mothers are often raised by grandparents or other relatives (Card 1981; Manlove 1997). Hence, a mother who was a teen mother might decrease the probability of a match. On the other hand, at least two other factors, which correlate with y^m, affect the probability of a match: adverse economic conditions (because they might discourage the daughter from leaving the parental home) and age of the mother (as women who were teen mothers are more likely to be younger than 49 and therefore interviewed). Hence, it also seems necessary to allow the missing process to depend on the mother's TCS, i.e., y^m. By conditioning the probability on y and y^m jointly, we also allow for interaction effects in how they affect the likelihood of a match.

Third, as shown in Table 3, differences between matched and unmatched pairs are only sizable in variables related to becoming a mother at a young age, which suggests that no additional conditioning in Assumption 1 is necessary. Nonetheless, as a robustness check, we relax Assumption 1 and estimate the model for subsamples based on the age of the daughter and the household wealth in Section 5.

The model parameters are: probit model parameters θ; conditional missing process parameters $\Pr\{I, I^m | y, y^m\}$; and marginal probabilities $\Pi_{y^m,x,x^m,z} \equiv \Pr\{y^m, x, x^m, z\}$. In Appendix A we derive the likelihood function. ML estimation will yield consistent and asymptotically efficient estimates of θ. The number of parameters in $\Pi_{y^m,x,x^m,z}$ grows exponentially with the number of controls and rapidly becomes computationally intractable. In our application, we reduce the number of parameters in $\Pi_{y^m,x,x^m,z} = \Pi_{y^m,x^m|x,z}\Pi_{x,z}$ by assuming that $\Pi_{y^m,x^m|x,z} = \Pi_{y^m,x^m|z}$, i.e., mother characteristics are conditionally independent of those of the daughter. We further reduce the number of parameters by assuming that $\Pi_{y^m,x^m|z} = \Pi_{y^m,v(x^m)|w(z)}$, where $v \subseteq x^m$ and $w \subseteq z$, i.e., the joint probability of y^m and x^m conditional on z only varies along a subset of x^m and a subset of z. For example, in our application for each country, v includes dummies for the age and the education of the mother, and w includes survey time dummies; therefore, the dimension of $\Pi_{y^m,v(x^m)|w(z)}$ does not increase with the number of controls. We avoid the estimation of $\Pi_{y^m,v(x^m)|w(z)}$ (whose potentially large number may risk sample identification) by replacing them with their values obtained from precise external data, referred to as TCRs in section 2.1.

[13] Allowing the conditional probability $\Pr\{I, I^m | y, y^m, x, x^m, z\}$ to additionally depend non-parametrically on a subset of (x, x^m, z) is theoretically feasible but in practice, might compromise sample identification as each additional discrete variable at least doubles the number of parameters related to the missing process.

4 | MAIN RESULTS

Table 4 shows country-specific estimates of Equation (1) under ignorability, i.e, using only observations from matched mother-daughter pairs. In this basic specification, in addition to the mother's TCS, we include survey year dummies and a small set of exogenous dummy variables: age of the daughter, whether the mother is 45 or older, and rural residence. All significant coefficients have the expected sign and direction, i.e., the mother TCS status ($y^m = 1$), older age of the daughter, and rural residence are associated with a higher probability of being a teen mother. For notational simplicity, we denote by AME (from Average Marginal Effect) the estimate of the associated change in the probability that the daughter is a teen mother when the mother's TCS, y^m, changes from zero to one.[14] For each country, the AME is positive and statistically significant. Having a mother who gave birth during her teenage years is associated with an increased probability of the daughter becoming a teenage mother which ranges from 2.0 to 5.0 percentage points (from 31 to 71 percent in relative terms). These AMEs are similar to the probability differences in teen childbearing between daughters whose mothers were and were not teen mothers implicit in Figure 1. The relative range in AMEs ($\frac{5.0-2.0}{2.0} \times 100 = 150\%$) is large, arguably because of country differences across cultural, economic, and legal characteristics, such as abortion legislation, human capital, and gender roles (see footnote 10 and Table B.1 in Appendix B).

We carry out ML estimation of equation (A.1) in Appendix A subject to (A.2) with the same specification as the probit under ignorability. Since we estimate the model with all mother-daughter pairs, including unmatched pairs, we refer to these results as "full sample ML estimates". Table 5 shows the results. Compared to the results under ignorability, the number of observations nearly doubles when we add unmatched pairs to the estimation sample. The average TCS of the daughter, $\bar{y}$, is also larger because unmatched daughters are more likely to be teen mothers, as shown in Figure 1. Turning our attention to the coefficient estimates, those related to y^m and to Mother 45+ are much larger, and both significantly increase the probability of being a teen mother. Additionally, the AME estimates range from 10.2 percentage points in Peru to 26.3 percentage points in the Dominican Republic. Comparing these AME estimates to those of Table 4, they are between 2.7 and 9.4 times larger. The relative range in AMEs ($\frac{26.3-10.2}{10.2} \times 100 = 158\%$) is similar to the range found under ignorability. However, two types of countries emerge: those with relative low AMEs (Peru and Colombia) and those with large AMEs (the remaining countries). This pattern is likely the result of country heterogeneity. For example, Peru and Colombia have higher levels of education, higher indexes of human development, and lower incidence of submissive beliefs on gender roles (Table B.1 in Appendix B).

[14] For the sake of comparability, the AME is always calculated for all observations where the mother's information is available. This sample includes both pairs where the mother and daughter are matched, as well as mothers for whom we cannot establish a match with a daughter.

TABLE 4 Probit estimates under ignorability. Matched mother-daughter pairs only.

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
y^m	0.414***	0.300***	0.289***	0.168**	0.231**	0.275***
	(0.112)	(0.037)	(0.056)	(0.079)	(0.092)	(0.039)
Mother 45+	0.126	0.000	0.063	-0.016	0.077	-0.017
	(0.120)	(0.043)	(0.069)	(0.093)	(0.095)	(0.045)
Age = 16	0.375**	0.281***	0.421***	0.387***	0.290*	0.310***
	(0.176)	(0.056)	(0.089)	(0.132)	(0.161)	(0.061)
Age = 17	0.707***	0.670***	0.698***	0.727***	0.829***	0.660***
	(0.169)	(0.053)	(0.087)	(0.128)	(0.149)	(0.059)
Age = 18	0.794***	0.967***	1.090***	0.978***	1.132***	0.946***
	(0.175)	(0.053)	(0.085)	(0.127)	(0.147)	(0.059)
Rural	0.685***	0.207***	0.094*	0.231***	0.127	0.275***
	(0.111)	(0.039)	(0.056)	(0.082)	(0.095)	(0.039)
AME	0.050***	0.048***	0.042***	0.020**	0.027**	0.033***
	(0.014)	(0.006)	(0.008)	(0.009)	(0.011)	(0.005)
$\overline{y}$	0.070	0.097	0.086	0.064	0.067	0.067
No.obs	1383	10381	4780	2741	2151	11247

Notes: *** p<0.01, ** p<0.05, * p<0.1. Estimation sample obtained from matched mother-daughter pairs from sample DHSs as described in Table 1. Dependent variable is daughter's TCS. All models include survey dummies. Standard errors in parenthesis. Variable y^m is the dummy for the mother's teen childbearing status. 'Mother 45+' is a dummy variable for mother 45 and older. 'Age = 16', 'Age = 17' and 'Age = 18' are age dummies (reference category is age 15). 'Rural' takes value 1 when the household is located in a rural area. Vector x^m includes 'Mother 45+' and 'Rural'. Vector z includes age of the daughter dummies and the survey year dummies. Vector x is empty. AME is the estimated Average Marginal Effect of y^m. $\overline{y}$ is the average value of the daughter's teen childbearing status in the estimating sample.

Under the header 'Ignorability tests', we report likelihood ratio tests for the null hypothesis that $\Pr(I = 1, I^m = 1 | y, y^m)$ are invariant to y^m and y.

