

Financial Aid, College Opportunity, and Teen Fertility: Evidence from *Ser Pilo Paga* in Colombia*

Michael D. Bloem[†] Jesús Villero[‡]

September 2021

Abstract

We study the effects on teen fertility of *Ser Pilo Paga* (SPP), a generous college financial aid program introduced in Colombia in 2014. We use a triple difference approach that leverages variation in the share of female students eligible for the program across municipalities and the fact that the introduction of SPP should not affect the education and fertility decisions of older women not targeted by the program. We find that after the introduction of SPP, fertility rates for women aged 15-19 years old decreased in more affected municipalities by about 10 percent relative to less affected municipalities. These effects are driven by municipalities that conceivably had greater excess demand for college before SPP, including areas located closer to SPP-eligible college institutions and areas with both low college-going expectations and high perceived returns to going to college. Our results suggest that increasing economic opportunities through expanding college access can contribute to lowering teen fertility rates.

JEL Codes: H51, I22, I23, I24, I25, J13, O15.

*We are grateful to Daniel Kreisman, Jonathan Smith, and seminar participants at Georgia State University and Universidad del Norte for useful feedback.

[†]Department of Economics, Andrew Young School of Policy Studies, Georgia State University. 55 Park Place, 6th Floor, Atlanta, GA 30303. E-mail: mbloem1@gsu.edu.

[‡]Department of Economics, Andrew Young School of Policy Studies, Georgia State University. 55 Park Place, 6th Floor, Atlanta, GA 30303. E-mail: jvilleroarocal@gsu.edu.

1 Introduction

Teen childbearing is associated with large private costs including lower educational attainment and poorer labor market outcomes, and large public costs including greater reliance on social programs. Furthermore, rates of teen childbearing are higher among low-income communities. Youth may be more likely to engage in risky behaviors when chances of economic mobility are low and opportunities to make investments in their own economic progress are limited. Thus, one possible way to break this cycle of early childbearing and poverty is to focus policies on reducing inequality in opportunities for youth to make investments in their own economic progress.

In this paper, we investigate how teen behavior responds to increases in educational opportunities by studying the effects on teen fertility of Colombia's 2014 introduction of *Ser Pilo Paga* (roughly translated as "Hard Work Pays Off"), a college financial aid program covering full tuition costs at high-quality institutions for high-achieving, low-income students. If students are facing binding financial constraints, newly available access to college aid may represent a positive shock to economic opportunity. Furthermore, an increase in educational opportunities may change the relative costs associated with engagement in risky activities.

This setting is suitable for studying this topic because *Ser Pilo Paga* (SPP) had large effects on college enrollment. SPP substantially relaxed credit constraints (and increased college supply) in a context with significant income-based gaps in college enrollment, high college tuition costs, and little pre-existing access to credit. [Londoño-Vélez, Rodríguez and Sánchez \(2020\)](#) show that SPP dramatically increased college enrollment on the eligibility margin (57 to 87 percent increases depending on the complier population), virtually eliminating the income-based gap in college enrollment among high-achieving students. SPP also increased college enrollment among low-income, aid-ineligible students (14 percent increase) due to colleges increasing supply to capture the additional demand. There is also evidence that the introduction of SPP altered human capital investment decisions before college through increased test scores ([Bernal and Penney, 2019](#); [Laajaj, Moya and Sánchez, 2018](#)) and reduced school dropout rates ([Basto-Aguirre, 2019](#)). Given this evidence, it is reasonable to expect that teens may also alter their behaviors on other dimensions.

Our primary contribution is documenting teen fertility responses to a large

change in schooling *opportunities*. An existing literature studies the effects of education on teen pregnancy using exogenous variation from school entry policies and mandatory schooling laws (e.g., [Black, Devereux and Salvanes \(2008\)](#); [McCrary and Royer \(2011\)](#)) and from the duration of school days ([Berthelon and Kruger, 2011](#)). Since these policies require additional time to be spent in school, evidence of declines in fertility in these settings may mostly be due to less available time to engage in risky behaviors—i.e., a pure incapacitation effect. Little is known about how increasing opportunities for schooling affects fertility decisions, where youth still have agency in their schooling choices. Moreover, little work exists in general about how college financial aid policies affect noneducational outcomes for teens, such as fertility.¹

We use a triple difference research design leveraging municipality-level variation in SPP eligibility rates determined prior to the introduction of the program and the fact that SPP should not affect the fertility decisions of older women. Eligibility for SPP was based on test scores on the national standardized exam and scores on a household wealth index. By the time SPP was announced, students had already taken the national standardized exam and there was not sufficient time to request a reevaluation of their household wealth index. Our empirical approach uses eligibility rates only from this first cohort of students, who could not influence their scores around the eligibility cutoffs.

We find that fertility rates for women aged 15-19 decreased by about 10 percent in more affected municipalities relative to less affected municipalities. We rule out the possibility that the effects we observe are entirely driven by the “direct” effects of SPP-eligible students enrolling in college and diverting births that otherwise would have occurred during their teen college-age years (i.e., 18 or 19 years old). An extreme assumption where all SPP-eligible female students would have had a birth in the absence of SPP can only account for 30 percent of the average decline in fertility rates we attribute to SPP. Thus, we interpret our findings as largely comprised of “indirect” effects of SPP, where the new college opportunities created by the program influenced teen fertility decisions before being able to benefit from the program. We also rule out that possible confounding events drive our results, including Colombia’s peace agreement with the Revolutionary Armed Forces of

¹An exception is unpublished work that studies the rollout of several Promise-type college scholarship programs in the United States which finds evidence of declining teen arrest rates and suggestive evidence of declining teen birth rates after the programs were announced ([Doleac and Gibbs, 2016](#)).

Colombia and the Zika virus epidemic.

We also find that the teen fertility impacts of SPP were driven by municipalities that conceivably had greater excess demand for college before SPP. First, we use pre-SPP individual-level survey data to create a municipality-level measure of excess demand for college based on the share of respondents who indicated having both a low likelihood of attending college *and* a high expected wage return if they did attend college. We estimate larger teen fertility impacts in municipalities with higher excess demand for college based on this measure (11.3 percent) than municipalities with lower excess demand (5 percent). Second, since most students attend a college close to home, we estimate the heterogeneity in SPP's effects by the municipality-level distance to the nearest SPP-eligible institution. We find that municipalities that are closer to an SPP-eligible institution had larger declines in teen fertility (11 percent) than municipalities that are further away (5.6 percent). These results support the interpretation that the declines in teen fertility were driven by an increase in college opportunities.

We add to the literature in three important ways. First, our analysis represents an empirical test of theoretical predictions that increases in economic opportunities (and increases in opportunity costs) influence the fertility decisions of young women (Becker, 1960; Willis, 1973; Kearney and Levine, 2014). Our results suggest that improving the future economic prospects of young women through the expansion of educational opportunities can successfully help reduce teen pregnancy and early childbearing. Second, our paper is related to a literature that studies the teen fertility impacts of interventions in developing countries that aim to improve economic opportunities and empowerment for adolescent women (Jensen, 2012; Duflo, Dupas and Kremer, 2015; Muralidhara and Prakash, 2017; Bandiera, Buehren, Burgess, Goldstein, Gulesci, Rasul and Sulaiman, 2020). We extend this body of work by presenting evidence from a large-scale setting in an upper middle-income country. Third, by examining understudied noneducational outcomes (Doleac and Gibbs, 2016), we build on the literature of the effects of the *Ser Pilo Paga* program (Londoño-Vélez et al., 2020; Bernal and Penney, 2019; Laajaj et al., 2018; Basto-Aguirre, 2019) and the effects of college financial aid programs more broadly on the decisions of high school students (Cáceres-Delpiano, Giolito and Castillo, 2018).

The remainder of this paper is organized as follows: In the next section, we describe the details of *Ser Pilo Paga* and discuss the Colombian context. The third

section describes the data sources we use, discusses the key variables used in our analyses, and presents trends in fertility rates in Colombia. The fourth section discusses our identification strategy and estimation approaches. The fifth section presents our core empirical results and tests for the sensitivity and robustness of those results. Finally, section six concludes.

2 Background

2.1 *Ser Pilo Paga* and higher education in Colombia

Ser Pilo Paga was announced by surprise on October 1st of 2014 by President Santos's administration. The program was publicly funded and covered recipients' full tuition cost of attending a four- or five-year undergraduate program at any university in Colombia with a High Quality Accreditation. The aid came in the form of a loan that is forgiven upon graduation, although only about 1.9 percent of SPP beneficiaries from the first three cohorts had dropped out of the program ([Londoño-Vélez et al., 2020](#)).² Additionally, SPP recipients would receive a biannual stipend of at least the national minimum wage.

Eligibility for SPP was based on both need and merit. First, students must score above a cutoff on the SABER 11, the Colombian equivalent of the SAT in the United States. The SABER 11 exam is taken by nearly all high school seniors regardless of their plans to attend an institution of higher education. SABER 11 scores play a significant role in college admissions, with about four-fifths of institutions using them in admissions considerations ([OECD and World Bank, 2012](#)). The SABER 11 cutoff score is placed at approximately the 91st percentile each year.

Second, students must be below a cutoff on the SISBEN, Colombia's wealth index used to target social welfare programs. The SISBEN cutoff varies by geographic location. The cutoff is 57.21 in the 14 main metropolitan areas, 56.32 in other urban areas, and 40.75 in rural areas. Between 2015 and 2018, there were about 10,000 SPP beneficiaries per year, which represents about one-third of students attending an institution with High Quality Accreditation.

In the first year of the program, students had already taken the SABER 11 exam before SPP was announced. Moreover, there was insufficient time to request a reevaluation of their household wealth index before determining eligibility for

²The SPP program considers students to have dropped out if they have not attended a high-quality institution for three or more consecutive semesters.

SPP. Thus, students in this first cohort had no opportunities to influence their test scores or wealth index scores in response to the SPP eligibility cutoffs.

While all institutions of higher education in Colombia are required to maintain minimum quality standards from the Ministry of Education, institutions can voluntarily apply for a certificate of High Quality Accreditation from the National Accreditation Council. Evidence shows that High Quality Accreditation is correlated with college exit tests and graduates' wage profiles ([Camacho, Messina and Uribe, 2016](#)). At the time SPP was announced, only 33 institutions (10 percent) had received High Quality Accreditation, although these "high quality" institutions enroll roughly one-third of all students in higher education.

Of the 33 high quality institutions, 21 were private and 12 were public. Tuition at the private universities is very expensive, both compared to private universities in other countries and to the public universities in Colombia ([OECD and World Bank, 2012](#)). Since the tuition at the high-quality public universities are relatively low, these institutions are historically oversubscribed, leading to highly selective admissions. Prior to SPP, there were very few financial aid opportunities for high-achieving, low-income students. Only 11 percent of first-year undergraduate students had a student loan before SPP ([Ferreyra, Avitabile, Botero Álvarez, Haimovich Paz and Urzúa, 2017](#)).

Historically, Colombia has been characterized by large income-based disparities in college enrollment. As of 2013, 25 percent of low-income students who completed secondary school enrolled in higher education, compared to 54 percent of middle- and upper-income students who completed secondary school ([Melguizo, Sánchez and Velasco, 2016](#)). While overall gross enrollment rates have increased from 32 percent in 2006 to 49 percent in 2014, most of this increase is from an expansion of low quality programs in low quality institutions with questionable labor market returns ([Camacho et al., 2016](#)).

2.2 Teen fertility in Colombia

Similar to many Latin American and Caribbean countries, teen fertility is high in Colombia. Estimated at 70.7 births per 1,000 women aged 15-19 years in 2014 (when SPP was announced), the adolescent fertility rate in Colombia was slightly higher than the Latin American average, more than twice that of other countries

with similar income levels and nearly three times higher than in the United States.³ These “higher-than-expected” adolescent fertility rates observed in Latin American countries are likely associated with the high levels of inequality of income (and opportunities) observed in the region (Azevedo et al., 2012).

In contrast, in 2014, Colombia had a lower *total* fertility rate than the mean Latin American country, similar to the overall fertility rates in other upper middle-income countries and the United States. As a result, early childbearing is a worrisome phenomenon and a policy concern in Colombia, given its association with worse prospects for the adolescent mothers and their children in terms of health, education, and labor market outcomes (Gaviria, 2010; Azevedo et al., 2012).

Early parenting in Colombia is primarily a female phenomenon. Data from the most recent Demographic and Health Survey (2015) show that adolescent women are 6.4 times more likely to have at least one child than adolescent men—13.6 percent versus 2.1 percent (Flórez and Soto, 2019). Furthermore, birth records data indicate that only 22 percent of births to adolescent women between 2008 and 2014 had a teenage father.⁴

While teenage pregnancy affects all income groups, it is particularly worrying among low-income women. Low-income Colombian teenagers are five times more likely to have ever been pregnant than their high-income peers (Flórez and Soto, 2019).

