Health at birth and the dynamics of racial inequality in human capital formation in Brazil

Alysson Lorenzon Portella
Insper Instituto de Ensino e Pesquisa
PhD Candidate
alyssonlp@al.insper.edu.br

Naercio Aquino Menezes-Filho
Insper Instituto de Ensino e Pesquisa
Full Professor

July 21, 2019

Abstract: We investigate the dynamics of racial inequalities in human capital formation in Brazil using data on children from their birth to 22 years old. We document racial inequalities in birth outcomes, years of schooling and socioemotional skills. Particularly for years of completed schooling, we observe increasing black-white unconditional gaps that start in half an year of schooling when children are 11 years old and grow to reach 1.3 years when children are 22 years old. When we condition the gap on previous schooling, together with a host of other control variables, we only observe significant gaps at ages 11 and 15, but no racial gaps at ages 18 and 22, which is evidence that the gap is formed early in life. The most important variables explaining the racial gap in schooling are family income, maternal education, number of prenatal visits, maternal smoking during pregnancy, and maternal mental health.

Key-words: Racial inequality, education, birth weight, human capital formation

Resumo: Neste trabalho, investigamos a dinâmica das desigualdades raciais na formação de capital humano no Brasil a partir de dados sobre crianças desde o nascimento até os 22 anos de idade. Documentamos desigualdades raciais em resultados ao nascer, em anos de estudo e habilidades socioemocionais. Em particular, observamos um aumento contínuo no hiato de anos de escolaridade entre brancos e negros que começa em meio ano de estudo aos 11 anos de idade e aumenta para 1,3 anos aos 22 anos. Quando o hiato é condicionado em escolaridade prévia e outras covariadas, ele continua significativo apenas aos 11 e 15 anos de idade, não sendo significativo aos 18 e 22 anos, indicando que o hiato racial é formado nos primeiros anos da infância e adolescência. As variáveis mais importantes para explicar o hiato racial em escolaridade são renda familiar, educação materna, número de visitas prenatal, tabagismo durante a gestação e saúde mental materna.

Palavras-chave: Desigualdade racial, educação, peso ao nascer, formação de capital humano

Área ANPEC: Área 12 - Economia Social e Demografia Econômica

Classificação JEL: I15, I24, J15

1 Introduction

Although many studies still find evidence of discriminatory behavior in the labor market[1], some authors highlight the importance of differences in skills and schooling as one of the major determinants of the wage gap between black and white workers. In the United States, both Heckman (2011) and Fryer (2011) suggest that public policy aiming at reducing racial economic disparities should focus on closing the gap in skills. Similarly, the Brazilian literature have pointed out that, contrary to what happens with respect to gender gaps in wages, a considerable part of differences in wages between black and white workers can be explained by observable differences in attributes, including schooling and test scores (Oliveira and Rios-Neto, 2006; Matos and Machado, 2006; Arraes et al., 2014; Curi and Menezes-Filho, 2014). A better understanding of how racial differences in skills and human capital emerge is thus crucial for reducing inequality in the labor market.

The Brazilian literature on human capital formation has collected evidence that there exist racial inequalities between black and white students in school attainment and test scores that do not vanish even when we control for a large set of covariates. For instance, Barros et al. (2001) find that Brazilian black students have 0,54 less years of schooling than their white counterpart. Menezes-Filho (2007) and Matos and Machado (2006) also find that black students have lower performance in math test score. Portella et al. (2017) show that black students have larger age-grade distortion and higher chances of ever failing a grade. These studies, however, only include a racial dummy as a control variable and do not deal particularly with issues of race nor do they try to explain the sources of these racial inequalities.