TABLE 5 Full sample ML estimates

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
y^m	1.235***	0.589***	1.253***	1.380***	1.195***	0.592***
	(0.200)	(0.046)	(0.059)	(0.081)	(0.154)	(0.053)
Mother 45+	0.787***	0.318***	0.842***	1.208***	0.468***	0.409***
	(0.105)	(0.041)	(0.050)	(0.060)	(0.083)	(0.043)
Age = 16	0.387***	0.407***	0.389***	0.351***	0.372***	0.366***
	(0.121)	(0.040)	(0.057)	(0.090)	(0.108)	(0.043)
Age = 17	0.851***	0.818***	0.829***	0.836***	0.877***	0.826***
	(0.113)	(0.038)	(0.054)	(0.086)	(0.102)	(0.041)
Age = 18	1.101***	1.152***	1.156***	1.062***	1.299***	1.152***
	(0.114)	(0.038)	(0.054)	(0.085)	(0.101)	(0.041)
Rural	0.814***	0.550***	0.275***	-0.000	-0.048	0.446***
	(0.110)	(0.035)	(0.050)	(0.068)	(0.084)	(0.039)
AME	0.213***	0.128***	0.263***	0.188***	0.193***	0.102***
	(0.018)	(0.006)	(0.006)	(0.006)	(0.010)	(0.006)
Ignorability test	204.678	4528.435	2814.760	3364.959	180.521	5170.959
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\overline{y}$	0.155	0.166	0.192	0.162	0.096	0.121
N. obs.	2982	19723	10793	4845	5675	20990

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis. Maximum Likelihood estimates of (A.1) subject to (A.2). Estimation sample obtained from all mother-daughter pairs (matched and unmatched) from sample DHSs as described in Table 1. Dependent variable is daughter's TCS. All models include survey dummies. Variable y^m is the dummy for the mother's teen childbearing status. 'Mother 45+' is a dummy variable for mother 45 and older. 'Age = 16', 'Age = 17', and 'Age = 18' are age of the daughter dummies (reference category is age 15). 'Rural' takes value 1 when the mother's household is located in a rural area. Vector x^m includes 'Mother 45+' and 'Rural'—$v(x^m)$ is 'Mother 45+'. Vector z includes age of the daughter dummies and the survey year dummies, the latter being $w(z)$. Vector x is empty. AME is the estimated Average Marginal Effect of y^m. Ignorability tests are likelihood ratio tests for the null hypothesis that conditional probabilities $\Pr(I = 1, I^m = 1|y, y^m)$ are invariant to the mother's TCS and the daughter's TCS, y^m and y, respectively, their p-values are in parenthesis. $\overline{y}$ is the average value of the daughter's teen childbearing status in the estimating sample.

We strongly reject the null in all countries, which shows that unmatched pairs are not ignorable. We conclude that ignoring the missing observations leads to a substantial negative selection bias in AMEs of the mother's TCS.

Because mothers' TCS may be correlated with relevant omitted variables, we enrich our basic specification by including more mother characteristics as controls in our estimation: (*i*) a dummy for whether the mother has no more than primary education (At most primary); (*ii*) a dummy for whether the mother's household size has at least seven members (Household size); (*iii*) a dummy for whether the mother's household belongs to the two poorest quintiles in the country (Poor); (*iv*) and a dummy that indicates that the mother's eldest child is the teenage daughter (Firstborn). The full sample ML estimates are shown in Table 7.

Generally, an increase in any of these additional controls results in a significant increase in the probability of teen motherhood. Regarding the AMEs, including the additional controls does not substantially change the results. In all countries, the new AMEs are still very large and statistically significant—especially when compared to the AMEs obtained under ignorability (reported in the bottom line of Table 7). They range from 9.1 percentage points in Peru to 23.7 percentage points in Dominican Republic (a relative range of 160%). Relative to the average probability of teen motherhood, ratios range from 75.2 percent in Peru to 187.5 percent in Haiti.

The comparison of Tables 4, 5, and 7 highlights the size of co-residential bias when only using the mother-daughter matched pairs. The comparison also suggests that omitted variable bias, if it exists, is an order of magnitude smaller than co-residential bias. If the problem of omitted variable bias were significant, the addition of maternal personal characteristics strongly associated with socioeconomic background should substantially vary the AMEs. However, the AMEs remain similar (Tables 5 and 7).

Table 6 reports full sample ML estimates of the probability of having a mother-daughter match, i.e., $\Pr(I = 1, I^m = 1|y, y^m)$, as well as the ignorability test.

TABLE 6 Missing process estimates & ignorability tests. Full sample ML estimation

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
$\Pr(I = 1, I^m = 1\|y = 0, y^m = 0)$	0.404	0.453	0.372	0.473	0.316	0.453
$\Pr(I = 1, I^m = 1\|y = 0, y^m = 1)$	0.857	0.999	0.878	0.951	0.675	0.928
$\Pr(I = 1, I^m = 1\|y = 1, y^m = 0)$	0.230	0.229	0.295	0.427	0.410	0.237
$\Pr(I = 1, I^m = 1\|y = 1, y^m = 1)$	0.205	0.364	0.170	0.169	0.181	0.404
Ignorability tests	201.724	24438.551	795.435	3479.408	179.483	5965.230
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: ML estimates for the probabilities of the missing process using the full sample ML estimates with additional controls reported in Table 7. Ignorability tests are likelihood ratio tests for the null hypothesis that conditional probabilities $\Pr(I = 1, I^m = 1|y, y^m)$ are invariant to the mother's TCS and the daughter's TCS, y^m and y, respectively, their p-values are in parenthesis.

The estimates vary considerably by y and y^m, showing that, as discussed in Section 3.2, y and y^m affect the probability of the match. Moreover, we can strongly reject the null hypothesis of ignorability in all countries. Three features stand out in Table 6. First, when the mother is not a teen mother, $y^m = 0$, in all cases except Haiti, the probability of a match decreases when daughters are teen mothers.

TABLE 7 Full sample ML estimates. Additional controls

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
y^m	1.186***	0.652***	1.233***	1.480***	1.169***	0.574***
	(0.215)	(0.046)	(0.091)	(0.082)	(0.150)	(0.059)
Mother 45+	0.750***	0.227***	0.693***	1.262***	0.453***	0.359***
	(0.118)	(0.044)	(0.058)	(0.063)	(0.088)	(0.051)
At most primary	0.195	0.038	0.400***	0.216*	0.421***	-0.063
	(0.159)	(0.038)	(0.065)	(0.125)	(0.140)	(0.052)
Age = 16	0.379***	0.321***	0.411***	0.329***	0.365***	0.365***
	(0.120)	(0.037)	(0.059)	(0.092)	(0.109)	(0.044)
Age = 17	0.835***	0.717***	0.847***	0.823***	0.872***	0.831***
	(0.115)	(0.036)	(0.056)	(0.088)	(0.104)	(0.042)
Age = 18	1.036***	1.040***	1.175***	1.027***	1.279***	1.140***
	(0.118)	(0.035)	(0.057)	(0.087)	(0.102)	(0.042)
Rural	0.122	0.176***	0.123**	-0.038	-0.074	0.054
	(0.192)	(0.044)	(0.054)	(0.073)	(0.112)	(0.064)
Household size	0.296***	0.570***	0.636***	0.123*	0.187**	0.619***
	(0.114)	(0.035)	(0.053)	(0.071)	(0.086)	(0.040)
Poor	0.688***	0.147***	0.055	0.128*	0.082	0.370***
	(0.194)	(0.044)	(0.057)	(0.074)	(0.106)	(0.071)
Firstborn	0.583***	0.247***	0.593***	0.416***	0.315***	0.371***
	(0.131)	(0.039)	(0.054)	(0.078)	(0.097)	(0.050)
N.obs	2982	19723	10793	4845	5675	20990
AME	0.208***	0.149***	0.237***	0.194***	0.180***	0.091***
	(0.020)	(0.006)	(0.009)	(0.006)	(0.009)	(0.006)
$\overline{y}$	0.155	0.166	0.192	0.162	0.096	0.121
AME under	0.040***	0.030***	0.018**	0.013	0.023**	0.019***
ignorability	(0.014)	(0.006)	(0.008)	(0.009)	(0.011)	(0.005)

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis. Maximum Likelihood estimates of (A.1) subject to (A.2). Estimation sample obtained from all mother-daughter pairs (matched and unmatched) from sample DHSs as described in Table 1. Dependent variable is daughter's TCS. All models include survey dummies.Variable y^m is the dummy for the mother's teen chilbearing status. 'Mother 45+' is a dummy variable for mother 45 and older. 'At most primary' is a dummy variable that takes value 1 if the mother has no more than primary education. 'Age = 16', 'Age =17', and 'Age = 18' are age dummies (reference category is age 15). 'Rural' takes value 1 when the mother's household is located in a rural area. 'Household size' takes value one when the mother's household has at least seven members. 'Poor' takes value 1 if the mother's household belongs to the two poorest quintiles in the country based on a continuous wealth measure produced by the *DHS*. Dummy variable 'Firstborn', obtained from the mother's birth history data, indicates that the mother's eldest child is the teenage daughter. Vector x^m includes 'Mother 45+', 'Rural', 'At most primary', 'Household size', 'Poor', and 'Firstborn'—$v(x^m)$ is 'Mother 45+' and 'At most primary'. Vector z includes age of the daughter dummies and the survey year dummies, the latter being $w(z)$. Vector x is empty. AME is the estimated Average Marginal Effect of y^m. $\overline{y}$ is the average value of y in the estimating sample. AME under ignorability stands for the average marginal effect of y^m using the probit estimates under ignorability for the same variable specification.