In the last decade and a half, Colombia has implemented several programs and policies directly aimed at reducing teenage pregnancies.⁵ Among the most relevant initiatives is the implementation of the Youth Friendly Health Services Model (SSAAJ, from the Spanish acronym) and the Program of Education in Sexuality and Construction of Citizenship (PESCC), both launched in 2007-2008 and scaled up nationally in subsequent years.⁶ In 2012, the national government additionally launched a strategic framework to address the issue comprehensively, articulating different actors within the public sector.⁷ On top of others not directly targeted at reducing fertility like *Familias en Acción*, the conditional cash transfer program in Colombia, these initiatives likely contributed to the downward trend in teenage

³As a region, Latin America and the Caribbean has the second highest fertility rate for teenagers in the world, second only to Sub-Saharan Africa. Data are from the World Bank’s WDI.

⁴The age of consent in Colombia is 14 years old.

⁵See part three of Vargas Trujillo, Flórez, Cortés and Ibarra, eds (2019) for a recent review.

⁶*Modelo de Servicios de Salud Amigables para Adolescentes y Jóvenes (SSAAJ)* and *Programa de Educación para la Sexualidad y Construcción de Ciudadanía (PESCC)* in Spanish.

⁷National Department of Planning. *Documento CONPES Social* No. 147.

fertility observed in the country since the mid-2000s after a concerning period of increase during the 1990s (Flórez and Soto, 2019; Attanasio et al., 2021).⁸

3 Data and key variables

This section describes our data sources and key variables. We gather data from publicly available sources on births and population counts in Colombia in order to calculate age-specific fertility rates. To compute a measure that indicates which municipalities were more or less affected by SPP, we collect SABER 11 test score data to calculate SPP eligibility rates.

3.1 Data sources

We use the universe of birth records and annual population estimates from the Colombian National Department of Statistics (DANE, from the Spanish acronym) from 2008 to 2019. Individual birth records contain information about the mother's age in 5-year intervals (i.e., 15-19, 20-24, 25-29, etc.) and about her municipality of *residence* (on top of the one where the birth took place). We use this data to create a municipality by age group panel dataset of age-specific fertility rates, which is our primary outcome.

We also use administrative data from the Colombian Institute for the Evaluation of Education (ICFES) containing student-level information of the national standardized high school exit exam, SABER 11, including test scores and socio-demographic characteristics. Importantly, this data includes the municipality of residence of the student and a variable on SPP eligibility for the years in which the program was in place.

Finally, we complement our data with pre-SPP municipality characteristics which we obtain from the Center for the Study of Economic Development (CEDE) from Universidad de los Andes, the Ministry of Education, and DANE.

3.2 Construction of analysis measures

We use the birth records and population estimates to create a municipality-of-residence by age group panel dataset of age-specific fertility rates, our primary

⁸Since all these policies were implemented years before SPP was introduced, we do not view them as threats to our identification strategy, but rather as possible factors explaining the decline in adolescent fertility observed before SPP.

outcome. Our main estimates use the natural log of these fertility rates.

We define our municipality-level treatment intensity measure as the rate of female SABER 11 test takers in 2014 who are eligible for SPP in each municipality. We use information only for test scores from the fall semester of 2014.⁹ For these students, the SPP program was announced after they had taken the SABER 11 exam. Thus, these eligibility rates avoid possible endogenous responses to the announcement of the program or its eligibility thresholds. We then separate the sample at the median, the top half representing the *treatment municipalities* and the bottom half representing the *comparison municipalities*.

We attempt to assess the validity of our treatment intensity measure by estimating whether it is associated with an increase in SABER 11 test scores after SPP is introduced. This is essentially testing whether we can replicate the results from [Bernal and Penney \(2019\)](#) and [Laajaj et al. \(2018\)](#) using our treatment measure. We use individual-level data on female SABER 11 test takers between 2010 and 2016 and estimate a triple difference model that compares standardized test scores of SISBEN-eligible students between treatment and comparison municipalities. See [Appendix B](#) for a full description of this analysis. Consistent with the existing evidence, we find that, after the introduction of SPP, SABER 11 test scores increased in treatment municipalities for SISBEN-eligible students by about 0.05 standard deviations, relative to comparison municipalities. These findings support the notion that our treatment intensity measure is adequately capturing the mechanisms underlying the introduction of SPP.

3.3 Analytic sample and summary statistics

We restrict our sample to municipalities (i) with at least one observed birth from each age group we use in our empirical analysis for all years between 2008 to 2019 and (ii) with SABER 11 information in 2014. By doing this, we drop extremely small municipalities from our sample. Our final analysis sample consists of a balanced panel of 1,061 municipalities (out of 1,122 in the country) for 2008-2019.

[Table 1](#) displays means and standard deviations of SPP eligibility rates in 2014

⁹The SABER 11 exam is administered by ICFES in both the spring and fall semester, with the vast majority of students taking the exam in the fall semester. SPP eligibility on the SABER 11 margin was based on exams taken in the fall semester. Typically, only students in a limited set of private schools whose academic calendar is synchronized with the United States take the SABER 11 exam during the first (spring) semester of the year. For example, in the year in which SPP was introduced 95.6 percent of the test takers took the test in the second (fall) semester.

for both treatment and comparison municipalities. We show unweighted means and means weighted by the number of births in the pre-period, which follows our main estimation approach. All comparison municipalities had zero female students who were eligible for SPP in 2014.¹⁰ We do not interpret this as indicating that comparison municipalities are completely “untreated.” Since, as [Londoño-Vélez et al. \(2020\)](#) document, SPP had effects on students throughout the distribution of SABER 11 test takers, our view is that students did not need to be eligible for SPP to be affected by the introduction of the program. We use eligibility rates (in 2014) to characterize municipalities as more or less affected by SPP.

Table 1. Summary statistics of key variables.

| | Weighted | | Unweighted | |
|--------------------------------------|----------------|----------------|----------------|----------------|
| | Comparison | Treatment | Comparison | Treatment |
| <i>Births per 1,000 women (2008)</i> | | | | |
| Age 15-19 | 91.2 (31.9) | 84.6 (24.8) | 80.3 (31.4) | 85.3 (29.2) |
| Age 25-29 | 87.3 (30.1) | 98.3 (21.6) | 82.5 (32.7) | 93.2 (29.3) |
| <i>SPP eligibility rates (2014)</i> | | | | |
| Per 1,000 female students | 0.0 (0.0) | 14.7 (9.1) | 0.0 (0.0) | 22.6 (24.7) |
| Number of municipalities | 521 | 540 | 521 | 540 |

Notes: This table shows means and standard deviations (in parentheses) for age-specific birth rates and SPP eligibility rates between treatment and comparison municipalities. Birth rates are from 2008 and SPP eligibility rates are from 2014. Treatment municipalities are above the median in female eligibility rates for SPP in 2014, while comparison municipalities are below the median. Weights are the number of births between 2008 and 2014.

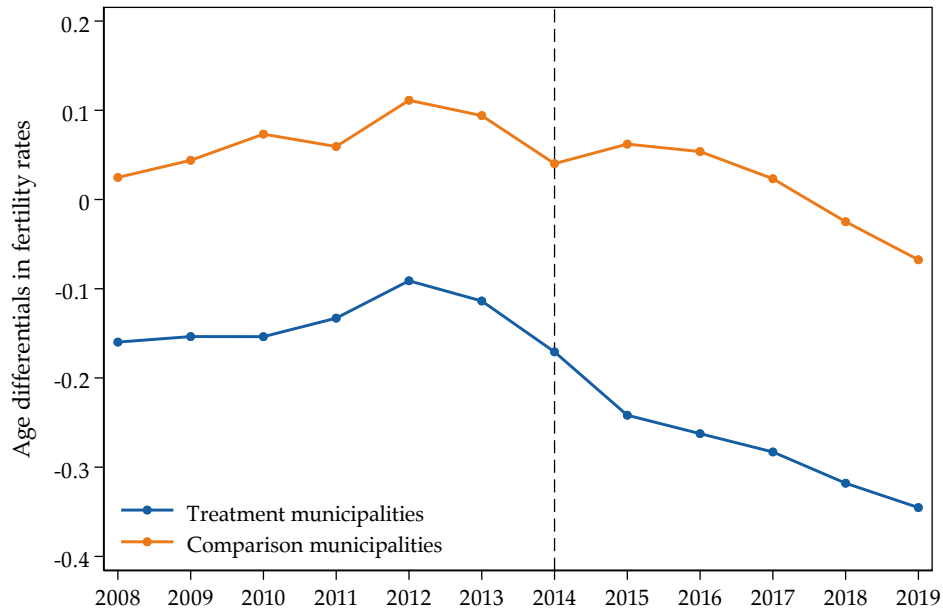
[Figure A2](#) visualizes the municipality level variation in the discrete version of SPP eligibility rates in 2014. While there are some clusters of treatment municipalities at a local level, there are treatment and comparison municipalities in every region of Colombia. [Table A1](#) suggests that these two groups of municipalities were different, on average, in terms of pre-SPP characteristics. For example, treatment municipalities have higher SABER 11 test scores, larger populations, and lower poverty levels. Importantly, since our identification strategy relies on an

¹⁰[Figure A1](#) plots the full distribution of SPP eligibility rates for the municipalities in our sample. The figure shows that about half of municipalities had zero female students eligible for SPP in 2014 and that there is not much variation among municipalities with at least one eligible female student.

assumption of parallel fertility rate trends in absence of SPP, these differences do not invalidate our empirical strategy.

Table 1 displays means of fertility rates in 2008 (the earliest pre-period year in our sample) for both 15-19 year olds and 25-29 year olds. These are the two age groups we use in our triple difference empirical strategy, which we describe in detail in the next section. Compared to the comparison municipalities, treatment municipalities have slightly lower fertility rates for women aged 15-19, but slightly higher fertility rates for women aged 25-29. Figure A3 plots the complete distribution of adolescent fertility rates, which shows a substantial amount of overlap between the distributions of treatment and comparison municipalities. Finally, Figure 1 shows the raw trends in the age-differentials in fertility rates by treatment and comparison municipalities. This figure mimics our triple difference research design. The age differences in fertility rates trend similarly between treatment and comparison municipalities through 2014, but diverge sharply in 2015, the first year after SPP was introduced.

Figure 1. Trends in fertility rate age-differentials (15-19 year olds minus 25-29 year olds) by treatment and comparison municipalities.



Notes: This figure plots trends in average age-differentials in log fertility rates between age groups for both treatment and comparison municipalities (15-19 year olds minus 25-29 year olds). Treatment municipalities are above the median in female eligibility rates for SPP in 2014, while comparison municipalities are below the median. The averages weight municipalities by the number of births between 2008 and 2014.

4 Empirical analysis

To estimate the effect of SPP on teen fertility, we follow a triple difference approach. Our design leverages the variation in the share of female students eligible for the program across municipalities and the fact that the introduction of SPP did not affect the education and fertility decisions of older women not targeted by the program. This section first describes the identifying assumptions behind our research design, and then presents the estimation procedure.

4.1 Identification strategy

We estimate the impact of SPP on the adolescent fertility rate by comparing the fertility rate of 15-19 year olds to the fertility rate of 25-29 year olds before and after the introduction of SPP among municipalities with different eligibility rates for the program. We choose 25-29 year olds as our within-municipality comparison group because it is the group closest in age to the 15-19 year olds that is likely not affected by the introduction of SPP. The 25-29 year old group cannot be SPP beneficiaries and most are likely past their college-going years. The 20-24 group is partially affected by the introduction of SPP during our sample period given the nature of our birth records data and is also more likely to be affected by the general equilibrium effects of SPP on the higher education market.

The identifying assumption for this triple difference design is that in the absence of the policy, the differentials in fertility outcomes between 15-19 and 25-29 years old in municipalities with higher SPP eligibility rates (treatment municipalities) would have evolved similarly to these differentials in municipalities with lower SPP eligibility rates (comparison municipalities). This is the usual parallel trends assumption underlying difference-in-differences designs applied to the triple difference case. We provide evidence in support of this assumption when we discuss our main results in [subsection 5.1](#).