Among the studies that try to uncover the reasons why Brazilian black students perform worse than white students are Soares and Alves (2003) and Soares (2006), who find that white students benefit more from better school structure. Simões and Ferrão (2005) observe a negative correlation between performance and self-evaluation of students. Since black students have lower self-evaluation, they point out that this can be a factor behind their relative poorer performance. Flores and Scorzafave (2014) find evidence that segregation and discrimination play a role in determining the lower scores of black students. Botelho et al. (2015) compare test scores of student in their schools with their scores in color-blind, standardized exams. They find a bias in grading by teachers against black students. This bias decrease with longer contact between student and teacher, supporting the existence of statistical discrimination. Rangel (2015) studies the possible effect that anticipated market discrimination can have on parental investments in their children's human capital. Using a within-family approach he finds that darker-skin children have lower chances of being enrolled in school between 5 and 14 years old. For co-residing children between 17 and 25 years old, darker-skin children are 5.7% less likely to have finished primary school, and the affect seems to be larger for boys, in line with observable returns to schooling. Marteleto and Dondero (2016) follow a similar strategy of observing variation within family, although they consider the case of twins between 12 and 18 years old that differ in their racial classification. Their twins fixed-effects estimation shows that children with darker complexion have on average 0.3 less years of study than their white siblings. Again, the effect is larger among boys.

Despite having different racial ideology than Brazil[2], the literature on racial inequalities in the United States may offer some insights on the possible reasons that help explain differences between black and white students. One important finding of this literature is that racial gaps in skills emerge quite soon in early childhood and widens during the first years of schooling. This, in turn, hinders long-term investment in human capital later in life. For instance, Fryer and Levitt (2013) observe differences in mental ability between black and white infants as young as 2 years old. Fryer and Levitt

[1] Audit and correspondence studies show that black applicant have lower return calls than their white counterparts. See for instance Heckman (1998); Guryan and Charles (2013). Hirata and Soares (2016) find evidence of taste discrimination in Brazil.

[2] See Telles (2004) for a comparison between the two countries.

(2004) are able to close the gap in cognitive tests for children enrolled in kindergarten after controlling for a series of covariates, but show that the gap starts growing after first grade, widening throughout the schooling years. Carneiro et al. (2005) summarize a series of papers showing that the racial gap in skills emerge early in life and present novel evidence on this matter for both cognitive and non-cognitive skills for the United States. Todd and Wolpin (2007) estimate a production function that allow for lagged inputs, parents' ability and unobserved endowments to affect cognitive achievement. They use their model to study racial gaps in test scores and find that mother's AFQT[3] account for most of the gap, together with differences in home inputs. Moreover, equalizing home inputs would close the gap in math and reading scores by 10-20%. Finally, Cameron and Heckman (2001) use a dynamic model of schooling choice to evaluate how income and family characteristics in early stages of life affect the chances of minorities students entering college in the United States. They find that the gap between races is closed and sometimes even reverted when the endowments of black students is equalized to that of white students. Including AFQT scores reduces the importance of family characteristics, particularly in later years of schooling. They conclude that short-term constraints are not as important in determining entrance in college as the long-term influence that family income and characteristics has through the formation of skills that prepare children to college.

Similarly, differences in health in early childhood may also explain part of the gap between black and white children. Chay et al. (2009, 2014) argue that improvements in health services available to black people during the 1960's were responsible for the convergence in test scores observed in the 1980's. The end of segregation in Southern states made better hospitals available to black people, leading to better health indicators for newborns. The improvement in test scores in the 1980's was concentrated precisely in those states and cannot be explained by other factors, such as school quality or family environment.

The main objective of this paper is to shed light on the dynamics of racial gaps in human capital in Brazil. We use a unique database that follows children from birth until 22 years old to investigate the evolution of human capital formation and how it differs between black and white children. We evaluate the relative importance of *in uterus* investments, birth outcomes, maternal psychological health, and previous achievement in explaining black-white gaps in schooling, from adolescence through early adulthood. Black and white children already differ when they are born, with black newborns weighting on average around 80 grams (0.15 standard deviations) less then white newborns. Differences in observables explain almost half of this gap.

We document growing racial gaps in years of schooling as children age. The unconditional gap in schooling is half a year when children are 11 years old (.41 standard deviations - SD) and grow to 1.3 years of schooling (.44 SD) when they are 22 years old. This increasing absolute difference between black and white children hide two important caveats: i) the normalized unconditional racial gap[4] is quite stable throughout the period, around 0,4 for all ages; ii) conditional on previous school attainment and other covariates, the racial gap cease to be significant after age 15, meaning that black-white differences in schooling are mostly determined up to that age. We also find racial gaps in socioemotional skills at age 11, measured using the Strength and Difficulties Questionnaire (SDQ), but we are not able to assess its dynamics since we only observe this score one period.