This strongly suggests that teen motherhood increases the probability that the daughter leaves the parental home to live with her partner. Second, when the mother is a teen mother and the daughter is not, it's highly probable that both mother and daughter are observed together—the likelihood ranges from 67.5 to 99.9%. This is partly because teen mothers tend to be younger, hence more likely to be interviewed. However, if the daughter is a teen mother, the probability of them being observed together dramatically decreases. This further underscores that teen motherhood tends to increase the likelihood of the daughter leaving the parental home, with the probability of a match now falling to a range of 16.9 to 40.4%. Third, this drop is smaller in Peru and Colombia, where the probability of a match is around double of the probability of the other countries (40.4 and 36.4% vs around 18%). Not surprisingly, for these two countries the AMEs estimates are closer to those obtained under ignorability.

| Interpretation of AMEs

In the Introduction, we emphasized the importance of estimating the intergenerational transmission of TCS. This remains true even if there is no causal relationship between being born to a teen mother and the likelihood of becoming one. However, there are compelling reasons to seek out the identification of this causal effect. Thus far, we have addressed coresidential bias and have controlled for mother's characteristics associated with her socioeconomic background. However, this may not be enough to ensure a causal interpretation of the average marginal effects (AMEs) because the mother's TCS may remain endogenous. For instance, if a teenager expects to be poor as an adult, she may be more likely to become a teen mother, and her poverty as an adult may persist irrespective of whether she has a child as a teenager. Furthermore, if her poverty affects her daughter's expectations of future poverty and increases her likelihood of becoming a teen mother, we observe intergenerational persistence. Controlling for expectations is challenging. While the DHS data lets us include daughter characteristics that are presumably strongly linked with her expectations, none of the available variables specific to the daughter serves as both a reliable predictor of her expectations and an exogenous control in the model.

So far, we included the age of the daughter and whether she is the firstborn child of her mother.[15] Other variables, such as her education or attitudes and beliefs about sex, are likely predictors of her expectations and simultaneous to her TCS and, consequently, endogenous. Nonetheless, to assess the potential extent of omitted variable bias it is interesting to see how the inclusion of these potentially endogenous characteristics changes the estimates of AME.

In Table 8, we present the results obtained with the full sample ML approach when adding four characteristics of the daughter, one at a time. They are the following: (*i*) 'At most primary' takes a value of one if the daughter has no more than primary education and zero otherwise; (*ii*) 'Submissive gender role' takes value one if the daughter agrees with "the partner having a sexually transmitted disease is no reason to use a condom" or "the partner having a sexual affair with another individual is no reason to refuse sex", or both; (*iii*) 'Fertility knowledge' takes value one if the daughter correctly states the moment during the menstrual cycle in which a woman is likely most fertile and zero otherwise; and (*iv*) 'Fertility preferences' takes value one if the daughter declares her ideal number of children to be larger than 2 and zero otherwise.

[15] Strictly speaking, 'Firstborn', although a characteristic of the daughter, is drawn from the birth history of mothers. Hence, it is only observed when the mother is observed.

TABLE 8 The role of potentially endogenous controls. Full sample ML estimates

Panel A: At most primary

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
$\widehat{\beta}_{\text{At most primary}}$	0.836***	0.928***	0.893***	0.526***	0.757***	0.820***
	(0.079)	(0.036)	(0.039)	(0.063)	(0.064)	(0.036)
AME	0.101***	0.106***	0.135***	0.185***	0.106***	0.095***
	(0.018)	(0.005)	(0.009)	(0.006)	(0.010)	(0.005)

Panel B: Submissive gender role

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
$\widehat{\beta}_{\text{Submissive gender role}}$	0.193*	-0.031	-0.102	-0.018	0.072	-0.167***
	(0.110)	(0.054)	(0.065)	(0.070)	(0.072)	(0.051)
AME	0.188***	0.103***	0.217***	0.199***	0.191***	0.090***
	(0.020)	(0.006)	(0.008)	(0.006)	(0.009)	(0.005)

Panel C: Fertility knowledge

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
$\widehat{\beta}_{\text{Fertility knowledge}}$	0.022	-0.114***	0.022	-0.013	-0.017	-0.132***
	(0.080)	(0.030)	(0.070)	(0.080)	(0.076)	(0.033)
AME	0.176***	0.098***	0.231***	0.198***	0.177***	0.082***
	(0.020)	(0.006)	(0.008)	(0.006)	(0.009)	(0.006)

Panel D: Fertility preferences

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
$\widehat{\beta}_{\text{Fertility preferences}}$	0.033	0.039	0.088	0.074	-0.133*	0.055
	(0.134)	(0.067)	(0.054)	(0.067)	(0.078)	(0.058)
AME	0.203***	0.105***	0.223***	0.199***	0.174***	0.088***
	(0.019)	(0.006)	(0.009)	(0.006)	(0.009)	(0.006)

Panel E: AME without potentially endogenous controls (from Table 7)

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
AME	0.208***	0.149***	0.237***	0.194***	0.180***	0.091***
	(0.020)	(0.006)	(0.009)	(0.006)	(0.009)	(0.006)

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors in parenthesis. Maximum Likelihood estimates of (A.1) subject to (A.2). Estimation sample obtained from all mother-daughter pairs (matched and unmatched) from sample DHSs as described in Table 1. All models use the same specification as in Table 7 adding one additional dummy variable as x. 'At most primary' takes value one if the daughter has no more than primary education. 'Submissive gender role' takes value one if the daughter agrees with "the partner having a sexually transmitted disease is no reason to use a condom" or "the partner having a sexual affair with another individual is no reason to refuse sex", or both. 'Fertility knowledge' takes value one if the daughter correctly states the moment during the menstrual cycle in which a woman is likely most fertile. 'Fertility preferences' takes value one if the daughter would have liked to have more than two children in her whole life, irrespective of the number she already has. $\widehat{\beta}_x$ reports the ML estimate of the parameter associated to x. AME is the estimated Average Marginal Effect of y^m.

Adding these controls does not cause the AME estimate to disappear. The new AMEs are generally similar to those obtained in Table 7, except for 'At most primary'. Thus, if there are endogeneity problems in y^m due to omitted variables, such variables are unlikely to be related to gender roles, reproductive health knowledge, or fertility preferences. In the case of the inclusion of the control 'At most primary', there are no substantial changes in the average marginal effects in Peru and Guatemala. This similarity strongly suggests that our estimates presented in Table 7 for these two countries are likely to closely represent the actual causal effects. In contrast, in the other four countries, it cannot be ruled out that omitted variables are related to 'At most primary' because the new estimate of the AME is lower. However, it is difficult to determine whether this decrease is due to a downward bias introduced by the endogeneity of 'At most primary' or because adding 'At most primary' actually corrects an original upwards omitted variable bias. Nevertheless, even if we assume that the new AMEs are the causal effects, the estimates are still considerably higher than those obtained under ignorability in Table 4.[16]

5 | EXTENSIONS

5.1 | Heterogeneity

In this section, we estimate the model using different subsamples defined by the age of the daughter and the wealth of the household. Our motivation for this exercise is twofold. First, estimating the model on different subsamples defined by the daughter's age or the household's wealth relaxes Assumption 1 as it allows the missing process to also be conditionally dependent on these variables. Second, the parameter α, may be heterogeneous. A way to allow for this heterogeneity is by estimating the model on different subsamples.[17] The estimation by country presented in Section 4 was already driven by these motivations.

5.1.1 | Age of the daughter

In the estimations presented thus far, we allow the age of the daughter to affect the probability of being a teen mother, i.e., it enters as a set of dummies in (1). The results in Section 4 show that the probability of teen motherhood increases significantly with the daughter's age. Given its range, from 15 to 18 years old, this result is not surprising.

Separating the results for 15 to 17-year-olds from those for 18-year-olds could provide valuable insights.[18] An 18-year-old daughter is likely to have a different living situation since she is considered an adult under the law and more likely to be past adolescence. In contrast, younger daughters are minors and more likely to live with their mothers. Additionally, using age dummies alone may not account for the issue of right censoring that affects the under-18 group. Presenting results separately for 18-year-olds eliminates the issue of right censoring in that particular sample and allows us to estimate different AME by age group.