We opt for a triple difference approach in part because our analyses determine that simple difference-in-differences (DD) approaches do not identify the effect of SPP. [Figure A4](#) plots DD event-study coefficients comparing the fertility rate of teenagers in treatment and comparison municipalities—a natural starting point. We see evidence that the teen fertility rate in treatment municipalities was trending down relative to comparison municipalities before the introduction of SPP, even though we indeed observe a substantial change in trends after 2014. In [Figure A5](#)

we also plot DD event-study coefficients that show the trends in the fertility rate differentials between age groups, with separate plots for treatment and comparison municipalities. Both treatment and comparison municipalities show that fertility rates were increasing for women aged 15-19 relative to women aged 25-29 leading up to 2012 but then decreasing precipitously afterward. These trends prior to the introduction of SPP prohibit interpreting these DD estimates as identifying the causal effects of SPP on adolescent fertility.

The fact that the trends observed in [Figure A5](#) are consistent between treatment and comparison municipalities indicates the presence of age-specific fertility trends that are common across municipalities. Our triple difference design differences out these nationwide age-specific fertility trends and also account for municipality-specific shocks affecting fertility independently of women’s age (e.g., local business cycles affecting everyone’s fertility decisions). After accounting for these trends, our identification then relies on SPP disproportionately affecting the fertility rate of teenagers in treatment municipalities.

The main identification threat to our strategy is the existence of other confounding events or policies that could have differentially affected the fertility rate of women in different age groups *and* are also correlated with our municipality-level treatment variable. We provide evidence that other majors events that occurred in the country around 2014 cannot explain our results in [subsection 5.3](#). Another potential issue would be if many students move from a comparison municipality to a treatment municipality when going to college. However, [Ferreyra et al. \(2017\)](#) illustrate that most students attend a college near their home, which mitigates this concern. Moreover, our data suggests that to the extent that students are moving across municipalities, this mobility is likely only occurring within treatment municipalities, not between comparison and treatment municipalities.

4.2 Estimation approaches

We implement the triple difference design with an event-study specification that we estimate by Ordinary Least Squares (OLS). More specifically, we use the following

specification:

$$\begin{aligned}
Y_{amt} = & \alpha + Teen_a \times SPP_m^* \times \sum_{\substack{\tau=2008 \\ \tau \neq 2014}}^{2019} \beta_\tau \mathbb{1}[t = \tau] \\
& + \gamma_1 (Teen_a \times \theta_m) + \gamma_2 (\theta_m \times \lambda_t) \\
& + \gamma_3 (Teen_a \times \lambda_t) + \varepsilon_{amt},
\end{aligned} \tag{1}$$

where Y_{amt} is the log fertility rate of age group $a \in \{15-19, 25-29\}$ in municipality m and year $t \in [2008, 2019]$. $Teen_a$ is an indicator for age group defined as $Teen_a = \mathbb{1}[a = 15-19]$. SPP_m^* denotes treatment and comparison municipalities and is defined as $SPP_m^* = \mathbb{1}[SPP_m > \text{median}(SPP_m)]$ with SPP_m being the rate of female students eligible for the program in a given municipality in 2014. θ_m and λ_t are municipality and year fixed effects, respectively. Finally, ε_{amt} is an error term.

In Equation 1, $\beta_{\tau, t > 2014}$ represents the average treatment effect of SPP on teen fertility at time $t = \tau$ after the introduction of the program. Equation 1 includes three two-way interactions between fixed effects for age group, municipalities, and years. The vectors γ_1 , γ_2 , and γ_3 contain the parameters associated with these interactions. The interaction between age group and municipality fixed effects ($Teen_a \times \theta_m$) controls for time-invariant, municipality-specific factors (both observed and unobserved) that affect fertility rates and that are potentially different by age groups. The interaction between municipality and year fixed effects ($\theta_m \times \lambda_t$) controls for municipality-specific trends in fertility rates common to all age groups. Finally, the interaction between age group and year fixed effects ($Teen_a \times \lambda_t$) accounts for age-specific trends in fertility that are common to all municipalities. As mentioned earlier, the remaining and identifying source of variation we leverage is the differential effect that SPP had on the adolescent fertility rate in the treatment municipalities.

To summarize the event-study estimates of SPP's effects in a single estimate, we also estimate a version of Equation 1 that replaces the year indicators with a single post-2014 indicator variable:

$$\begin{aligned}
Y_{amt} = & \tilde{\alpha} + \beta (Teen_a \times SPP_m^* \times Post_t) + \tilde{\gamma}_1 (Teen_a \times \theta_m) \\
& + \tilde{\gamma}_2 (\theta_m \times \lambda_t) + \tilde{\gamma}_3 (Teen_a \times \lambda_t) + u_{amt},
\end{aligned} \tag{2}$$

where $Post_t = \mathbb{1}[t > 2014]$ and everything else defined as in Equation 1. In Equation 2, β is the summary triple difference parameter across all post-SPP years.

In all our regressions, we cluster the standard errors at the municipality level and weight each cell by the number of births in the municipality and age group between 2008 and 2014 (pre-SPP). This weighting approach gives more importance to municipalities that historically contribute more to the overall fertility rates and limits the influence of municipalities with small populations (Guerra-Cújar, Prem, Rodríguez-Lesmes and Vargas, 2020). However, we report robustness to other weighting procedures, including no weights.

5 Results

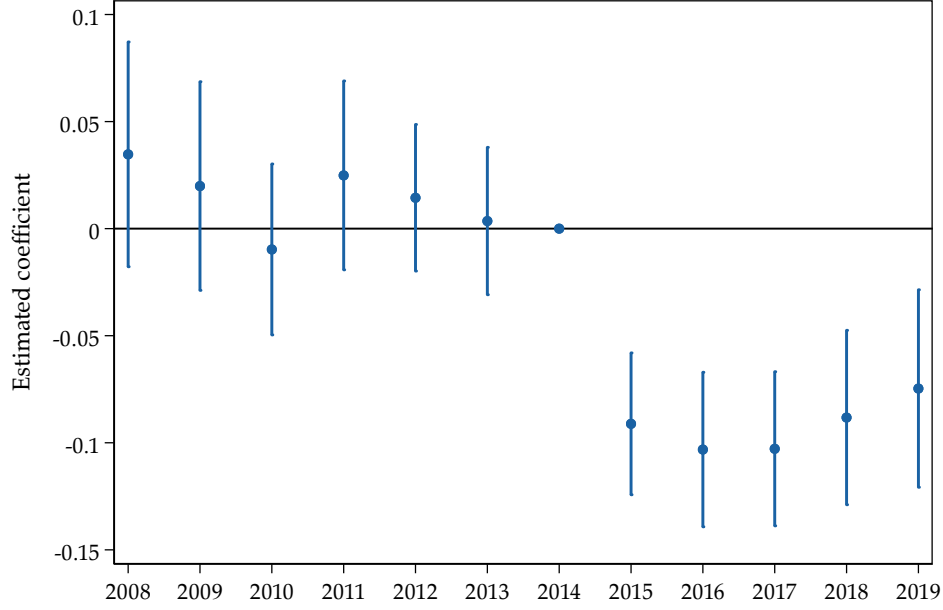
This section reports and discusses our results. We begin by presenting our main estimates of the teen fertility impacts of *Ser Pilo Paga*. We then present supporting evidence for our main estimates by examining the heterogeneity in the teen fertility impacts by municipality characteristics. Finally, we discuss and report a variety of analyses to assess the robustness of our estimates.

5.1 Main results

Figure 2 displays our event-study estimates of β_τ in Equation 1. None of the coefficients in the pre-period are statistically significant at the 5 percent level. Following Borusyak, Jaravel and Spiess (2021), we conduct a more formal test of pre-period trends by estimating our main triple difference specification using only the set of untreated observations and running a joint F-test of these coefficients. With a p -value of 0.232, this test cannot reject the null hypothesis that the pre-trend coefficients are jointly equal to zero. This provides empirical support in favor of the parallel trends assumption of our research design.

Starting in 2015, the first year after the introduction of SPP, there is a stark decrease in the fertility rate of women aged 15-19 in treatment municipalities relative to comparison municipalities. The effects are more indicative of a level shift in the fertility rate trend rather than a change in the slope of the trend. The coefficients begin to attenuate in later years, perhaps indicating a decreasing relevance of the initial eligibility rates in 2014 as a proxy for more highly affected municipalities. We report the summary estimates using Equation 2 in column 1 of Table 2. The triple difference estimate in this specification indicates that SPP reduced fertility rates of women aged 15-19 in treatment municipalities by 10 percent relative to

Figure 2. Triple difference event study estimates.



Notes: This figure plots the triple difference event study estimates of β_τ from Equation 1. The dots represent the estimated coefficients and the vertical lines represent 95 percent confidence intervals. All estimates are weighted by the number of births between 2008 and 2014 for each age group in each municipality. Standard errors are clustered at the municipality level.

comparison municipalities.¹¹

We attempt to assess whether the magnitude of this 10 percent effect is reasonable by translating this effect size into an approximation for how many students would have to change from having a child to not having a child. To do this, we first apply this 10 percent effect to the pre-period weighted average birth rate of women aged 15-19 in treatment municipalities to get the number of fewer births per 1,000 women implied by this estimate. We then compare this to the average number of female SABER 11 test takers in 2014 in treatment municipalities. These calculations indicate that about one percent of female students would be required to change from having a child to not having a child to reach our estimated effect size. This back-of-the-envelope calculation seems to be consistent with estimates of the overall college enrollment effect of SPP from Londoño-Vélez et al. (2020), who estimate a four percentage point increase in immediate college enrollment.

Since our data includes mother's age in intervals, our main estimates may only reflect the direct effect of students receiving SPP, going to college, and reducing

¹¹The exact percentage changes implied by our estimates are given by $100 \times [\exp(\hat{\beta}) - 1]$.

(or delaying) childbearing that would have occurred during their teen college-age years (i.e., 18 or 19 years old). To assess whether this direct channel fully explains our estimates, we conduct an exercise that calculates how large our estimates would be under an extreme assumption where all SPP-eligible female students would have had a birth in absence of SPP.

To do this, we first impute counterfactual fertility rate trends for treatment municipalities using the imputation estimator derived in [Borusyak et al. \(2021\)](#). Then, we subtract from these counterfactual trends the number of births implied by an assumption that all SPP-eligible female students in each year would have experienced a birth and compare this “simulated” fertility rate trend to the actual trend we observe. We plot the results of this exercise in [Figure A6](#). The solid line represents the observed fertility rate trend in treatment municipalities. The dashed line represents the estimated counterfactual fertility rate trend. Finally, the dotted line shows the simulated fertility rate trend if effects were only driven by SPP-eligible females.

The results of this exercise reveal that an assumption where all SPP-eligible female students would have had a birth in absence of SPP can account for only 30 percent of the average decline in fertility rates we attribute to SPP. The vertical difference between counterfactual and simulated trends in [Figure A6](#) represents an upper bound of the direct effect of SPP eligibility on fertility. This indicates that at least 70 percent of the effects we observe are “indirect” effects, where the new opportunities created by the program influenced teen fertility decisions for students before they were even able to benefit directly from the program.

5.2 Heterogeneity

In this subsection, we probe the credibility of our estimated fertility impacts of SPP by assessing the heterogeneity in these estimates by municipality characteristics. If our main estimates are in fact identifying the effect of SPP, we would expect larger effects in areas with higher excess demand for college, where there are more students who perhaps were facing credit constraints prior to SPP that were prohibiting attending college. We would also expect larger effects in areas that are closer to SPP-eligible institutions, since most students who attend college in Colombia do so at an institution near their home ([Ferreyra et al., 2017](#)). We analyze both of these characteristics in turn below.

Excess Demand for College — To assess the heterogeneous teen fertility impacts

Table 2. Triple difference summary estimates.