Previous education is the main determinant of years of education, highlighting the cumulative nature of skills and the importance of early investments. Families' socioeconomic status (SES) at the time children were born and when they are 11-years-old are the main factors behind the differences in schooling through their impact on years of schooling at age 11. *In uterus* investments and mother's psychological health also play a role in explaining these differences. Although there exist unexplained racial inequalities in birth weight, and birth weight is correlated with schooling when children are 11, it does not seem to account for much of the racial gap in schooling.

[3]Armed Forces Qualification Test. This is a test widely used in research as an indicator of cognitive skills in the United States.

[4]That is, when we normalize years of schooling at each year to have mean zero and standard deviation one.

We also investigate whether the dynamics of the gap vary according to families SES, based on *per capita* family income and maternal education by the time of delivery, and by gender. We observe that the dynamics of the gap change across socioeconomic groups, with significant conditional racial gaps among children in high favorable at the age 22, while such conditional gap is not significant for the whole population. We interpret these results as suggesting that black children may face barriers to finish high school and enter college that are not present for white children, and it is only significant among those with high SES because only families in the highest socioeconomic group are able to support the studies of their children past high school. Moreover, the racial gap grows much faster for boys between 11 and 15 years old, whereas for girls the racial gap grows continuously, although it reaches the same level as among boys by the age of 22.

This paper proceeds as follows. Section 2 presents the data and discuss the theoretical background in which we conduct our empirical analysis. In Section 3 we discuss the main results, starting with racial differences in birth outcomes, then moving to differences in schooling and SDQ scores, and finally evaluating the evolution of racial gaps from 15 to 22 years. In section 4 we evaluates possible heterogeneities in the dynamics of the black-white gap in attained education based on family SES and gender. Finally, in Section 5 we conclude and discuss possible policy implication.

2 Methods and data

2.1 Data

Data comes from the Pelotas' 1993 Cohort Study, collected by the Research Center of Epidemiology from the Universidade Federal de Pelotas[5]. These data follow every children born in the town of Pelotas in the year of 1993. They contain information collected on many rounds of interviews with more than 5 thousand children since they were born up to 22 years old. The variables cover a wide range of themes, including mother's habits during pregnancy, family characteristics, physical and psychological health indicators, information on education and labor market participation, among other topics. Here we limit our analysis to five rounds of this survey: soon after children were born (perinatal) and when children were 11, 15, 18, and 22 years old[6]

From the round of interviews in 2004 (when children were 11 years old) we obtain information regarding children's self-identified racial identity. Each children interviewed had to select one of the following items as their race or skin color: a) white, b) black, c) brown, d) yellow (Asian), and e) indigenous[7]. We focus exclusively on children who classify themselves as being white, black or brown, and join together the black and brown categories into a non-white category. This restriction leaves us with a sample of 4212 children out of the original 5249, as can be seem in Table 1. This table also displays the sample sizes under many criteria, together with the relative size of this sample with respect to the original number of observations. Due to attrition, 27% of the original sample is lost when children complete 22 years. This value increases to one third when we exclude individuals that self-classified as either yellow or indigenous. Finally due to missing values in variables included in the regressions, the sample is reduced by around 50%.

Table 2 provides descriptive statistics regarding differences between white and non-white individuals in the cohort of 1993. We observe that non-white children lag behind in every measure of human capital displayed. This includes birth weight, scores from the Strength and Difficulties Questionnaires (SDQ)[8], and years of completed schooling. Black children are born nearly 90 grams

[5]Details of this research are available in Victora et al. (2006).

[6]We also have access to data from surveys conducted when children were 6, 12 and 48 months old, but these data is available only for 20% of the total sample.

[7]This is the same racial classification used by Brazil's Bureau of Statistics (IBGE).