In Table 9, we show the results for the two age subsamples. The AMEs for the whole sample in Table 7 fall within the estimates shown in Table 9 with the two subsamples. They are closer to those in the 15 to 17-year-olds subsample possibly because this subsample is

[16]Notably, when potentially endogenous controls are added to the probits under ignorability, the AME estimates drop slightly, but the changes are not statistically significant.

[17]Even when the estimated α's are similar, the AMEs may differ, as other characteristics—x, x^m, and z—may also vary by subsample.

[18]We always observe the age of the daughter because, when the daughter's information is missing, we can retrieve it from the mothers' birth histories.

TABLE 9 Full sample ML estimates. Subsamples by age.

	Bolivia		Colombia		DR		Guatemala		Haiti		Peru	
	15-17	18	15-17	18	15-17	18	15-17	18	15-17	18	15-17	18
y^m	0.729**	1.547***	0.407***	0.824***	1.310***	1.197***	1.642***	1.149***	1.414***	1.999***	0.540***	0.925***
	(0.294)	(0.197)	(0.077)	(0.072)	(0.112)	(0.103)	(0.110)	(0.128)	(0.210)	(0.210)	(0.081)	(0.076)
	[0.15,1.31]	[1.16,1.93]	[0.26,0.56]	[0.68,0.96]	[1.09,1.53]	[0.99,1.40]	[1.43,1.86]	[0.90,1.40]	[1.00,1.83]	[1.59,2.41]	[0.38,0.70]	[0.78,1.07]
AME	0.108***	0.357***	0.070***	0.261***	0.212***	0.361***	0.169***	0.270***	0.158***	0.550***	0.089***	0.262***
	(0.024)	(0.020)	(0.009)	(0.011)	(0.009)	(0.013)	(0.007)	(0.015)	(0.009)	(0.013)	(0.009)	(0.011)
	[0.06,0.16]	[0.32,0.40]	[0.05,0.09]	[0.24,0.28]	[0.19,0.23]	[0.34,0.39]	[0.16,0.18]	[0.24,0.30]	[0.14,0.17]	[0.52,0.58]	[0.07,0.11]	[0.24,0.28]
AME under	0.032**	0.072*	0.026***	0.047**	0.013*	0.049*	0.019**	-0.022	0.014	0.049	0.019***	0.017
ignorability	(0.014)	(0.037)	(0.006)	(0.019)	(0.008)	(0.027)	(0.009)	(0.030)	(0.010)	(0.037)	(0.005)	(0.016)
	[0.00,0.06]	[-0.00,0.15]	[0.01,0.04]	[0.01,0.08]	[-0.00,0.03]	[-0.00,0.10]	[0.00,0.04]	[-0.08,0.04]	[-0.01,0.03]	[-0.02,0.12]	[0.01,0.03]	[-0.01,0.05]
$\bar{y}$	0.122	0.256	0.122	0.308	0.146	0.341	0.122	0.289	0.063	0.202	0.087	0.239
N.obs	2258	724	14993	4730	8252	2541	3699	1146	4307	1368	16400	4590

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis. 95 percent confidence intervals in brackets. Maximum Likelihood estimates of (A.1) subject to (A.2). Estimation sample obtained from all mother-daughter pairs (matched and unmatched) from sample DHSs as described in Table 1. Dependent variable is daughter's TCS. "15-17" refers to the subsample of daughters aged between 15 and 17. "18" refers to the subsample of daughters aged 18. All models include survey dummies, a dummy variable for mother 45 and older, a dummy variable that takes value 1 if the mother has no more than primary education, a dummy that takes value 1 when the household is located in a rural area, a dummy variable for parents households with at least seven members, and a dummy variable that indicates whether the daughter is the eldest child of the mother. For the 15-17 subsamples age dummies of the daughter are also included. Variable y^m is the dummy for the mother's teen chilbearing status. AME is the estimated Average Marginal Effect of y^m. $\bar{y}$ is the average value of y in the estimating sample.

larger. The estimated AMEs for daughters aged 18 are between 1.6 and 3.7 times greater than for those aged 15 to 17. Nevertheless, as a higher proportion of the 18-year-old daughters are already mothers (an average 2.5 times higher), the AMEs relative to the mean $\overline{y}$ increase by only around 10%. Another difference with respect to previous results is the varying magnitude of the ignorability bias. For the 18-year-olds, the bias is, on average, 11.4% higher than that obtained with the 15-17 year-old subsample. These larger biases can be attributed to a higher proportion of unmatched mother-daughter pairs among the 18-year-old daughters. This is because 18-year-olds are less likely to live with their mothers and more likely to have become mothers themselves, compared to younger daughters.[19] Moreover, large AMEs among the 18-year-old daughters, for whom right censoring is not an issue, imply that obtaining large estimates with the full sample ML estimator cannot be driven by right censoring.

Regarding the estimation of the missing process parameters (results available upon request), the pattern found for both subsamples broadly replicate the results for the pool sample presented in Table 6.

Overall, although the results are sensitive to age, the main conclusions remain valid for both age groups: the AMEs are large and statistically significant, and the ignorability bias is still sizable, especially for the 18-year-old daughters.

5.1.2 | Wealth

The probability of observing a matched pair may depend on the socioeconomic conditions of the mothers' household, conditional on y and y^m. For example, in a poorer family, the daughter may lack the financial support to become independent, making the mother-daughter match more likely. Alternatively, one of the two women may be forced to migrate in search of a job, resulting in an unmatched pair.

Similar to age, wealth may also affect the level of inertia, i.e., the parameter α. For example, daughters may have a greater probability of becoming adolescent mothers in areas with a high prevalence of teen motherhood. Since low-income families live in segregated areas where teen motherhood is more frequent, we would expect α to be higher for this group.

Ideally, we want to stratify the sample by the level of wealth in the household where the daughter grew up. However, this information is not available. In the case of a mother-daughter match or when the daughter is missing, we observe the wealth of the mother's current household, not necessarily the household where the daughter grew up. When the mother's information is missing, we observe only the wealth of the daughter's household, and if the daughter is married or cohabits with her partner, most likely she does not live in the home where she grew up. Despite these caveats, we proxy the wealth of the household where the daughter grew up by the observed current wealth of the household. Then, we use it to split the sample between those belonging to the two poorest quintiles in the country and those belonging to the two richest quintiles.[20]

Table 10 reports the results by country and wealth. As expected, the incidence of teen childbearing, $\overline{y}$, is higher among poorer households in all countries.

The confidence intervals of $\widehat{\alpha}$ and AME generally overlap across countries for the rich samples.

[19] The proportion of 18-year-old daughters who are matched with their mothers is, on average, only 39%, compared to 49% in the whole sample and 51% in the 15- to 17-year-old subsample.

[20] When we estimate the model only with the households of the richest quintile, the results become unstable in the countries with the smallest samples and there are issues of sample identification. We have done the same exercise re-defining nonpoor households as those in the 3rd and 4th quintile and the results are similar.

TABLE 10 Full sample ML estimates. Subsamples by household wealth

	Bolivia		Colombia		DR		Guatemala		Haiti		Peru	
	Poor	Rich	Poor	Rich	Poor	Rich	Poor	Rich	Poor	Rich	Poor	Rich
y^m	1.279***	1.008***	0.168**	1.092***	0.903***	0.623***	1.586***	1.358***	1.066***	1.471***	0.250***	1.145***
	(0.154)	(0.359)	(0.066)	(0.076)	(0.075)	(0.212)	(0.101)	(0.099)	(0.206)	(0.219)	(0.077)	(0.072)
	[0.98,1.58]	[0.30,1.71]	[0.04,0.30]	[0.94,1.24]	[0.76,1.05]	[0.21,1.04]	[1.39,1.78]	[1.16,1.55]	[0.66,1.47]	[1.04,1.90]	[0.10,0.40]	[1.00,1.29]
AME	0.300***	0.122***	0.039***	0.173***	0.210***	0.085***	0.228***	0.161***	0.169***	0.188***	0.044***	0.136***
	(0.015)	(0.026)	(0.008)	(0.009)	(0.008)	(0.016)	(0.008)	(0.008)	(0.015)	(0.009)	(0.008)	(0.006)
	[0.27,0.33]	[0.07,0.17]	[0.02,0.06]	[0.16,0.19]	[0.19,0.23]	[0.05,0.12]	[0.21,0.24]	[0.15,0.18]	[0.14,0.20]	[0.17,0.21]	[0.03,0.06]	[0.12,0.15]
AME under	0.108***	0.012	0.034***	0.026***	0.020	0.015	0.020	0.007	0.012	0.033**	0.025***	0.014***
ignorability	(0.033)	(0.013)	(0.009)	(0.008)	(0.014)	(0.010)	(0.017)	(0.011)	(0.018)	(0.014)	(0.008)	(0.005)
	[0.04,0.17]	[-0.01,0.04]	[0.02,0.05]	[0.01,0.04]	[-0.01,0.05]	[-0.00,0.03]	[-0.01,0.05]	[-0.01,0.03]	[-0.02,0.05]	[0.01,0.06]	[0.01,0.04]	[0.00,0.02]
$\bar{y}$	0.249	0.105	0.220	0.104	0.284	0.111	0.216	0.125	0.118	0.080	0.176	0.077
N.obs.	1041	1941	10533	9190	5065	5728	1973	2872	2437	3238	9349	11641