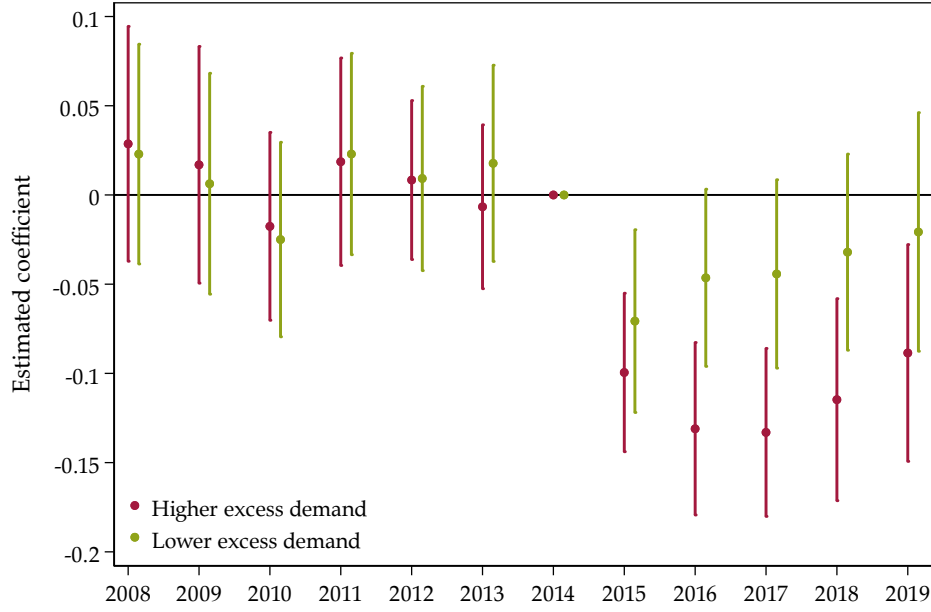
| Sample: | Log fertility rate | | | | |
|-------------------------------|----------------------|------------------------------|---------------------|---|----------------------|
| | Full sample (1) | By excess demand for college | | By distance to SPP eligible institution | |
| | | Higher (2) | Lower (3) | Shorter (4) | Longer (5) |
| $Teen \times SPP \times Post$ | -0.105*** (0.020) | -0.120*** (0.025) | -0.051** (0.020) | -0.116*** (0.026) | -0.058*** (0.020) |
| Observations | 25,464 | 12,936 | 12,432 | 12,744 | 12,720 |
| Treatment municipalities | 540 | 213 | 327 | 288 | 252 |
| Comparison municipalities | 521 | 305 | 212 | 243 | 278 |
| Pre-trends test p -value | 0.232 | 0.527 | 0.508 | 0.316 | 0.442 |

Notes: This table presents summary triple difference estimates using Equation 2. Column 1 presents the full sample estimates. Columns 2 and 3 splits the sample by whether municipalities are above or below the median of a measure of excess demand for college as described in the text. Columns 4 and 5 splits the sample by whether municipalities are above or below the median distance to the nearest SPP-eligible institution. All estimates are weighted by the number of births between 2008 and 2014 for each age group in each municipality. Standard errors are clustered at the municipality level and presented in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

of SPP by excess demand for college, we utilize survey responses on the college expectations of a 10 percent random sample of SABER 11 test takers in 2013 and 2014 to create a proxy measure for excess demand. We use female students' responses to questions that ask about how likely they are to attend college immediately after high school and about the wages they expect to earn if they completed either high school or college. At the municipality level, we calculate the share of respondents whose responses indicate they are unlikely to attend college, but they also have a relatively high perceived return to going to college as measured by the difference between students' expected wages from completing college and completing high school. Finally, we group municipalities by whether they are above or below the median share of students who meet this criteria and estimate Equation 1 separately for these two subsets of municipalities. The triple difference coefficients are presented in Figure 3.

Using subsets of municipalities necessarily changes the composition of municipalities in each comparison group. Thus, it is important to ensure that the parallel trends assumption is still reasonably satisfied. None of the pre-period coefficients are statistically different from zero. Furthermore, using the pre-trends test described in subsection 5.1, we cannot reject the null hypothesis that all pre-period

Figure 3. Triple difference event study estimates by excess demand for college.



Notes: This figure presents triple difference event study estimates of β_τ from Equation 1 where the sample is split by whether municipalities are above or below the median of a measure of excess demand for college as described in the text. The dots represent the estimated coefficients and the vertical lines represent 95 percent confidence intervals. All estimates are weighted by the number of births between 2008 and 2014 for each age group in each municipality. Standard errors are clustered at the municipality level.

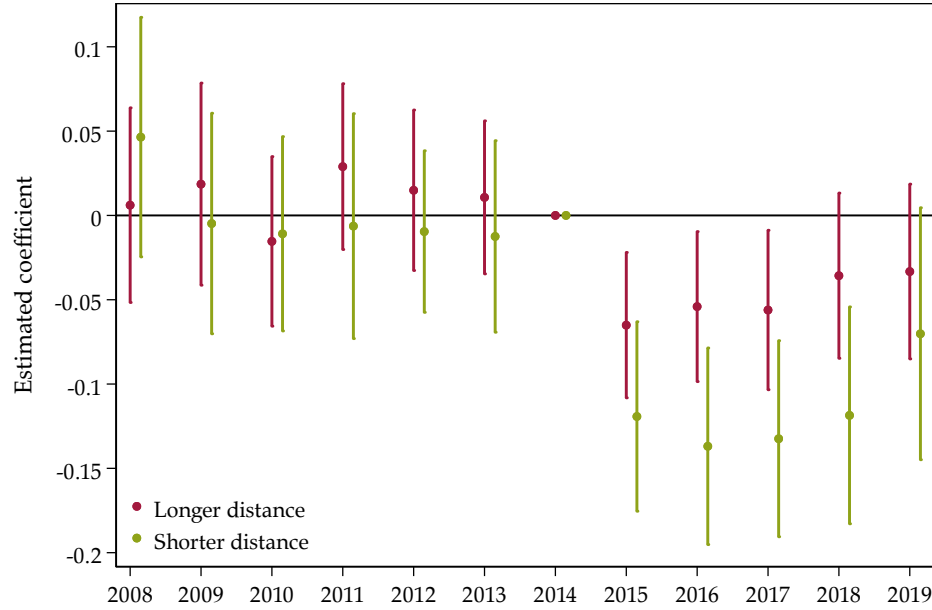
coefficients for both the treatment and comparison excess demand subsets are zero (p -values are 0.527 and 0.508, respectively).

Figure 3 also shows that the post-period effects of SPP are larger for the subset of municipalities that had relatively higher excess demand for college based on our measure. Summary estimates of the effects, reported in columns 2 and 3 of Table 2, are negative 5 percent for municipalities with lower excess demand and negative 11.3 percent for municipalities with higher excess demand. These estimates are statistically different at the 5 percent level.

Distance to Nearest SPP-Eligible Institution — To assess the heterogeneous teen fertility impacts of SPP by distance to the nearest SPP-eligible institution, we first locate each of the institutions with High Quality Accreditation where students could receive SPP benefits. We then calculate at the municipality level the distance to each of these institutions while identifying the institution that is the closest. Then, we group municipalities by whether their distance to the nearest SPP-eligible

institution is above or below the median distance among all municipalities. Finally, we estimate Equation 1 separately for these two subsets of municipalities.

Figure 4. Triple difference event study estimates by distance to nearest SPP-eligible institution.



Notes: This figure presents triple difference event study estimates of β_τ from Equation 1 where the sample is split by whether municipalities are above or below the median distance to the nearest SPP-eligible institution. The dots represent the estimated coefficients and the vertical lines represent 95 percent confidence intervals. All estimates are weighted by the number of births between 2008 and 2014 for each age group in each municipality. Standard errors are clustered at the municipality level.

Figure 4 plots the triple difference estimates for these two subgroups of municipalities. Pre-period coefficients are all statistically insignificant, and our pre-trend test cannot reject a null hypothesis that the pre-period coefficients are all zero for either subgroups (p -values of 0.316 and 0.442). The results show that, while both subsets experience a significant decrease in fertility rates starting in 2015, there are larger effects for the subset of municipalities that are relatively closer to the nearest SPP-eligible institution. Summary estimates of these effects, reported in columns 4 and 5 of Table 2, are negative 5.6 percent for municipalities that are relatively further away from the nearest SPP-eligible institution and negative 11 percent for municipalities that are relatively closer. These summary estimates are statistically different at the 10 percent level.

Overall, we take these heterogeneity analyses to support the interpretation

of our main estimates as identifying the effect of SPP on teen fertility. The larger fertility impacts in municipalities that are relatively closer to the nearest SPP-eligible institution and in municipalities that have relatively higher pre-SPP excess demand for college provides more evidence that the introduction of SPP led to significant behavioral responses, primarily through the mechanism of the increased college opportunities provided by the program.

5.3 Robustness

This subsection overviews a series of analyses that assess the robustness of our results to possibly confounding events, alternative specification choices, and other sensitivity checks.

Possible confounding events — Since our setting involves a single treatment time period, we are potentially vulnerable to events that happened simultaneously (or around the same time) as the introduction of SPP. Although, we note that to truly be a threat to identification, these simultaneous events would have to differentially affect women of the different age groups and be correlated with SPP eligibility rates. Nevertheless, we assess whether three events that occurred at a similar time might be driving our results: 1) the unilateral permanent ceasefire by the Revolutionary Armed Forces of Colombia (FARC, from the Spanish acronym) in December 2014 as part of the by then ongoing peace process between the guerrilla group and the Colombian government, 2) the Zika virus epidemic, which occurred from October 2015 to July 2016, and 3) the *Jornada Única* initiative, which gradually transitioned some public secondary schools that were operating half-day shifts into full school days beginning in 2015.

For each of these possibly confounding events, we re-estimate our main specification using only a subset of municipalities that were likely unaffected by the relevant event. If these events are not driving our results, we would expect to see estimates based on these subsets of municipalities that are similar to our main estimates. We provide a full description of these analyses in [Appendix C](#) and report these estimates in [Table C1](#). Indeed, we consistently estimate large and statistically significant effects of SPP in each of these subsample analyses. We conclude that these three events cannot explain the effects we observe.

Alternative specification choices — We assess the robustness of our results to alternative specification choices. Our main estimates weight observations by the pre-period number of births. In doing so, we give heavier weight to municipalities

that traditionally contribute more to Colombia’s birth rates. We also report estimates instead using number of females as weights and using the number of births or women in a single pre-period year (2008) as weights. Finally, following [Solon, Haider and Wooldridge \(2015\)](#), we report estimates without using weights. These estimates are reported in [Table A2](#). The results show that estimates are consistent across alternative weights. With no weights, the triple difference estimate drops to -0.041, but is still statistically significant at the 1 percent level. This implies that SPP’s effects on teen fertility are relatively larger in more populous municipalities.¹²

Column 3 of [Table A2](#) also presents results using a continuous and standardized version of our treatment intensity measure. The triple difference estimate is smaller, which is to be expected since a one standard deviation in SPP eligibility is less than the average difference in eligibility rates between treatment and comparison municipalities.¹³ Finally, to be sure our results are not driven by a small number of municipalities, we re-estimate our main specification while each time excluding municipalities in a single department (reported in [Figure A7](#)).¹⁴ We also estimate our main specification by excluding all municipalities that include a department’s capital city (reported in column 2 of [Table A2](#)). The results from each of these regressions produce estimates that are very similar to our main estimates.

Other sensitivity checks — Although our empirical strategy does not rely on a staggered rollout design, we implement an imputation estimator developed by [Borusyak et al. \(2021\)](#) that is robust to the issues with estimating difference-in-differences designs with fixed effects documented in the literature ([Goodman-Bacon, 2021](#); [Callaway and Sant’Anna, 2020](#); [de Chaisemartin and D’Haultfœuille, 2020](#); [Baker, Larcker and Wang, 2021](#)). One advantage of using this estimator is that it allows us to include unit-specific trends into the estimating equation. Since OLS estimates unit-specific trends using both pre and post period data, the estimates are potentially contaminated by the dynamics of the treatment effects ([Borusyak et al., 2021](#)). Results using this alternative estimator are reported in [Table A3](#). The triple difference estimate using the imputation estimator is -0.102, nearly identical to using OLS. When including unit-specific trends the estimate dips slightly to -0.079, but is still statistically significant at the 1 percent level.

¹²Column 4 of [Table A3](#) also reports results in which we do not use regression weights during the estimation but instead use the weights to recover a weighted ATT from the set of individual, municipality-specific treatment effects. This estimate is -0.080 (0.017).

¹³We note, however, that our pre-trends test rejects the null hypothesis that the pre-period coefficients are all zero when using the continuous treatment measure.

¹⁴Departments in Colombia are similar to states in the United States.

To assess whether the decline in teen fertility we observe in treatment municipalities is the result of pure chance, we perform a permutation test that randomly assigns municipalities to be treatment or comparison municipalities. We then compare our main estimate to a distribution of estimates across 5,000 randomly assigned groups of treatment municipalities. To do this, we use the randomization inference routine developed by [Heß \(2017\)](#) and the specification in [Equation 2](#). We report the results in [Figure A8](#). Reassuringly, we see that our main estimate is in the far left tail of the distribution of estimated triple difference coefficients.

Finally, we perform a placebo strategy to further support the validity of the parallel trends assumption required for our estimates to have a causal interpretation. In [Table A4](#), we use 2008-2014 data and estimate the same specification in [Equation 2](#) pretending that SPP was introduced in years 2008-2013. The overall estimated effects after each of the placebo treatment years are always statistically insignificant and close to zero. The parallel trends assumption is an assumption about counterfactuals and, therefore, untestable. We have shown robust evidence that the differentials in fertility rates between younger and older women were not trending differently among treatment and comparison municipalities before SPP, which adds support that this would have been the case during the post-period had the program not been introduced.