[8]The Strength and Difficulties Questionnaire provides information on children's mental health. It is divided in five scales: conduct problems, hyperactivity, emotional problems, peer relations, and pro-social behavior. The first

Table 1: Sample sizes according given criteria

Age	Full Sample	Percent	Race Dummy	Percent	Analytic Sample	Percent	Common Sample	Percent
Birth	5249	100%	4212	80%	3464	66%	2447	47%
11	4452	85%	4212	80%	3382	64%	2447	47%
15	4349	83%	4033	77%	3159	60%	2447	47%
18	4106	78%	3766	72%	2786	53%	2447	47%
22	3813	73%	3443	66%	2487	47%	2447	47%

Full Sample refers to all individuals that participated in each round. Race Dummy refers to the remaining individuals that self-classified as being white, brown or black. Analytic Sample refers to observations used in the regressions of outcomes from that particular round of interviews, thus excluding observations with missing values on the dependent or independent variables. Finally, Common Sample contains only observations with no missing value in any dependent or independent variables used in every regression in this work.

lighter and by age 11 they have 0.28 SD lower scores on SDQ. Absolute differences in years of schooling grow with time, from half a year at age 11 to 1.3 years when they are 22 years old. However, these racial differences in schooling are almost stable in terms of standard deviations, varying from 0.46 SD at 11, to 0.57 SD when they are 15, then dropping to 0.48 SD when they are 22 (using as reference the standard deviation among white children). At the same time that we observe considerable differences in outcomes between non-white and white children, we also observe differences in backgrounds between them. Even before black children are born, they are already disadvantaged, as we can see by the number of prenatal appointments mothers had and whether or not they smoked cigarettes during pregnancy. Mothers of black children are also younger, have lower education and report more mental health problems (measure by the Self-Reporting Questionnaire with 20 items - SRQ-20[9]). Families of black children are larger and poorer than their white counterparts. Although we do not include in our descriptive analysis, we also have data on mother's health during pregnancy (whether or not she had any abortion threat, anemia, urinary infection, high pressure, or diabetes), number of previous children (parity), whether any previous birth was low birth weight (LBW) or premature, mother's height and body mass index (BMI) before the pregnancy, gestational weight gain, whether the mother worked during pregnancy, whether any partner lived with her during pregnancy and when the child was 11 years old, and hospital of delivery. Other outcomes at birth apart from weight are also available, such as baby's length, head, thoracic, and abdominal circumferences, gestation length, Dubowitz score[10], and whether the child was taken to a intensive unit care.

2.2 Theoretical Framework

This paper uses as main theoretical background the idea of cumulative formation of skills, expressed in works such as Cunha and Heckman (2007) and Todd and Wolpin (2003, 2007). The literature on skills and human capital formation has departed from the idea of innate and immutable skills that children are born with and now considers skill formation as a life cycle process that starts in the womb and continues throughout childhood and adolescence. Childhood is not considered to be a one-stage period in which investments are made, but as a multi-stage period that allows for

four scales made up total scores. The first two scales can also be grouped to form the externalizing subcomponent, whereas emotional problems and peer relations can be grouped to form the internalizing subcomponent. Each scale has five questions and scores are computed by adding answers. Here, we normalize total and subcomponents scores to have mean zero and standard deviation 1, with positive values meaning better mental health. This questionnaire was applied when children were 11 years old, both to them directly and to their mothers. For more details, see Goodman et al. (2010)

[9]The SRQ-20 is an instrument used since the 1980's to assess mental disorders in adults. See Harding et al. (1980).

[10]The Dubowitz scores is a measure of gestational age.