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis. 95 percent confidence intervals in brackets. Maximum Likelihood estimates of (A.1) subject to (A.2). Estimation sample obtained from all mother-daughter pairs (matched and unmatched) from sample DHSs as described in Table 1. Dependent variable is daughter's TCS. "Poor" refers to the subsample of households belonging to the two poorest quintiles of the country's population based on a continuous wealth measure produced by the *DHS*. "Rich" refers to the remaining subsample of households. All models include survey and age dummies, a dummy variable for mother 45 and older, a dummy variable that takes value 1 if the mother has no more than primary education, a dummy that takes value 1 when the household is located in a rural area, a dummy variable for parents households with at least seven members, and a dummy variable that indicates whether the daughter is the eldest child of the mother. Variable y^m is the dummy for the mother's teen chilbearing status. AME is the estimated Average Marginal Effect of y^m. $\bar{y}$ is the average value of y in the estimating sample.

Among the poor samples in Bolivia, DR, Guatemala, and Haiti, AMEs are generally larger than those of the rich samples, which can be attributed to the higher incidence of teen mothers and daughters. However, in Colombia and Peru, the AMEs of the poor are markedly lower than those of the full and rich samples. This is because, as shown for the whole sample in Table 6, the probability of matched pairs in Peru and Colombia when both the mother and daughter are teen mothers is large relative to the proportion of teen motherhood in the population, and the percentage of matched pairs is relatively high compared to other countries in our sample. These features are also found in the subsample of poor households. As a result, the results produced by the ML approach do not differ much from those obtained from matched pairs alone (i.e., under ignorability), as observed in Table 10 for the poor samples in Colombia and Peru.[21]

In summary, direct examination of the role of household wealth during the daughter's upbringing is not possible due to the lack of retrospective information. While some of the estimated α's and AMEs may be sensitive to wealth, the main conclusions drawn in Section 4 remain valid when current household wealth is used as a proxy for retrospective household wealth: (i) the AME is always positive, large, and statistically significant in all countries and wealth groups, and (ii) the bias is large in all subsamples except for the poor households in Colombia and Peru, where results under ignorability are similar.

5.2 | Other outcomes

Teen childbearing does not occur in isolation; it is typically correlated with other outcomes arising from the same underlying process. In section 3.1, we highlighted notable differences between teenagers living with their mothers and those who are not. These differences extend beyond TCS, affecting rates of early sexual initiation and the prevalence of marriage or cohabitation during adolescence. Each of these variables significantly impacts the probability of a mother being interviewed when her daughter is also interviewed (see Table B.3 in Appendix B). We consider four related outcomes. The first two outcomes examine whether the daughter had sex before the age of 16 and the daughter's TCS itself. The third outcome is a binary variable indicating whether the daughter was married before the age of 19. Last, the fourth outcome is a school dropout indicator of whether the daughter has no more than primary education. The first three outcomes exhibit strong correlations, with most correlations exceeding 0.6. The correlations between the first three outcomes and 'At most primary' are relatively weaker, ranging from 0.2 to 0.4, particularly for Haiti. Collectively, these findings strongly suggest a sequence of interrelated events, all shaped by the mother's TCS.

Considering that all outcomes, including 'At most primary', likely arise from the same underlying process, Assumption 1 should hold, enabling us to use the same modeling strategy. Therefore, the full-sample ML estimator remains consistent, even if the outcome variable for the daughter is not TCS. Table 11 presents the average marginal effect of the mother's TCS on these four related outcomes.

[21] It could be argued that one potential explanation for the low AMEs in Colombia and Peru among the poor is that the use of current household wealth as our proxy for household wealth during a daughter's childhood may lead to inaccuracies in classifying households as "high-wealth" or "low-wealth" in these countries as they experience higher levels of intergenerational mobility and economic opportunities (Berniell et al., 2021, Cuesta et al., 2011, Ferreira and Gignoux, 2011). However, when we define the poor as those in the first quintile and the rich as those of the fifth quintile, results are similar.

TABLE 11 Other outcomes. Full sample ML estimates

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
Panel A: Sex before 16						
AME	0.301***	0.102***	0.236***	0.244***	0.345***	0.094***
	(0.013)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)
$\overline{y}$	0.220	0.385	0.296	0.206	0.302	0.196
N.obs.	2980	19721	10767	4843	5674	20990
Panel B: TCS						
AME	0.208***	0.099***	0.237***	0.194***	0.180***	0.091***
	(0.020)	(0.007)	(0.009)	(0.006)	(0.009)	(0.006)
$\overline{y}$	0.155	0.166	0.192	0.162	0.096	0.121
N.obs.	2982	19723	10793	4845	5675	20990
Panel C: Ever married						
AME	0.224***	0.084***	0.221***	0.191***	0.157***	0.103***
	(0.018)	(0.006)	(0.006)	(0.006)	(0.008)	(0.005)
$\overline{y}$	0.135	0.151	0.257	0.179	0.077	0.112
N.obs.	2982	19723	10793	4845	5675	20990
Panel D: At most primary						
AME	0.126**	0.180***	0.302***	0.200***	0.234***	0.059***
	(0.056)	(0.005)	(0.007)	(0.010)	(0.016)	(0.011)
$\overline{y}$	0.313	0.116	0.420	0.463	0.464	0.124
N.obs.	2982	19723	10793	4845	5675	20990

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis. Estimation sample obtained from all mother-daughter pairs (matched and unmatched) from sample DHSs as described in Table 1. Maximum Likelihood estimates of (A.1) subject to (A.2). Each panel refers to the estimation of the model with a different dependent variable. 'Sex before 16' is a dummy variable that takes value one if the daughter reports to have had sex before age 16 and zero otherwise. 'TCS' is the daughter's TCS. 'Ever married' indicates whether daughter has ever been married or lived with her partner. 'At most primary' takes value one if the daughter has no more than primary education. All models use the same control specification as in Table 7. AME is the estimated Marginal Effect of the mother's TCS. $\overline{y}$ is the average value of y in the estimating sample.

All marginal effects exhibit large, positive, and statistically significant values.[22] The magnitudes, in absolute terms, resemble each other across the different outcomes and align with the patterns discussed in Table 7. The marginal effects range from 27% to 204% of the average incidence of each outcome. These findings can be interpreted in at least two ways. First, the mother's TCS may contribute to the daughter's early exposure to risky sexual activity, resulting in unintended consequences such as adolescent childbearing, early marriage, and school dropout. Alternatively, it is possible that the mother's influence on the daughter shapes values and expectations, leading to a stronger preference for early marriage and school dropout. Once the daughter marries before the age of 19 and discontinues her education, early sexual activity and adolescent motherhood become logical consequences.

Regardless of the interpretation one favors, the consistency of signs, magnitudes, and statistical significance of the AMEs strengthens the hypothesis that all the outcomes share a common driver. This driver could be the mother's TCS itself or a factor highly correlated with it, as we discussed in detail at the end of Section 4.

6 | CONCLUSIONS

Latin America has experienced high rates of teen childbearing for decades. Using DHS comparable data from six countries in the region—Bolivia, Colombia, Dominican Republic, Guatemala, Haiti, and Peru—we study the existence of a mother-daughter association in teen childbearing. Regrettably, in the DHS, a significant portion of mother-daughter pairs (49.7 percent) cannot be matched. If the process of matching was independent of teen childbearing, estimation using only matched pairs would not introduce a coresidential bias. Yet, the assumption of independence is questionable, as teenagers who become pregnant and subsequently mothers are arguably more likely to leave their parental home. Consistent with this, our data shows that unmatched daughters are nearly three times more likely to be mothers compared to those who are matched.

Certain questionnaires avoid the need to construct mother-daughter pairs by soliciting information about the respondents' mothers. For instance, the US National Survey of Family Growth (NSFG) queries the age of the respondent's mother at her first childbirth (Kahn and Anderson 1992). This approach depends on the completeness and accuracy of the retrospective data provided by the child.[23] Longitudinal surveys such as the Panel Survey of Income Dynamics (PSID), where the children of the original respondents are followed regardless of their residence, provide alternatives to cross-sectional surveys. However, they potentially face attrition issues, particularly with teen childbearing.[24] Another option, administrative data (as employed in Aizer et al. 2020, for Norway), is primarily available in countries where teen childbearing incidence is low. In the absence of these three alternatives, a methodology that mitigates coresidential bias from household surveys or census data becomes necessary.