6 Conclusion and Discussion

In this paper, we study the teen fertility impacts of *Ser Pilo Paga*, Colombia's generous college financial aid program for high-achieving, low-income students. After the 2014 introduction of the program, we find that teen fertility rates decreased by about 10 percent in municipalities more affected by SPP relative to less affected municipalities. Consistent with an interpretation of these effects being the result of increases in college opportunities, we estimate larger effects in municipalities with higher pre-SPP excess demand for college.

Our results suggest that increasing future economic opportunities for young women can lead to meaningful reductions in teen fertility. Prior to SPP, Colombia was characterized by large socio-economic gaps in college enrollment due to severe financial constraints, low access to credit, and high college tuition costs. We posit that, in countries with high inequality, college financial aid programs like SPP that decrease inequality of opportunity can have behavioral effects on teen childbearing

and perhaps other outcomes. The characteristics of SPP—namely its generosity, salience, and simplicity—would seem to be important in accounting for the far-reaching impacts of the program, which is consistent with the college financial aid literature more broadly (e.g., [Bettinger, Long, Oreopoulos and Sanbonmatsu \(2012\)](#); [Dynarski, Libassi, Micheltore and Owen \(2021\)](#)).

In 2018, under a new presidential administration, the Colombian government announced that the *Ser Pilo Paga* program would no longer accept new beneficiaries and the program would be replaced. The program gained controversy during its four years due to its high cost to the government and the fact that most SPP beneficiaries attended private institutions. Our findings illustrate important indirect benefits of SPP. While such benefits alone may not justify the program's costs, they should be included in a full accounting of the program's costs and benefits, including the effects on college enrollment outcomes documented in [Londoño-Vélez et al. \(2020\)](#). Future research should explore other shorter-term outcomes and the longer-term outcomes of SPP, such as labor market outcomes, as data becomes available.

References

- Attanasio, Orazio, Lina Cardona Sosa, Carlos Medina, Costas Meghir, and Christian Manuel Posso-Suárez,** “Long Term Effects of Cash Transfer Programs in Colombia,” Working Paper 29056, National Bureau of Economic Research July 2021.
- Azevedo, Joao Pedro, Marta Favara, Sarah E. Haddock, Luis F. López-Calva, Miriam Muller, and Elizaveta Perova,** *Teenage pregnancy and opportunities in Latin America and the Caribbean: On teenage fertility decisions, poverty and economic achievement*, World Bank, Washington, DC, 2012.
- Baker, Andrew C., David F. Larcker, and Charles C.Y. Wang,** “How Much Should We Trust Staggered Difference-In-Differences Estimates?,” *Working Paper*, 2021.
- Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman,** “Women’s Empowerment in Action: Evidence from a Randomized Control Trial in Africa,” *American Economic Journal: Applied Economics*, 2020, 12 (1), 210–259.
- Basto-Aguirre, Nathalie,** “The Unexpected Effects of a Merit-Based Scholarship: Evidence from Ser Pilo Paga,” *Working Paper*, 2019.
- Becker, Gary S.,** “An Economic Analysis of Fertility,” in Gary S. Becker, ed., *Demographic and Economic Change in Developed Countries*, Princeton, NJ: Princeton University Press, 1960, pp. 209–231.
- Bernal, Gloria L. and Jeffrey Penney,** “Scholarships and Student Effort: Evidence from Colombia’s Ser Pilo Paga Program,” *Economics of Education Review*, 2019, 72, 121–130.
- Berthelon, Matias E. and Diana I. Kruger,** “Risky behavior among youth: Incapacitation effects of school on adolescent motherhood and crime in Chile,” *Journal of Public Economics*, 2011, 95 (1), 41–53.
- Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu,** “The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment,” *Quarterly Journal of Economics*, 2012, 127 (3), 1205–1242.
- Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes,** “Staying in the Classroom and Out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births,” *The Economic Journal*, 2008, 118 (530), 1025–1054.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess,** “Revisiting Event Study Designs: Robust and Efficient Estimation,” *Working Paper*, 2021.

- Callaway, Brantly and Pedro H.C. Sant’Anna**, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2020.
- Camacho, Adriana, Julián Messina, and Juan Pablo Uribe**, “The Expansion of Higher Education in Colombia: Bad Students or Bad Parents?,” Discussion Paper 452, Inter-American Development Bank 2016.
- Cáceres-Delpiano, Julio, Eugenio Giolito, and Sebastián Castillo**, “Early impacts of college aid,” *Economics of Education Review*, 2018, 63, 154–166.
- de Chaisemartin, Clément and Xavier D’Haultfœuille**, “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 2020, 110 (9), 2964–2996.
- Doleac, Jennifer L. and Chloe R. Gibbs**, “A Promising Alternative: How Making College Free Affects Teens’ Risky Behaviors,” *Unpublished manuscript*, 2016.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer**, “Education, HIV, and Early Fertility: Experimental Evidence from Kenya,” *American Economic Review*, 2015, 105 (9), 2757–2797.
- Dynarski, Susan, CJ Libassi, Katherine Micheltore, and Stephanie Owen**, “Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students,” *American Economic Review*, 2021, 111 (6), 1721–1756.
- Ferreira, María Marta, Ciro Avitabile, Javier Botero Álvarez, Francisco Haimovich Paz, and Sergio Urzúa**, “At a Crossroads—Higher Education in Latin America and the Caribbean,” Report, Washington, DC: World Bank Group 2017.
- Flórez, Carmen Elisa and Victoria Eugenia Soto**, “Tendencias del embarazo en la adolescencia y sus determinantes próximos en Colombia.” In [Vargas Trujillo et al., eds \(2019\)](#) pp. 59–98.
- Gamboa, Luis Fernando and Paul Rodríguez-Lesmes**, “The Fertility-Inhibiting Effect of Mosquitoes: Socio-economic Differences in Response to the Zika Crisis in Colombia,” *Economics and Human Biology*, 2019, 35, 63–72.
- Gaviria, Alejandro**, “Cambio social en Colombia durante la segunda mitad del siglo XX,” Documentos CEDE 30, Universidad de los Andes, Facultad de Economía, CEDE October 2010.
- Goodman-Bacon, Andrew**, “Difference-in-Differences with Variation in Treatment Timing,” *Journal of Econometrics*, 2021.
- Guerra-Cújar, María Elvira, Mounu Prem, Paul Rodríguez-Lesmes, and Juan F. Vargas**, “The Peace Baby Boom: Evidence from Colombia’s Peace Agreement with the FARC,” *LACEA Working Paper Series*, October 2020, (0052).

- Heß, Simon**, “Randomization inference with Stata: A guide and software,” *Stata Journal*, 2017, 17 (3), 630–651.
- Hincapié, Diana**, “Do Longer School Days Improve Student Achievement? Evidence from Colombia,” IDB Working Paper Series IDB-WP-679, Washington, DC 2016.
- Jensen, Robert**, “Do Labor Market Opportunities Affect Young Women’s Work and Family Decisions? Experimental Evidence from India,” *Quarterly Journal of Economics*, 2012, 127 (2), 753–792.
- Kearney, Melissa S. and Phillip B. Levine**, “Income Inequality and Early Nonmarital Childbearing,” *Journal of Human Resources*, 2014, 49 (1), 1–31.
- Laajaj, Rachid, Andrés Moya, and Fabio Sánchez**, “Equality of Opportunity and Human Capital Accumulation: Motivational Effect of a Nationwide Scholarship in Colombia,” Documentos CEDE 26, Universidad de los Andes, Facultad de Economía, CEDE May 2018.
- Londoño-Vélez, Juliana, Catherine Rodríguez, and Fabio Sánchez**, “Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia,” *American Economic Journal: Economic Policy*, May 2020, 12 (2), 193–227.
- McCrary, Justin and Heather Royer**, “The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth,” *American Economic Review*, 2011, 101, 158–195.
- Melguizo, Tatiana, Fabio Sánchez, and Tatiana Velasco**, “Credit for Low-Income Students and Access to and Academic Performance in Higher Education in Colombia: A Regression Discontinuity Approach,” *World Development*, 2016, 80, 61–77.
- Muralidhara, Karthik and Nishith Prakash**, “Cycling to School: Increasing Secondary School Enrollment for Girls in India,” *American Economic Journal: Applied Economics*, 2017, 9 (3), 321–350.
- OECD and World Bank**, “Reviews of National Policies for Education—Tertiary Education in Colombia,” Report, Paris: OECD Publishing 2012.
- Prem, Mounu, Juan F. Vargas, and Olga Namen**, “The Human Capital Peace Dividend,” *Journal of Human Resources*, 2021, *Forthcoming*.
- , **Santiago Saavedra, and Juan F. Vargas**, “End-of-conflict deforestation: Evidence from Colombia’s peace agreement,” *World Development*, 2020, 129, 104852.
- Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge**, “What Are We Weighting For?,” *Journal of Human Resources*, 2015, 50 (2), 301–316.

Vargas Trujillo, Elvia, Carmen Elisa Flórez, Darwin Cortés, and Marta Carolina Ibarra, eds, *Embarazo temprano: Evidencias de la investigación en Colombia*, Universidad de los Andes, Ediciones Uniandes-Universidad del Rosario, 2019.

Willis, Robert J., "A New Approach to the Economic Theory of Fertility Behavior," *Journal of Political Economy*, 1973, 81 (2), S14–S64.

A Appendix Tables and Figures

Table A1. Municipality characteristics.

| | With weights | | Without weights | |
|---|--------------------------------|---------------------------------------|--------------------------------|--------------------------------------|
| | Comp. munic. mean (1) | Treat. munic. difference (2) | Comp. munic. mean (3) | Treat munic. difference (4) |
| SPP eligible on SABER 11 margin (all students) | 0.017 (0.001) | 0.075*** (0.013) | 0.021 (0.001) | 0.040*** (0.002) |
| SPP eligible on SABER 11 margin (female students) | 0.009 (0.001) | 0.067*** (0.011) | 0.011 (0.001) | 0.041*** (0.002) |
| Ln(population) (2008) | 9.695 (0.048) | 2.833*** (0.512) | 9.153 (0.030) | 0.861*** (0.059) |
| Share of women aged 15-19 (2008) | 0.098 (0.000) | -0.005*** (0.001) | 0.094 (0.000) | -0.002*** (0.001) |
| Share of women aged 25-29 (2008) | 0.074 (0.000) | 0.008*** (0.001) | 0.071 (0.000) | 0.003*** (0.001) |
| Rural share of population (2008) | 0.604 (0.015) | -0.396*** (0.039) | 0.644 (0.009) | -0.142*** (0.014) |
| Distance to department's capital (km) | 93.207 (4.281) | -54.439*** (8.584) | 85.404 (2.560) | -16.075*** (3.422) |
| Altitude (meters above sea level) | 809.458 (47.535) | 296.639 (255.707) | 1079.906 (39.329) | 115.700** (55.931) |
| Multidimensional poverty index (0-1) (2005) | 0.772 (0.008) | -0.291*** (0.040) | 0.751 (0.006) | -0.115*** (0.009) |
| Public expenditure per capita (2008) | 893.103 (32.170) | 150.703 (104.141) | 1011.499 (24.762) | 35.142 (175.334) |
| Tax revenue per capita (2008) | 92.251 (5.577) | 224.879*** (72.389) | 93.969 (4.154) | 47.799*** (8.200) |
| Public investment in education per capita (2008) | 290.892 (140.126) | 19.161 (146.478) | 325.065 (162.825) | -95.344 (198.534) |
| Net enrollment rate 10-11 grade (2011) | 0.280 (0.007) | 0.172*** (0.020) | 0.302 (0.006) | 0.093*** (0.009) |
| Exposed to FARC (0/1) (2011-2014) | 0.177 (0.025) | -0.039 (0.044) | 0.134 (0.015) | -0.024 (0.020) |

Notes: This table compares pre-SPP characteristics between treatment and comparison municipalities. Columns 1 and 2 present results of a regression of a municipality characteristic on an indicator for being a treatment municipality, using births between 2008 to 2014 as weights. Column 1 shows the coefficients on the intercept term and represents the mean of comparison municipalities. Column 2 shows coefficients on the treatment indicator and represents the mean difference between treatment and comparison municipalities. Columns 3 and 4 present results of the same regressions without weights. Money variables are measured in 2019 thousand Colombian pesos. Robust standard errors are presented in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). Significance stars are suppressed for coefficients on the intercept term.