Table 2: Descriptive statistics for selected variables

	(1) White		(2) Non-white		Difference
Variable	N	Mean/SD	N	Mean/SD	(1)-(2)
Birth Weight	2947	3204.989	1258	3116.923	88.067***
		(520.609)		(514.711)	
Years of Schooling at 11	2951	3.674	1258	3.173	0.501***
		(1.077)		(1.186)	
Years of Schooling at 15	2803	6.686	1203	5.761	0.924***
		(1.632)		(1.766)	
Years of Schooling at 18	2622	8.995	1137	7.850	1.145***
		(2.152)		(2.250)	
Years of Schooling at 22	2375	10.491	1064	9.146	1.345***
		(2.817)		(2.719)	
Normalized SDQ Total Score	2857	0.094	1216	-0.189	0.283***
		(0.985)		(0.995)	
Number of Prenatal appointments	2949	8.220	1256	6.660	1.560***
		(3.501)		(3.591)	
Mother Smoked During Pregnancy	2953	0.319	1259	0.373	-0.054***
		(0.466)		(0.484)	
Parity in 1993	2953	1.060	1259	1.381	-0.322***
		(1.310)		(1.595)	
Mother's Schooling	2948	7.143	1257	5.636	1.507***
		(3.598)		(2.979)	
Mother's Age	2953	26.295	1259	25.385	0.909***
		(6.398)		(6.364)	
Per Capital Family Income in 1993	2951	551.776	1258	316.866	234.910***
		(856.754)		(370.302)	
Per Capital Family Income in 2004	2937	482.645	1252	249.620	233.025***
		(1053.162)		(675.378)	
Mother's SRQ-20	2920	0.302	1248	0.338	-0.036***
		(0.205)		(0.216)	
Per Capital Family Income in 2011	2628	792.761	1138	476.072	316.689***
		(1071.091)		(773.212)	
Studying at 18	2628	0.561	1138	0.520	0.040**
		(0.497)		(0.502)	
Works at 18	2628	0.759	1138	0.804	-0.045***
		(0.428)		(0.397)	
Real Income at 18	1792	655.589	816	589.087	66.502***
		(575.720)		(342.436)	
Has child at 18	2628	0.075	1138	0.143	-0.068***
		(0.264)		(0.350)	
Studying at 22	2375	0.396	1064	0.267	0.129***
		(0.499)		(0.445)	
Works at 22	2208	0.688	1019	0.644	0.045**
		(0.463)		(0.479)	
Real Income at 22	1532	1369.566	673	1111.872	257.694***
		(1118.139)		(616.649)	
Has children at 22	2377	0.227	1066	0.395	-0.168***
		(0.419)		(0.489)	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

self-productivity in the process of forming skills, that is, previously accumulated skills foster the formation of new skills. Unlike previous works on economics that used a scalar to summarizes people's abilities, now the literature highlights the importance of multiple skills in many life outcomes, particularly cognitive and non-cognitive skills and health (Cunha et al., 2006; Glewwe and Miguel, 2008; Currie and Almond, 2011). These different skills may also help in the development of other skills, in what is called dynamic complementarities. For instance, children with less non-cognitive skills may find it more difficult to focus in classroom and thus acquire less cognitive skills (Cunha et al., 2006; Cunha and Heckman, 2007).

Next, we follow closely Cunha and Heckman (2007) in developing a simple model that guides our empirical analysis and interpretation of results. We assume households consist of an adult parent and her children. Childhood lasts for T periods, during which parents must choose the level of investment to be made in their children. After T periods of childhood, they become adults with skills $h = g(\theta_T)$, that they sell in the labor market, and they also give birth to children of their own. At conception, children endow a vector of k skills $\theta_0 = (\theta_{0,1}, \theta_{0,2}, \ldots, \theta_{0,k})$, which develops in the womb to form a vector of skills at birth, $\theta_1 = (\theta_{1,1}, \theta_{1,2}, \ldots, \theta_{1,k})$. This development process already depends on parental investments (such as prenatal care or smoking during pregnancy) and random shocks that may affect children *in uterus* (such as placenta positioning or some maternal physiological condition)[11]. In the following stages, parents continue to invest in their children to form skills, always building up on previous skills. This process can be formalized as follows:

$$\theta_{k,t+1} = f_t(\theta_t, I_t, h, X_t, \eta_{k,t}) \tag{1}$$

That is, skill k in period $t+1$ depends on the vector of skills in period t, θ_t, on parental investments at time t, I_t, parental skills, h, and other family and child characteristics X_t, together with random shocks $\eta_{k,t}$ that are independent across skills and periods. Parental investments account for both monetary and time investments. Parental skills include human capital that may increase the productivity of investments (e.g. parents with more schooling may be more able to help their children with homework). With respect to other children and family characteristics, we are particularly interested in a child's race, although other characteristics such as gender and presence of both parents may also be relevant. The relationship between these inputs and skills in the next periods is mediated by a technology of skill formation, f_t, which is time-variant. Cunha and Heckman (2007) discuss conditions on this production function so that empirical findings such as dynamic complementarity, self-productivity, and critical and sensitive periods may arise. Particularly, a CES production function is able to capture all of these characteristics under certain values for its parameters.