[22]Using the National Longitudinal Survey of Youth of the US, Levine et al. (2001) find that being born to a teen mother has a negative association to early sexual initiation, skipping school, and fighting behaviors, even after controlling for a rich specification of background controls.

[23]The information on the mother's age at first birth might be subject to nonignorable missing answers or contain measurement errors due to several reasons. First, children of teen mothers are less likely to live with their biological parents during adolescence due to high divorce rates among teen mothers (Eshbaugh 2008). Second, they are more likely to reside with alternative guardians (Card 1981; Manlove 1997).

[24]Smaller longitudinal datasets that track mothers and children over extended periods also exist. For instance, Barber (2001) provides an example of such a dataset, comprising married mothers who gave birth in Detroit in July 1961. Despite the long timespan, attrition is relatively low, with only 15% of mothers and 18% of children no longer participating. However, the small sample size (427 daughters) constrains the potential for generalizing the results.

We propose a maximum likelihood estimator that mitigates coresidential bias. Employing this method, we show that limiting the sample to matched mother-daughter pairs results in a substantial negative coresidential bias. Our findings provide substantial evidence of the intergenerational transmission of teen childbearing from mothers to daughters. We observe that daughters of teen mothers face a significantly increased risk of teen childbearing, ranging from 9.1 (in Peru) to 23.7 (in Dominican Republic), or equivalently, between 75 and 123% relative to their average incidence of teen childbearing. This high level of intergenerational transmission is at the core of the persistence of elevated teenage childbearing rates in Latin America. As a result, our findings suggest supplementing policies aimed at teenagers with strategies targeting those most at risk—specifically, the daughters of teen mothers.

Our paper relates to studies on the determinants of being a teen mother. This literature has identified multiple causes, from the distribution of parental income during childhood and adolescence to compulsory education (Carneiro et al. 2021, Chetty et al. 2011, Black et al. 2008, and An et al., 1993). Recent studies based on Latin American countries reveal associations between poverty, poor family structure, low educational inputs, low aspirational objectives, low sexual literacy, poor neighborhood, high levels of violence, and teen pregnancy and childbearing status (Aguia-Rojas et al. 2020, Alzate et al. 2020, Drewry and Garcés-Palacio 2020, Tsaneva and Gunes 2020, Dongarwar and Salihu 2019, and Millán 2015). In our extended specification, we consider the role of poverty, birth order, family structure, and low educational inputs. Our findings concerning the impacts of these factors align with the conclusions drawn in these studies. Furthermore, we contribute to the existing literature by providing empirical evidence regarding the role of intergenerational transmission in teen childbearing.

The magnitude of intergenerational transmission of teen childbearing observed across our six Latin American countries is relatively similar to previous estimates for the US (Haveman et al., 2008, Hoffman and Scher, 2008). However, it surpasses the estimates for other developed countries (Francesconi, 2008, for the UK, and Aizer et al., 2020, for Norway). These international differences, along with the variation we report among the six Latin American countries in our study, suggest that the intergenerational transmission of teen childbearing is sensitive to factors such as the availability of public programs, contraceptive use, the legality of abortion, as well as broader cultural values and educational achievements. Further research on the role of these factors would be desirable.

ACKNOWLEDGEMENTS

We acknowledge the financial support of CAF through the call for research projects on "Intergenerational Mobility in Latin America". We thank Jorge Agüero, Lucila Berniell, Damian Clarke, Miles Corak, Dolores De la Mata, Raquel Fernandez, Jorge Galego, Juliana Jaramillo, Kevin Lang, Francesc Ortega, Debraj Ray, Martin Rotemberg, Alejandro Sanchez Becerra, Ernesto Schargrodsky, Yochanan Shachmurove, Jan Stuhler, Wim Vijverberg and the audiences at NYU development seminar, CUNY Queens College Economics and Graduate Center seminars, the 2022 American European Health Economics Study Group, the 2021 LACEA-LAMES conference, Ridge 2022, CAF 2021 workshop, and seminars at UAB and Instituto de Economia de la Universidad de la República de Uruguay for helpful comments.

REFERENCES

Aguia-Rojas, K., Gallego-Ardila, A. D., Bonilla, M. V. E. and Rodriguez-Nino, J. N. (2020) Individual and contextual factors associated with teenage pregnancy in Colombia: A multilevel analysis. *Maternal and Child Health Journal*, **24**, 1376–1386.

Aizer, A., Devereaux, P. and Salvanes, K. (2020) Grandparents, moms, or dads? why children of teen mothers do worse in life. *Journal of Human Resources*, 1019–10524R2.

Alzate, M. M., Dongarwar, D., Matas, J. L. and Salihu, H. M. (2020) The effect of sexual literacy on adolescent pregnancy in Colombia. *Journal of Pediatric and Adolescent Gynecology*, **33**, 72–82.

An, C.-B., Haveman, R. and Wolfe, B. (1993) Teen out-of-wedlock births and welfare receipt: The role of childhood events and economic circumstances. *The Review of Economics and Statistics*, 195–208.

Ashcraft, A., Fernández-Val, I. and Lang, K. (2013) The consequences of teenage childbearing: Consistent estimates when abortion makes miscarriage non-random. *The Economic Journal*, **123**, 875–905.

Azevedo, J. P., Favara, M., Haddock, S. E., López-Calva, L. F., Muller, M. and Perova, E. (2012) *Teenage pregnancy and opportunities in Latin America and the Caribbean: on teenage fertility decisions, poverty and economic achievement*. World Bank, Washington, DC.

Azevedo, V. M. and Bouillon, C. P. (2010) Intergenerational social mobility in latin america: a review of existing evidence. *Revista de Análisis Económico–Economic Analysis Review*, **25**, 7–42.

Barber, J. S. (2001) The intergenerational transmission of age at first birth among married and unmarried men and women. *Social Science Research*, **30**, 219–247.

Berniell, L., Bonavida, C., de la Mata, D., Schargrodsky, E. et al. (2021) La movilidad educativa intergeneracional en el siglo xx en américa latina y el caribe. *Tech. rep.*, CAF Development Bank Of Latinamerica.

Black, S. E. and Devereux, P. J. (2011) Recent developments in intergenerational mobility. vol. 4 of *Handbook of Labor Economics*, 1487–1541. Elsevier. URL: https://www.sciencedirect.com/science/article/pii/S0169721811024142.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2008) Staying in the classroom and out of the maternity ward? the effect of compulsory schooling laws on teenage births. *The Economic Journal*, **118**, 1025–1054.

Card, J. J. (1981) Long-term consequences for children of teenage parents. *Demography*, **18**, 137–156.

Carneiro, P., Garcia, I. L., Salvanes, K. G. and Tominey, E. (2021) Intergenerational mobility and the timing of parental income. *Journal of Political Economy*, **129**, 757–788.

Carro, J. M., Machado, M. P. and Mora, R. (2023) The role of mothers on female labour force participation: an approach using historical parish records. *Empirical Economics*, 1–40.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W. and Yagan, D. (2011) How does your kindergarten classroom affect your earnings? evidence from project STAR. *The Quarterly Journal of Economics*, **126**, 1593–1660.

Cuesta, J., Ñopo, H. and Pizzolitto, G. (2011) Using pseudo-panels to measure income mobility in latin america. *Review of Income and Wealth*, **57**, 224–246.

Daude, C. and Robano, V. (2015) On intergenerational (im) mobility in latin america. *Latin American Economic Review*, **24**, 1–29.

Dongarwar, D. and Salihu, H. M. (2019) Influence of sexual and reproductive health literacy on single and recurrent adolescent pregnancy in Latin America. *Journal of Pediatric and Adolescent Gynecology*, **32**, 506–513.

Drewry, J. and Garcés-Palacio, I. C. (2020) Socio-demographic characteristics and female empowerment as determinants of adolescent pregnancy in colombia. *International Journal of Sexual Health*, **32**, 293–305.

Eshbaugh, E. M. (2008) Teen mothers: Marriage, cohabitation, and educational achievement. *Journal of Family Social Work*, **11**, 3–16.

Ferreira, F. H. and Gignoux, J. (2011) The measurement of inequality of opportunity: Theory and an application to latin america. *Review of income and wealth*, **57**, 622–657.

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

Francesconi, M. (2008) Adult outcomes for children of teenage mothers. *Scandinavian Journal of Economics*, **110**, 93–117.