Table A2. Robustness to alternative specifications.

| | Log fertility rate | | | | | | |
|-------------------------------|----------------------|----------------------|---------------------|----------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| $Teen \times SPP \times Post$ | -0.105*** (0.020) | -0.081*** (0.016) | -0.034** (0.017) | -0.098*** (0.019) | -0.110*** (0.020) | -0.115*** (0.020) | -0.041*** (0.015) |
| Treatment measure | Discrete | Discrete | Continuous | Discrete | Discrete | Discrete | Discrete |
| Weights | Births 2008-14 | Births 2008-14 | Births 2008-14 | Births 2008 | Women 2008 | Women annual | None |
| Exclude capital cities | No | Yes | No | No | No | No | No |
| Observations | 25,464 | 24,696 | 25,464 | 25,464 | 25,464 | 25,464 | 25,464 |
| Treatment municip. | 540 | 510 | 540 | 540 | 540 | 540 | 540 |
| Comparison municip. | 521 | 519 | 521 | 521 | 521 | 521 | 521 |
| Pre-trends test p -value | 0.232 | 0.436 | 0.008 | 0.210 | 0.074 | 0.073 | 0.279 |

Notes: This table reports results from variations of our main specification. In column 3, we replace SPP_m^* in Equation 2 by the (standardized) rate of female SABER 11 test takers in 2014 who are eligible for SPP in each municipality. Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Table A3. Borusyak et al. (2021)'s imputation estimator and unit-specific trends.

| | Log fertility rate | | | |
|-------------------------------|----------------------|----------------------|----------------------|----------------------|
| Estimation: | OLS | Imputation estimator | | |
| | (1) | (2) | (3) | (4) |
| $Teen \times SPP \times Post$ | -0.105*** (0.020) | -0.102*** (0.017) | -0.079*** (0.018) | -0.080*** (0.017) |
| Observations | 25,464 | 25,464 | 25,464 | 25,464 |
| Treatment municipalities | 540 | 540 | 540 | 540 |
| Comparison municipalities | 521 | 521 | 521 | 521 |
| Unit-specific trends | No | No | Yes | No |
| Regression weights | Yes | Yes | Yes | No |
| Pre-trends test p -value | 0.232 | 0.232 | 0.210 | 0.279 |

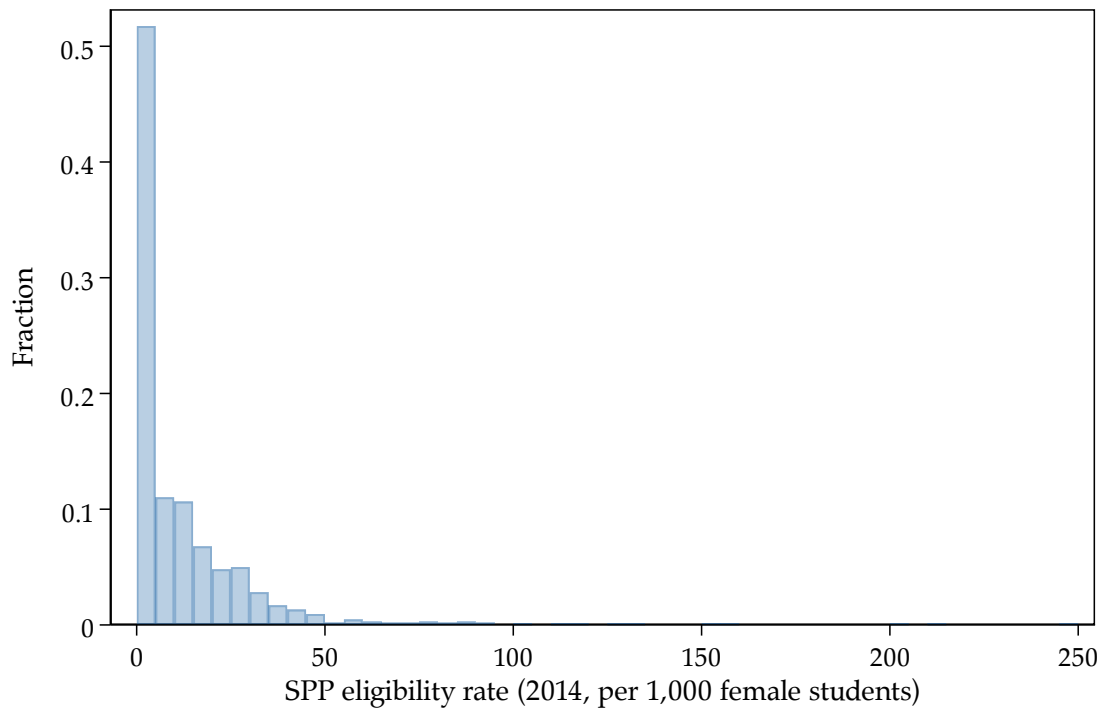
Notes: This table reports results using Borusyak et al. (2021)'s imputation estimator. Column 1 replicates our main results using OLS to estimate Equation 2. Columns 2-4 report the treatment effect using the imputation estimator including the same set of fixed effects from Equation 2. Column 3 additionally includes unit-specific trends by adding the term $\tilde{\gamma}_4 (Teen_a \times \theta_m \times t)$ to Equation 2. Estimation in columns 1-3 is done by a weighted regression using the number of births between 2008 and 2014 for each age group in each municipality. Column 4 uses these same weights only to construct a weighted ATT from the set of individual, municipality-specific treatment effects. See Borusyak et al. (2021) for details. Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Table A4. Robustness to placebo treatment years.

| | Log fertility rate | | | | | |
|-------------------------------|--------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Placebo year (h): | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Teen \times SPP \times Post$ | -0.026 (0.018) | -0.021 (0.016) | -0.004 (0.014) | -0.011 (0.016) | -0.015 (0.017) | -0.015 (0.018) |
| Observations | 14,854 | 14,854 | 14,854 | 14,854 | 14,854 | 14,854 |
| Treatment municipalities | 540 | 540 | 540 | 540 | 540 | 540 |
| Comparison municipalities | 521 | 521 | 521 | 521 | 521 | 521 |

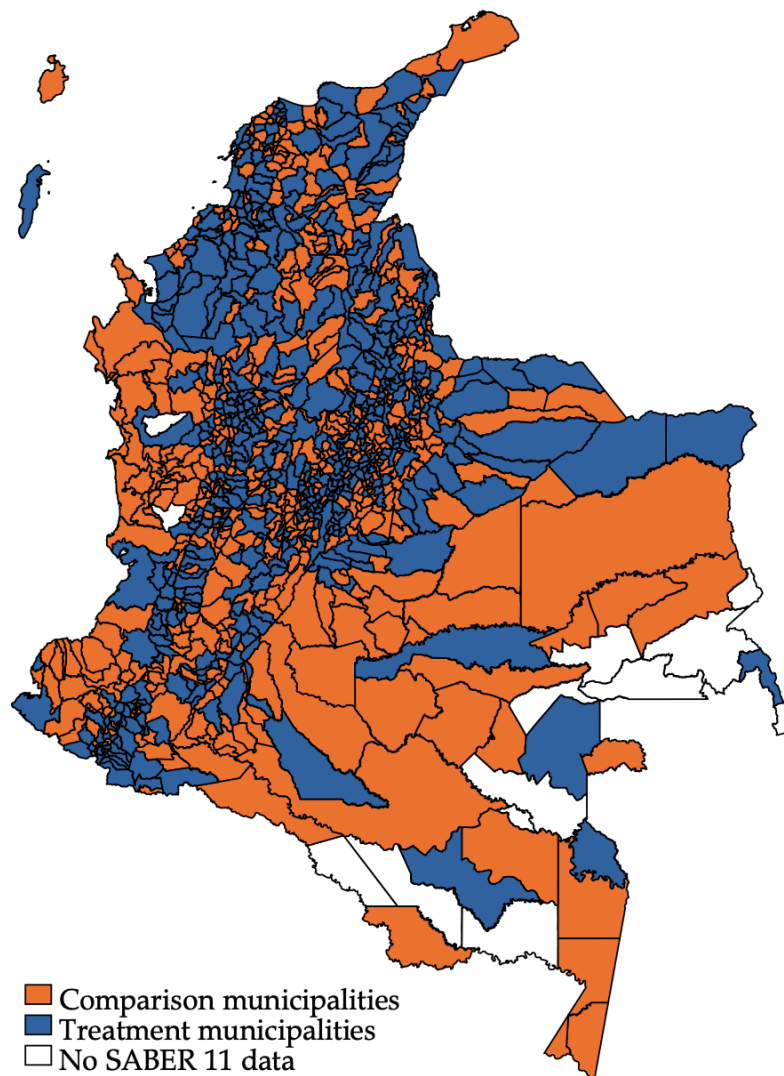
Notes: Each column in this table assumes that SPP was introduced in year h instead of 2014 and estimates Equation 2 with $Post_t = \mathbb{1}[t > h]$ and $t \in [2008, 2014]$. All estimates are weighted by the number of births between 2008 and 2014 for each age group in each municipality. Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Figure A1. Distribution of SPP eligibility rates.



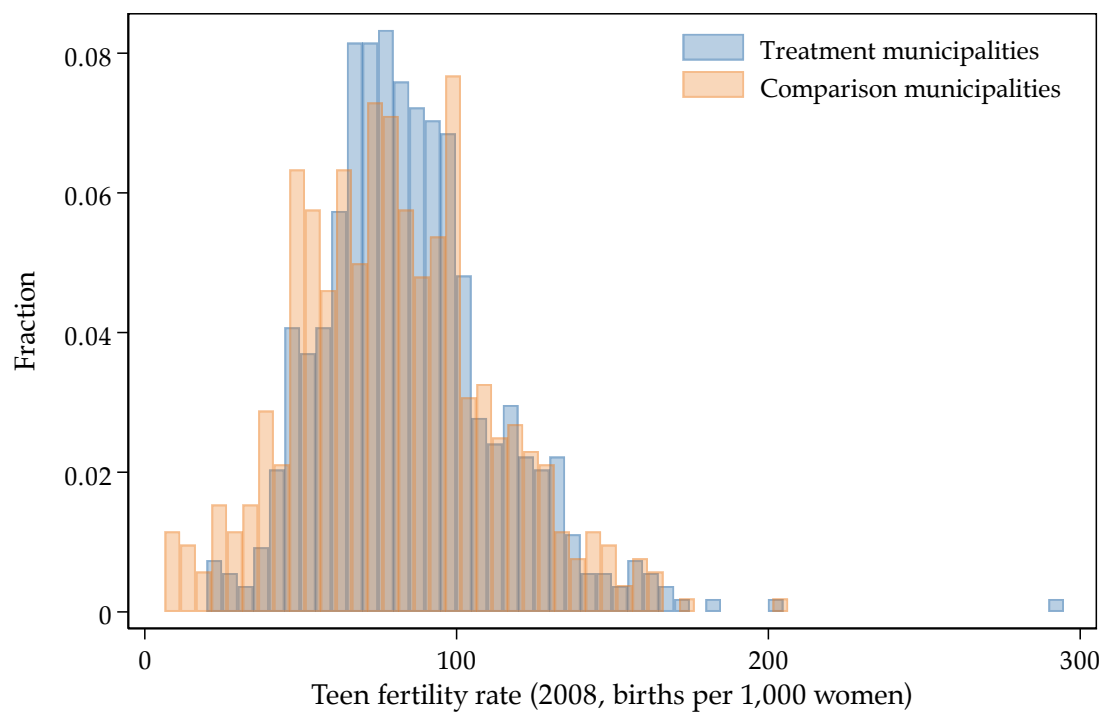
Notes: This histogram shows the distribution of SPP eligibility rates in 2014 for the 1,061 municipalities in our final sample. The minimum value is 0 and the maximum value is 250 (eligible female students per 1,000 female students). The 95th percentile is 42.

Figure A2. Municipality-level variation in SPP eligibility.