We can write the process of formation for skills at period $t+1$ recursively as:

$$\theta_{k,t+1} = m_t(h, \theta_0, I_0, \ldots, I_t, X_0, \ldots, X_t, \eta_{k,0}^t), \text{ for } t = 0, 1, \ldots, T. \tag{2}$$

that is, skills k at period $t+1$ depends on the initial vector of skills, parental skills, and the whole history of investments $(I_0, \ldots, I_t)$, family and child characteristics $(X_0, \ldots, X_t)$, and independent shocks across time $(\eta_{k,0}^t = (\eta_{k,0}, \ldots, \eta_{k,t}))$ that affect skill accumulation. At period T, this set of skills is converted into a stock of human capital $h = g(\theta_T)$.

Cunha and Heckman (2007) illustrate how investments by optimizing parents can vary under different circumstances. For this, they develop a simple over-lapping generations model with four periods. Upon reaching adulthood, parents receive a bequest b and adult human capital h, which they offer inelastically in the labor market for wage w. They also give birth to a children with initial skill levels θ_1. Parents must then decide how much to consume in periods 1 and 2 (c_1 and c_2), how much to invest in their children (I_1 and I_2) and how much to leave as bequest to them (b'). Assuming

[11] Cunha and Heckman (2007) are open to the possibility of including *in uterus* investments in their model, although they do not do so explicitly.

interest rates r and scalar skills for both parents and children, the budget constraint for parents is written as:

$$c_1 + I_1 + \frac{c_2 + I_2}{(1+r)} + \frac{b'}{(1+r)^2} = wh + \frac{wh}{(1+r)} + b \tag{3}$$

Parents are altruistic towards their children and want to maximize their welfare. Assuming a discount factor β, parental altruism δ, and utility function $u(.)$, we can formulate parents' recursive problem as:

$$V(h, b, \theta_1) = \max\{u(c_1) + \beta u(c_2) + \beta^2 \delta E[V(h', b', \theta_1')]\} \tag{4}$$

where h', b' and θ_1' represent their children's adult skills, bequest, and grandchild's initial vector of skills.

By maximizing equation (4) subject to budget (3) and technology (2), Cunha and Heckman (2007) are able to derive optimal ratios of early to late investment. They point out three market imperfections that may affect this ratio of investments and therefore children's accumulation of skills. The first is that children cannot choose the optimal level of parental skills, h, that match their initial endowment of skills, θ_1. Children of parents with low levels of education, for instance, would acquire less skills than children with similar levels of initial skills, but with highly educated parents. This market imperfection is present even under unconstrained borrowing. Second, parents cannot leave negative bequests, that is, they cannot borrow against their children's future income, which may lead to suboptimal investment in poorer families. Third, under incomplete markets, parents may be unable to borrow against their own future incomes, what could inhibit adequate investments in the first periods, thus hindering later development due to dynamic complementarity. These market failures help explain why children born in poor families under-invest in education. As we have seem in our descriptive analysis, black children are born in less advantaged environments, what can explain their worse results in schooling and other skills compared to their white counterparts.

To empirically test to what extent black-white gaps in birth weight, socioecomotional skills, or years of schooling are due to differences in family background we estimate simple versions of equation (1). We follow Todd and Wolpin (2003, 2007) and Cunha and Heckman (2008) in proposing a linear specification for this equation. We define six periods in our setting, period 0, relative to pregnancy, period 1, from perinatal to 10 years, period 2, between 11 and 14 years, period 3, between 15 and 17 years, period 4, between 18 and 21 years, and period 5, from 22 years on. For period one, we estimate the following equation:

$$\theta_{1,i} = \alpha_0 + \delta_0 D_i + \beta_0 I_{0,i} + \gamma_0 X_{0,i} + \varepsilon_{0,i} \tag{5}$$