Francesconi, M. and Nicoletti, C. (2006) Intergenerational mobility and sample selection in short panels. *Journal of Applied Econometrics*, **21**, 1265–1293.

Haveman, R., Wolfe, B. and Peterson, E. (2008) Consequences of teen childbearing for the life chances of children, 1968–88. In *Kids having kids: Economic costs and social consequences of teen pregnancy*, 324–342. Urban Institute Press Washington, DC.

Heckman, J. J. and Mosso, S. (2014) The economics of human development and social mobility. *Annual Review of Economics*, **6**, 689–733.

Hoffman, S. D. and Scher, L. S. (2008) Consequences of teen childbearing for the life chances of children, 1979-2002. *Kids having kids: Economic costs and social consequences of teen pregnancy*, 342–358.

Jäntti, M. and Jenkins, S. P. (2015) Income mobility. In *Handbook of Income Distribution*, vol. 2, 807–935. Elsevier.

Kahn, J. R. and Anderson, K. E. (1992) Intergenerational patterns of teenage fertility. *Demography*, **29**, 39–57.

Lang, K. and Nuevo-Chiquero, A. (2012) Trends in self-reported spontaneous abortions: 1970–2000. *Demography*, **49**, 989–1009.

Lang, K. and Weinstein, R. (2015) The consequences of teenage childbearing before Roe v. Wade. *American Economic Journal: Applied Economics*, **7**, 169–97.

Levine, J. A., Pollack, H. and Comfort, M. E. (2001) Academic and behavioral outcomes among the children of young mothers. *Journal of Marriage and Family*, **63**, 355–369.

Manlove, J. (1997) Early motherhood in an intergenerational perspective: The experiences of a british cohort. *Journal of Marriage and the Family*, 263–279.

Millán, J. (2015) Drugs, guns and early motherhood in Colombia.

Neidhöfer, G., Serrano, J. and Gasparini, L. (2018) Educational inequality and intergenerational mobility in latin america: A new database. *Journal of Development Economics*, **134**, 329–349.

Ramalho, E. A. and Smith, R. J. (2013) Discrete choice non-response. *The Review of Economic Studies*, **80**, 343–364.

Schwarcz, R. and Fescina, R. (2000) Maternal mortality in latin america and the caribbean. *The Lancet*, **356**, S11. URL: https://www.sciencedirect.com/science/article/pii/S0140673600919974. One World, Many Voices - The Lancet Perspectives.

Solon, G. (1999) Intergenerational mobility in the labor market. In *Handbook of Labor Economics*, vol. 3, 1761–1800. Elsevier.

Stuhler, J. et al. (2018) A review of intergenerational mobility and its drivers. *Publications Office of the European Union, Luxembourg*.

Torche, F. (2014) Intergenerational mobility and inequality: The latin american case. *Annual Review of Sociology*, **40**, 619–642.

Tsaneva, M. and Gunes, P. M. (2020) The effect of violent crime on teenage pregnancy in Mexico. *Review of Economics of the Household*, **18**, 141–164.

UNDP (2016) Human development report 2016. *United Nations Development Programme*. URL: http://report2016.archive.s3-website-us-east-1.amazonaws.com.

UNFPA (2022) Motherhood in childhood: The untold story. *Tech. rep.*, United Nations Population Fund, New York.

A | THE LIKELIHOOD FUNCTION

Let y_i and y_i^m be the daughter's TCS and her mother's TCS respectively. We assume the following probit specification:

$$\Pr\{y_i = 1 \,| y_i^m, x_i, x_i^m, z_i\} y_i = \Phi\left(\alpha y_i^m + x_i\beta + x_i^m\beta^m + z_i\gamma\right)$$

Control vectors x_i, x_i^m, and z_i are discrete. Vector x_i includes variables that are missing when the daughter is not interviewed. Similarly, x_i^m includes variables that are missing when the mother is not interviewed. Vector z_i includes controls which are always observed. The aim is to estimate parameter vector $\theta \equiv \{\alpha, \beta, \beta^m, \gamma\}$.

Let $F\left(y_i, y_i^m, x_i, x_i^m, z_i; \theta\right) \equiv \Pr\left\{y_i \left| y_i^m, x_i, x_i^m, z_i\right.\right\}$ and define I_i and I_i^m as binary indicators of observability of the daughter's and of the mother's information, respectively. We have three cases. In the first case, $I_i = I_i^m = 1$, the joint probability of the observation is:

$$\begin{aligned}\Pr\left\{I_i = I_i^m = 1, y_i, y_i^m, x_i, x_i^m, z_i\right\} = \;& \Pr\left\{I_i = I_i^m = 1 \left| y_i, y_i^m, x_i, x_i^m, z_i\right.\right\} \times \\ & F\left(y_i, y_i^m, x_i, x_i^m, z_i; \theta\right) \times \Pr\left\{y_i^m, x_i, x_i^m, z_i\right\}.\end{aligned}$$

In the second case, $I_i = 0, I_i^m = 1$, the joint probability is:

$$\begin{aligned}\Pr\left\{I_i = 0, I_i^m = 1, y_i^m, x_i^m, z_i\right\} = \;& \sum_{\{y,x\}} [\Pr\{I_i = 0, I_i^m = 1 \,| y, y_i^m, x, x_i^m, z_i\} \times \\ & F\left(y, y_i^m, x, x_i^m, z_i; \theta\right) \times \Pr\{y_i^m, x, x_i^m, z_i\}].\end{aligned}$$

In the last case, $I_i = 1, I_i^m = 0$, the joint probability is:

$$\begin{aligned}\Pr\left\{I_i = 1, I_i^m = 0, y_i, x_i, z_i\right\} = \;& \sum_{\{y^m,x^m\}} [\Pr\{I_i = 1, I_i^m = 0 \,| y_i, y^m, x_i, x^m, z_i\} \times \\ & F\left(y_i, y^m, x_i, x^m, z_i; \theta\right) \times \Pr\{y^m, x_i, x^m, z_i\}].\end{aligned}$$

Let $H_{y_i y_i^m}^{I_i I_i^m} \equiv \Pr\left\{I_i, I_i^m \left| y_i, y_i^m\right.\right\}$, with I_i, I_i^m, y_i, and $y_i^m \in \{0, 1\}$, and $\Pi_{y_i^m, x_i, x_i^m, z_i} \equiv \Pr\left\{y_i^m, x_i, x_i^m, z_i\right\}$, where $\Pi_{y_i^m, x_i, x_i^m, z_i} \in [0, 1)$. For observations with information for mother and daughter, Assumption 1 implies that:

$$\Pr\{I_i = I_i^m = 1, y_i, y_i^m, x_i, x_i^m, z_i\} = H_{y_i y_i^m}^{11} F\{y_i, y_i^m, x_i, x_i^m, z_i; \theta\} \Pi_{y_i^m, x_i, x_i^m, z_i}.$$

When only the daughter's information is observed, the joint probability is:

$$\begin{aligned}\Pr\{I_i = 1, I_i^m = 0, y_i, x_i, z_i\} &= \sum_{y^m, x^m} \left(H_{y_i y^m}^{10} F\{y_i, y^m, x_i, x^m, z_i; \theta\} \Pi_{y^m, x_i, x^m, z_i}\right) \\ &= \sum_{y^m} \left(H_{y_i y^m}^{10} \sum_{x^m} \left(F\{y_i, y^m, x_i, x^m, z_i; \theta\} \Pi_{y^m, x_i, x^m, z_i}\right)\right)\end{aligned}$$

where $F\{y_i, y^m, x_i, x^m; \theta\}$ and Π_{y^m, x_i, x^m, z_i} are evaluated at values y_i, x_i, and z_i and all potential combinations of running values y^m and x^m. Finally, the joint probability of an observation without daughter's information is:

$$\Pr\{I_i = 0, I_i^m = 1, y_i^m, x_i^m, z_i\} = \sum_y \left(H^{01}_{yy_i^m} \sum_x \left(F\{y, y_i^m, x, x_i^m, z_i; \theta\} \Pi_{y_i^m, x, x_i^m, z_i} \right) \right)$$