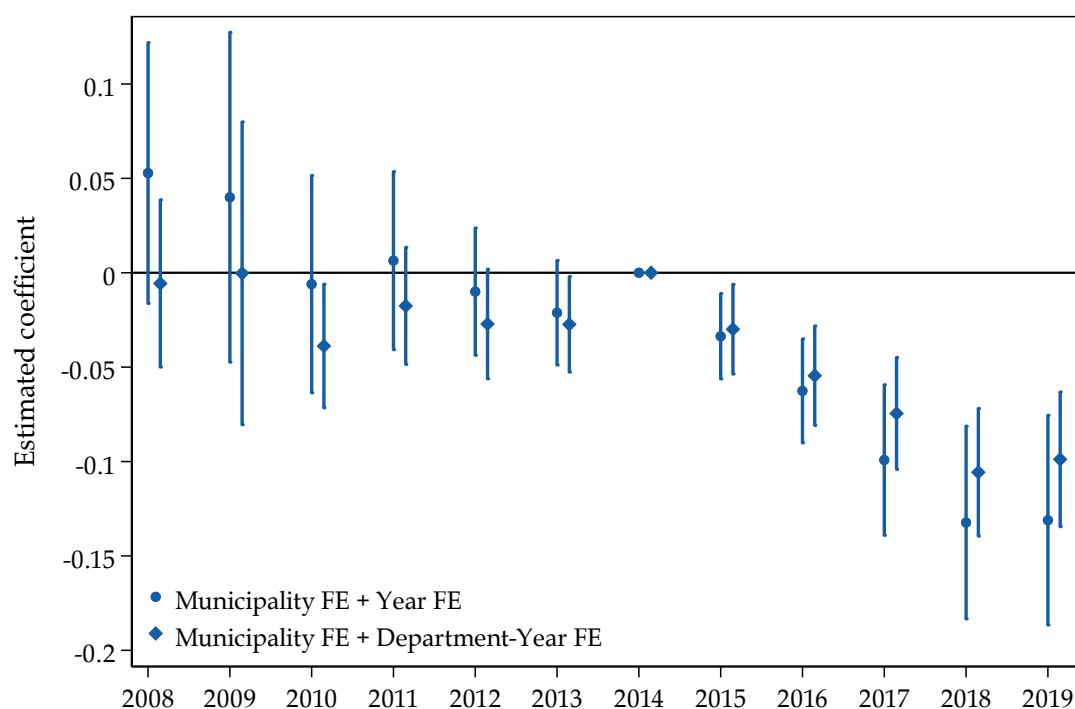
Notes: This map displays the geographic distribution of treatment and comparison municipalities. Treatment municipalities are above the median in female eligibility rates for SPP in 2014, while comparison municipalities are below the median.

Figure A3. Distribution of teen fertility rates.



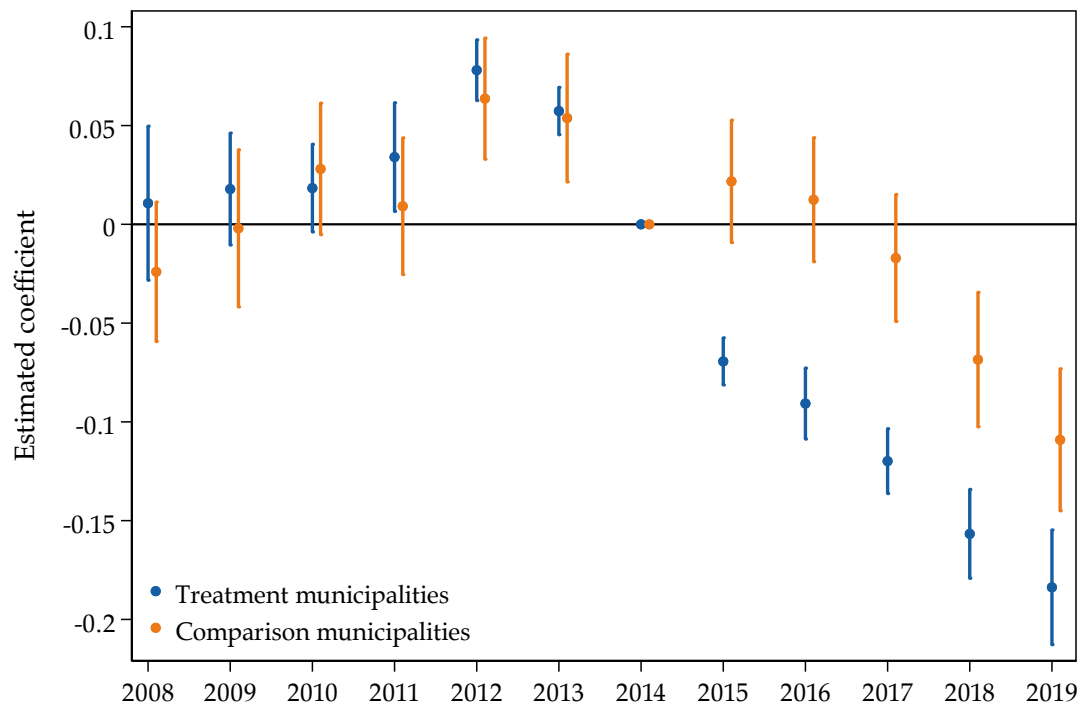
Notes: This histogram shows the distribution of teen eligibility rates in 2008 for the 1,061 municipalities in our final sample separately for treatment and comparison municipalities. Treatment municipalities are above the median in female eligibility rates for SPP in 2014, while comparison municipalities are below the median.

Figure A4. Difference-in-differences event study estimates for 15-19 years old.



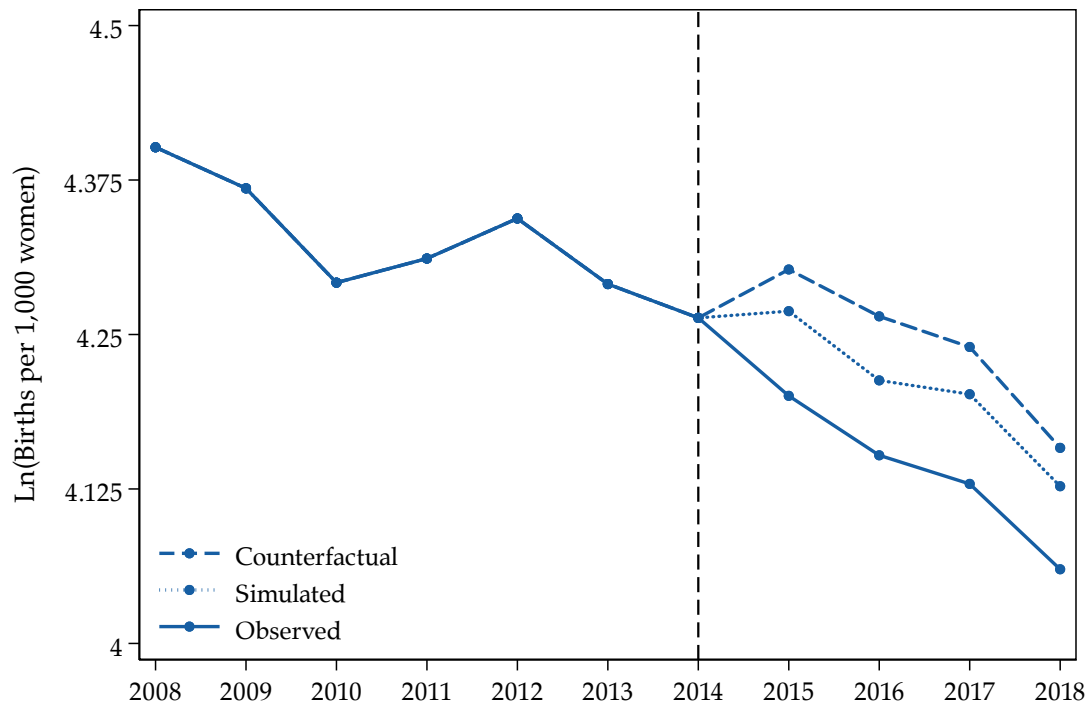
Notes: This figure presents difference-in-differences event study estimates comparing the log fertility rate of 15-19 years old in treatment and comparison municipalities. Dots and diamonds represent estimated coefficients, and vertical lines represent 95 percent confidence intervals. Departments in Colombia are similar to states in the United States. All estimates are weighted by the number of births between 2008 and 2014 in each municipality. Standard errors are clustered at the municipality level. We reject the null hypothesis of no pre-trends (p -value<0.05).

Figure A5. Difference-in-differences event study estimates by treatment and comparison municipalities (15-19 vs. 25-29 years old).



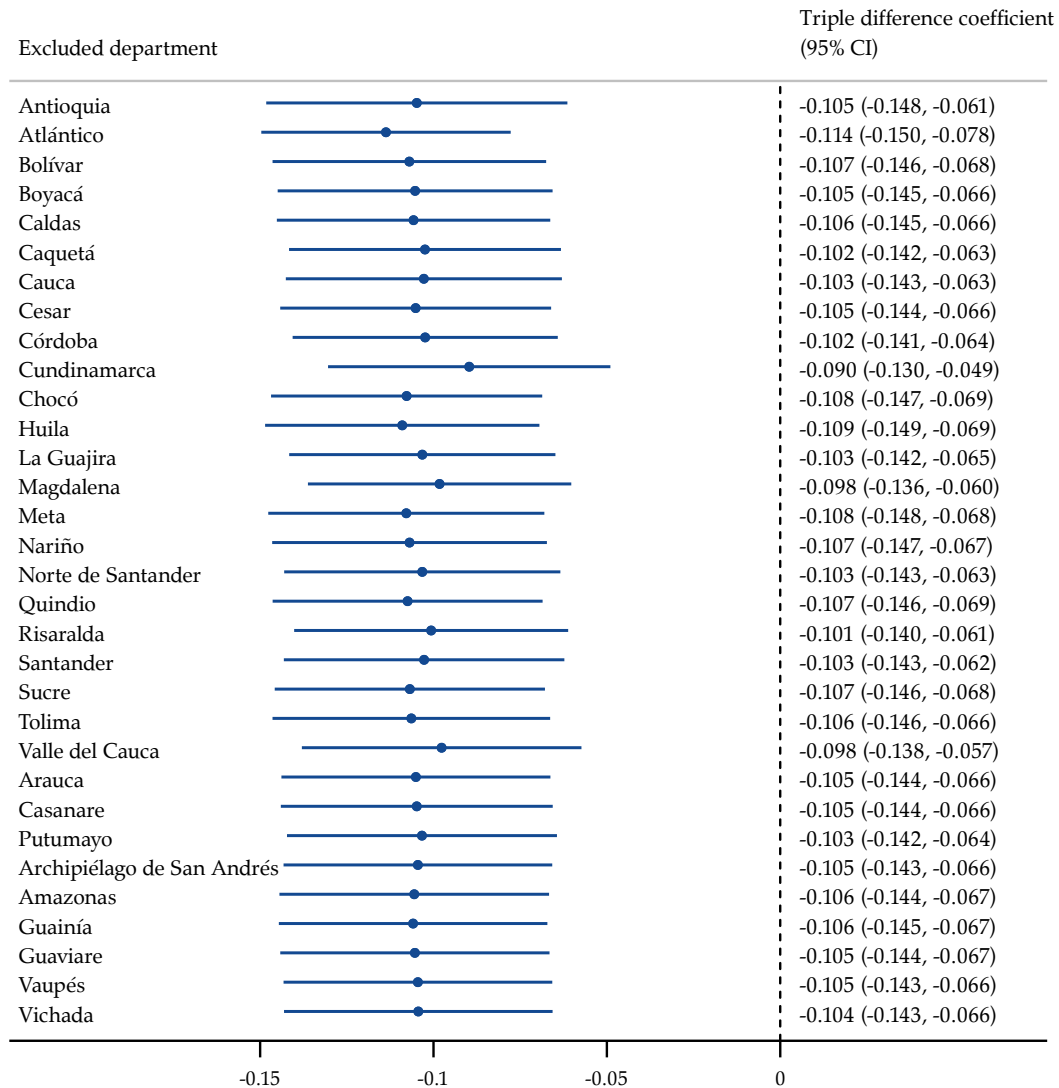
Notes: This figure presents difference-in-differences event study estimates comparing the log fertility rate of 15-19 years old to 25-29 years old separately for treatment and comparison municipalities. Dots represent estimated coefficients and vertical lines represent 95 percent confidence intervals. The regressions include municipality by age group fixed effects and municipality by year fixed effects. All estimates are weighted by the number of births between 2008 and 2014 for each age group in each municipality. Standard errors are clustered at the municipality level.

Figure A6. Simulation exercise: Teen fertility rate in treatment municipalities.



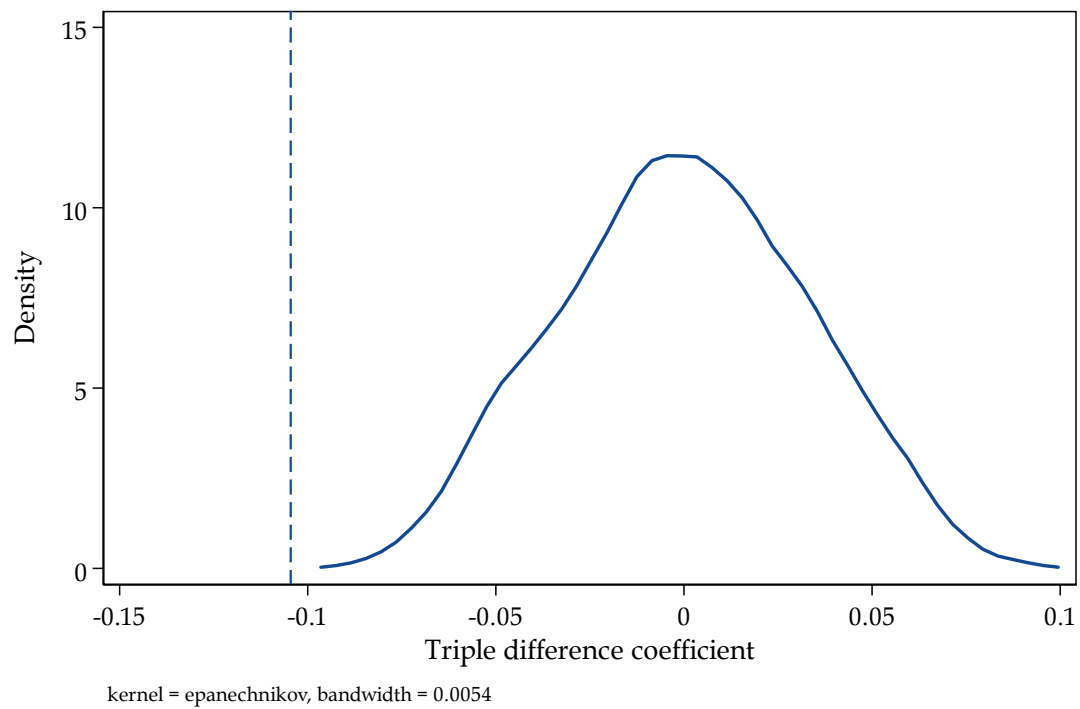
Notes: This figure presents the result of a simulation exercise in which we first impute counterfactual fertility rate trends for teenagers in treatment municipalities using the specification in [Equation 1](#) and the imputation estimator derived in [Borusyak et al. \(2021\)](#). Then, we subtract from these counterfactual trends the number of births implied by an assumption that all SPP-eligible female students in each year would have experienced a birth and compare this “simulated” fertility rate trend to the actual trend we observe. The solid line represents the observed fertility rate trend in treatment municipalities. The dashed line represents the estimated counterfactual fertility rate trend. Finally, the dotted line shows the simulated fertility rate trend if effects were only driven by SPP-eligible females.

Figure A7. Robustness to excluding all municipalities in a given department.



Notes: This figure reports triple difference event study estimates of β from [Equation 2](#) where a single department is excluded in each regression. Dots represent estimated coefficients and horizontal lines represent 95 percent confidence intervals. All estimates are weighted by the number of births between 2008 and 2014 for each age group in each municipality. Standard errors are clustered at the municipality level.

Figure A8. Randomization Inference (RI): Distribution of placebo treatments.



Notes: This figure presents the distribution of placebo treatments after 5,000 random permutations of the treatment assignment (i.e., we randomize municipalities to be treatment or comparison municipalities). We run the regressions using our summary specification in [Equation 2](#). The vertical dashed line represents the original estimated coefficient. RI-based p-value = 0.000. The procedure was implemented using the routine by [Heß \(2017\)](#).

B Validation of Treatment Intensity Measure

We attempt to probe the validity of our treatment intensity measure by estimating whether it is associated with an increase in SABER 11 test scores after SPP is introduced. This is essentially testing whether we can replicate the results from [Bernal and Penney \(2019\)](#) and [Laajaj et al. \(2018\)](#) using our treatment measure. We use individual-level test scores on the SABER 11 exam for female students (15-19 years old) between 2010 and 2016.¹⁵ We use a triple difference empirical approach that leverages the same municipality-level variation in SPP eligibility as in our fertility analysis (see [Equation 1](#)) and variation between students who are eligible for SPP on the SISBEN margin and those who are not.

Specifically, we estimate the following equation by OLS:

$$\begin{aligned} StdTestScore_{imt} = & \zeta + \phi \left(SISBEN_i^{1-2} \times SPP_m^* \times Post_t \right) \\ & + \psi_1 \left(SISBEN_i^{1-2} \times \theta_m \right) + \psi_2 \left(\theta_m \times \lambda_t \right) \\ & + \psi_3 \left(SISBEN_i^{1-2} \times \lambda_t \right) + v_{imt}, \end{aligned} \quad (B1)$$

where i denotes a student, m denotes a municipality, and t denotes year. The $StdTestScore_{imt}$ variable is students' SABER 11 test score standardized by test year.¹⁶ $SISBEN_i^{1-2}$ indicates whether the student is categorized as SISBEN level 1 or 2. A SISBEN level of 1 or 2 is roughly equivalent to being eligible for SPP on the SISBEN margin, whereas students with higher SISBEN levels or not categorized are ineligible. SPP_m^* denotes treatment and comparison municipalities and is defined as $SPP_m^* = \mathbb{1} [SPP_m > \text{median}(SPP_m)]$ with SPP_m being the rate of female students eligible for the program in a given municipality. In the standard way for triple difference specifications, we include the three two-way interactions between fixed effects for SISBEN levels ($SISBEN_i^{1-2}$), municipalities (θ_m), and years (λ_t). The

¹⁵A consistent SISBEN level variable is only available in the SABER 11 data for these years. The SISBEN level is self-reported by the student. We use data from fall semesters.

¹⁶The overall individual SABER 11 score is a linear combination of scores in different subjects. We follow ICFES and [Londoño-Vélez et al. \(2020\)](#) (see Online Appendix) and calculate this individual score as follows:

$$TestScore_i = \frac{Chem_i + Bio_i + Phys_i + 2SocSci_i + Philo_i + 3Lang_i + 3Math_i + Eng_i}{13} \text{ for 2010-2013, and}$$

$$TestScore_i = 5 \times \left(\frac{3Math_i + 3Reading_i + 3NatSci_i + 3SocSci_i + Eng_i}{13} \right) \text{ for 2014-2016.}$$

We then normalize these scores by year using the mean and standard deviation from the whole sample of students (males and females) in each year.

vectors ψ_1 , ψ_2 , and ψ_3 contain the parameters associated with these interactions. Finally, v_{imt} is an error term. In [Equation B1](#), ϕ is our parameter of interest, measuring the effect of SPP on test scores.

Table B1. Triple difference estimates on SABER 11 test scores.

| | Standardized SABER 11 test score (1) |
|---------------------------------------|---|
| $SISBEN^{1-2} \times SPP \times Post$ | 0.052** (0.024) |
| Observations | 1,862,511 |
| Treatment municipalities | 540 |
| Comparison municipalities | 521 |
| Pre-trends test p -value | 0.690 |
| Pre-SPP socioeconomic achievement gap | 0.700 |

Notes: The table above reports triple difference estimates of ϕ from [Equation B1](#). Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

[Table B1](#) presents the results from this regression. We find that, after the introduction of SPP, test scores increase for SISBEN-eligible female students in treatment municipalities by 0.052 standard deviations relative to comparison municipalities. This increase represents about 7.5 percent of the pre-SPP test score gap between SISBEN levels 1-2 and higher SISBEN levels. This estimate is smaller but qualitatively comparable with the estimates in [Bernal and Penney \(2019\)](#) and [Laajaj et al. \(2018\)](#).¹⁷

¹⁷Since [Bernal and Penney \(2019\)](#) and [Laajaj et al. \(2018\)](#) use a regression discontinuity design, their estimates reflect local average treatment effects. Since our triple difference estimates represent average treatment effects, it is reasonable to expect our estimates to be somewhat smaller.

C Robustness to Possibly Confounding Events

This appendix provides a full description of the analyses we conduct to assess whether events and policies that occurred around the time SPP was introduced are driving our results. We consider three events: 1) the unilateral permanent ceasefire by the Revolutionary Armed Forces of Colombia (FARC, from the Spanish acronym) in December 2014 as part of the by then ongoing peace process between the guerrilla group and the Colombian government, 2) the Zika virus epidemic, which occurred from October 2015 to July 2016, and 3) the *Jornada Única* initiative, which gradually transitioned some public secondary schools that were operating half-day shifts into full school days beginning in 2015.

[Guerra-Cújar et al. \(2020\)](#) finds evidence that the peace agreement with FARC led to a “baby boom” in municipalities that experienced more FARC conflict before the peace agreement, and other studies find effects of the peace agreement on educational outcomes and deforestation ([Prem, Vargas and Namen, 2021](#); [Prem, Saavedra and Vargas, 2020](#)). While [Guerra-Cújar et al. \(2020\)](#) find that the relative increase in fertility rates does not seem to be driven by any particular age group, we assess whether the effects of the FARC peace agreement are driving our results. We use data from [Prem et al. \(2020\)](#) on the locations of FARC presence in the years before the ceasefire and estimate our main specification with the subset of municipalities that did not experience any FARC-related violence in the period 2011–2014. We report these results in column (2) of [Table C1](#). The triple difference estimate for this subset of municipalities is -0.109, nearly identical to the estimate with all municipalities.

The Zika virus can be spread from a pregnant woman to her baby, which can result in birth defects. [Gamboa and Rodríguez-Lesmes \(2019\)](#) studies the effect of the Zika virus epidemic in Colombia on birth rates, finding a 10 percent decline. We assess whether the Zika virus could be driving our results by estimating our main specification on the subset of municipalities that have an altitude of at least 1,800 meters above sea level, where mosquitos—which transmit the virus—are less prevalent. These results are reported in column (3) in [Table C1](#). Our estimate for this subset of municipalities is -0.137, even larger than our main estimate. Together, these results indicate that our estimated teen fertility impacts of SPP are not driven by the FARC ceasefire or the Zika virus epidemic.

Finally, since 2015 the Colombian Ministry of Education has been gradually

Table C1. Robustness to possible confounding events.

| Municipalities: | Log fertility rate | | | |
|------------------------------------|----------------------|----------------------|----------------------|--------------------------------|
| | All (1) | No FARC (2) | High altitude (3) | No <i>Jornada Única</i> (4) |
| <i>Teen × SPP × Post</i> | -0.105*** (0.020) | -0.109*** (0.022) | -0.137*** (0.031) | -0.073*** (0.016) |
| Observations | 25,464 | 22,296 | 6,696 | 15,816 |
| Treatment municipalities | 540 | 480 | 152 | 291 |
| Comparison municipalities | 521 | 449 | 127 | 368 |
| Pre-trends testing <i>p</i> -value | 0.232 | 0.476 | 0.290 | 0.582 |

Notes: Column 1 reproduces the main results. Column 2 reports results from our summary specification in [Equation 2](#) for municipalities that did not experience any violent events by FARC from 2011 to 2014 using data from [Prem et al. \(2020\)](#). Column 3 reports results from our summary specification in [Equation 2](#) for municipalities that are located above 1,800 meters above sea level and therefore have a low likelihood of Zika infection. Column 3 does not include Bogotá, the country's capital city, because of its relative size compared to others high altitude municipalities. The coefficient with Bogotá is -0.152 (0.025). Column 4 excludes municipalities in which female students were exposed to *Jornada Única* ("Full School Day") at any point between 2015-2018. All estimates are weighted by the number of births between 2008 and 2014 for each age group in each municipality. Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

implementing an initiative to transition public schools from half-day shifts (morning and afternoon) to full school days to extend the duration and the quality of instruction. This policy is called *Jornada Única* or "Full School Day."¹⁸ There is evidence from other contexts that lengthening the school day can reduce adolescent pregnancies via an incapacitation effect ([Berthelon and Kruger, 2011](#)). Accordingly, we also test that our results are robust to the expansion of *Jornada Única*. Given the high costs associated with this strategy, its expansion has been very gradual over time and was not adopted in all municipalities during our period of interest. In 2015, less than 0.04 percent of the female students taking the SABER 11 test attended a school with *Jornada Única*. This share increased to 0.46 percent in 2016, 6 percent in 2017, and 8 percent in 2018. We, therefore, do not expect this policy to explain the sharp decline in teen fertility observed right after the introduction of SPP in 2014. Column 4 of [Table C1](#) corroborates this. It presents our summary

¹⁸See [Hincapié \(2016\)](#) for a review of the length of the school day in Colombia around the time of the implementation of *Jornada Única*. But, shortly, in many public schools, two different groups of students attend the same institution (i.e., use the same physical infrastructure), one in the morning and one in the afternoon. So there are two "shifts," particularly in schools serving basic secondary (grades 6 to 9) and mid secondary (grades 10 and 11) students.

triple difference estimate excluding the municipalities in which female students were exposed to *Jornada Única* at any point between 2015-2018. We still find a big, negative and significant impact of SPP on the sample of municipalities not exposed to full-day shifts due to the *Jornada Única* initiative.