The model parameters are: (*i*) probit model parameters θ; (*ii*) conditional missing process parameters $\left\{ H^{II^m}_{yy^m} \right\}$; and (*iii*) marginal probabilities $\{\Pi_{y^m,x,x^m,z}\}$. The conditional likelihood $\mathcal{L}_i$ for any given observation i is:

$$\begin{aligned} \mathcal{L}_i = \quad & \left(H^{11}_{y_i y_i^m} F\left\{ y_i, y_i^m, x_i, x_i^m, z_i; \theta \right\} \Pi_{y_i^m, x_i, x_i^m, z_i} \right)^{I_i I_i^m} \times \\ & \left(\sum_{y^m} \left(H^{10}_{y_i y^m} \sum_{x^m} \left(F\{y_i, y^m, x_i, x^m, z_i; \theta\} \Pi_{y^m, x_i, x^m, z_i} \right) \right) \right)^{I_i (1 - I_i^m)} \times \\ & \left(\sum_{y} \left(H^{01}_{y y_i^m} \sum_{x} \left(F\left\{ y, y_i^m, x, x_i^m, z_i; \theta \right\} \Pi_{y_i^m, x, x_i^m, z_i} \right) \right) \right)^{(1 - I_i) I_i^m} . \end{aligned} \tag{A.1}$$

The log-likelihood function results from the sum of the log of $\mathcal{L}_i$, $\log(\mathcal{L}) = \sum_{i=1}^N \log(\mathcal{L}_i)$ and is maximized subject to the following constraints:

$$\begin{aligned} H^{II^m}_{yy^m}, \Pi_{y^m,x,x^m,z} \in [0,1) \; & \text{ for all } I, I^m, y, y^m, x, x^m, z \\ \sum_{I, I^m} H^{II^m}_{yy^m} & = 1 \text{ for all } y, y^m \\ \sum_{y^m, x, x^m, z} \Pi_{y^m,x,x^m,z} & = 1 \end{aligned} \tag{A.2}$$

B | ADDITIONAL TABLES

TABLE B.1 Country indicators

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
Teen childbearing status:						
TCS by 15	0.103	0.079	0.122	0.11	0.086	0.073
TCS by 18	0.3	0.257	0.335	0.328	0.227	0.242
Sexual behavior:						
Age at first sex	16	14.6	16.1	14	14.6	14.8
Sex by 15	0.234	0.238	0.266	0.19	0.267	0.162
Sex by 18	0.562	0.58	0.561	0.484	0.627	0.48
Number of sex partners	1.8	4.1	3.4	1.5	2.6	1.7
Household formation:						
Age at first child	20.3	20.7	20	19.7	20.9	21.1
Married by 18	0.202	0.187	0.264	0.266	0.153	0.163
Parent is household head	0.732	0.735	0.633	0.718	0.611	0.746
Number of children	2.4	1.6	2	2.1	1.9	1.9
Contraception and gender roles:						
Uses contraception	0.313	0.663	0.623	0.435	0.286	0.407
Fertility knowledge	0.391	0.332	0.122	0.205	0.249	0.362
Submissive gender role	0.25	0.078	0.123	0.267	0.366	0.09
Ideal number of children	3.8	2.6	3.9	3.6	3.3	2.6
Demographics:						
Years of education	8.6	9.1	9	6.1	6.5	9.6
Rural	0.34	0.216	0.288	0.547	0.532	0.265
More than six household members	0.221	0.176	0.17	0.352	0.322	0.21
Household characteristics:						
House has finished floor	0.717	0.937	0.958	0.705	0.635	0.629
House has modern toilet facilities	n.a.	0.93	0.623	0.604	0.118	0.576
House has minimum toilet facilities	n.a.	0.939	0.953	0.941	0.734	0.855
House has electricty	0.804	0.973	0.947	0.879	0.381	0.839
Human development:						
Human development index (2016)	0.674	0.727	0.722	0.478	0.493	0.740

Notes: The Human development index (2016) is taken from UNDP (2016). All other variables are country weighted averages using all records in the DHS household and women questionnaires for all countries and years covered in the estimation sample. Variables 'TCS by 15' and 'TCS by 18' indicate whether the woman was pregnant or had a live birth by the age of 15 and 18, respectively. 'Age at first sex' is the age at first sexual intercourse. Variables 'Sex by 15' and 'Sex by 18' indicate whether the woman had had sexual intercourse by the age of 15 and 18, respectively. 'Number of sex partners' is the total lifetime number of sexual partners (conditional on having had sexual relations). 'Age at first child' is the age at first birth (conditional on having had a birth). 'Married by 18' indicates whether the woman is currently married or lives with her partner (only for women aged 18). 'Parent is household head' indicates whether a teen's parent is the household head (only for women 18 or younger). 'Number of children' is the number of children ever born. 'Uses contraception' indicates whether a teen woman currently uses a modern contraceptive methods (only for women 18 or younger who are fecund and not currently pregnant). 'Fertility knowledge' takes value one if the woman correctly states the moment during the menstrual cycle in which a woman is likely most fertile. 'Submissive gender role' takes value one if the woman agrees with 'the partner having a sexually transmitted disease is no reason to use a condom' or 'the partner having a sexual affair with another individual is no reason to refuse sex', or both. 'Ideal number of children' is the number of children the respondant would have liked to have in her whole life, irrespective of the number she already has. 'Years of education' is the number of years of complete education. 'Rural' takes value one if the household is located in a rural area. 'More than six household members' takes value one if the household has more than six members. 'House has finished floor' indicates that the household's floor is not mud, sand, or dirt. 'House has modern toilet facilities' indicates that the household has a toilet connected to a sewer system or a septic well. 'House has minimum toilet facilities' indicates that the household has access to at least a pit latrine. 'House has electricty' indicates whether the house has electricity.

TABLE B.2 Terminations & live births

PANEL A: During adolescence

	% women with terminations					
	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
Only no teen-mothers	1.918	3.578	2.634	1.946	1.369	1.907
Only teen-mothers	3.694	7.879	8.707	7.246	2.795	4.935
All	2.498	4.947	5.010	3.496	1.677	2.724

	% women sexually active with terminations					
	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
Only no teen-mothers	3.074	4.728	3.898	2.927	1.869	2.988
Only teen-mothers	3.631	7.823	8.678	7.353	2.812	4.956
All	3.310	5.937	6.178	4.692	2.126	3.715

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
Terminations rate	7.360	14.432	12.300	11.419	7.396	9.593
Chilbearing Rate	32.644	31.841	39.128	29.237	21.596	27.009

PANEL B: After adolescence

	% sexually active women with terminations					
	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
All	6.151	9.836	13.525	4.047	3.828	7.186

Notes: Pooled data from all surveys post 2000. Women aged 23 during the year of interview. Terminations include abortions, misscarriages, and still births. *Childbearing rate* is the percentage of women who have had at least one live birth during adolescence. *Terminations rate* is the percentage of women with at least one termination during adolescence over all women who were pregnant during adolescence.

TABLE B.3 Average marginal effects for the probability of the mother being interviewed

	Bolivia	Colombia	DR	Guatemala	Haiti	Peru
			Panel A: TCS			
AME	-0.302***	-0.237***	-0.310***	-0.378***	-0.105***	-0.242***
	(0.026)	(0.009)	(0.012)	(0.017)	(0.023)	(0.011)
N.obs.	2884	18821	10181	4714	5600	20183
			Panel B: Sex before 16			
AME	-0.249***	-0.161***	-0.294***	-0.339***	-0.057***	-0.191***
	(0.021)	(0.007)	(0.010)	(0.015)	(0.015)	(0.008)
N.obs.	2881	18818	10147	4712	5599	20183
			Panel C: Ever married			
AME	-0.433***	-0.404***	-0.370***	-0.465***	-0.168***	-0.368***
	(0.029)	(0.009)	(0.010)	(0.015)	(0.027)	(0.011)
N.obs.	2884	18821	10181	4714	5600	20183
			Panel D: At most primary			
AME	-0.185***	-0.206***	-0.185***	-0.140***	-0.083***	-0.163***
	(0.019)	(0.011)	(0.010)	(0.014)	(0.013)	(0.010)
N.obs.	2884	18821	10181	4714	5600	20183

Notes: *** $p<0.01$. Standard errors in parenthesis. Average marginal effects are derived from probit estimates, indicating the probability of the mother being interviewed when the daughter is also interviewed. All models include the daughter's age and time dummies as part of the basic specification controls. Each panel presents country-specific average marginal effect (AME) estimates for an additional control variable beyond the basic variable specification. For example, in Panel A, the AME represents the average marginal effect of the daughter's Teen Childbearing Status (TCS) on the likelihood of the mother being interviewed, while controlling for the daughter's age, time dummies, and the daughter's TCS. 'TCS' is the daughter's TCS. 'Sex before 16' is a dummy variable that takes value one if the daughter reports to have had sex before age 16 and zero otherwise. 'Ever married' indicates whether daughter has ever been married or lived with her partner. 'At most primary' takes value one if the daughter has obtained at most primary education.
The number of observations reflects the subset of cases where the daughter is interviewed and all relevant controls are observed.