where θ_1 represent some health outcome at birth, particularly birth weight, α_0 is a constant, D_i is an indicator for race, such that $D_i = \mathbb{1}[\text{Race} = \text{Non-white}]$ and δ_0 is the estimated black-white gap in this skill. *In uterus* parental investments, $I_{0,i}$, is a vector that includes the number of prenatal appointments, an indicator variable for maternal smoking during pregnancy, an indicator variable for mother working during pregnancy, and gestational weight gain. Finally, we add many family and individual controls: indicator for baby being male, maternal BMI before pregnancy, maternal height, whether a child born before that pregnancy was low birth weight or premature, the number of previous births (parity), dummies for whether the mother was diagnosed with diabetes, high pressure, urinary infection, anemia, or had an abortion threat. Vector $X_{0,i}$ also includes SES variables: log of *per capita* family income, mother's schooling, maternal age, and a dummy for living with partner. We assume the error term $\varepsilon_{1,i}$ is independent and identically distributed. Theoretically, this error term includes both an initial endowment of skills, $\theta_{0,i}$, as well as a random shock $\eta_{1,i}$. However, we cannot distinguish these objects in our empirical setting.

The inclusion of these controls are motivated by the literature, particularly Kramer (1987). He reviews the literature on determinants of birth weight and argue that there are well-established causal impacts on intrauterine growth that include infant sex and race, maternal height and weight, gestational weight gain, cigarette smoking, parity, previous cases of low birth weight, episodic illness and other conditions[12]. Socioeconomic status seem to only play a role as indirect factor, having no causal impact on birth weight alone. However, we include it due to its correlation with our racial dummy. Similar factors are also associated with gestational duration.

After investigating birth outcomes we turn to SDQ scores at 11 years and, most importantly, years of schooling across age. For age 11 we estimate the following functions:

$$\theta^k_{2,i} = \alpha^k_{1,i} + \delta^k_{1,i} D_i + \beta^k_1 I_{0,i} + \gamma^k_{1,1} X_{1,i} + \gamma^k_{1,0} X_{0,i} + \theta_{1,i} + \varepsilon^k_{1,i} \quad (6)$$

where k correspond to SDQ scores or years of schooling. Apart from the variables already included in the first regression (D, I_0, X_0) we also include: i) a measure of health at birth $\theta_{1,i}$, generally birth weight; ii) family characteristics in period 1, $X_{1,i}$, which include: log of *per capita* family income, number of children born up to that period[13], a dummy variable for mothers living with a partner, and mother's SRQ-20. Although we do not include variables for investments in the first period, the covariates included act as inputs in the demand equations for investments and help dealing with potential omitted-variable bias, what Todd and Wolpin (2007) call hybrid specifications of the production function. The inclusion of *in uterus* investments and birth weight, together with other covariates related to early health, is motivated by recent developments that support the *fetal origins hypothesis* (Currie and Moretti, 2007; Case and Paxson, 2010; Currie, 2011; Currie and Almond, 2011; Almond and Currie, 2011). This literature has shown that early life condition, measured through indicator of health at birth such as birth weight, are related to outcomes later in life, including schooling and performance in test scores.

For ages 15, 18, and 22 (periods 3, 4, and 5) we also include previous schooling and SDQ score at 11 years as independent variables, and for ages 18 and 22 we include log of *per capita* family income at 18 years old. The introduction of these lagged dependent variables aims at reducing problems of omitted-variable bias due to lack of data on lagged inputs (Todd and Wolpin, 2003).

3 Main Results

This section presents the main results regarding racial gaps in the outcomes of interest based on the empirical analysis discussed in the previous section. We first present and discuss the finds related to birth outcomes, assessing to what degree black and white children differ at the moment they are born and whether such differences can be explained by differences in family characteristics and investments during pregnancy. Next, we analyze differences in years of schooling and socioemotional skills at age 11. We also analyze the evolution of the racial gap in years of schooling from childhood up to early adulthood.

3.1 Differences at birth

Table 3 displays regressions of birth weight on several variables. Birth weight is normalized to have mean zero and standard deviation (SD) one across the whole sample (one standard deviation is approximately 550 grams). We use all available information and restrict estimation to the sample of children that are singleton and not premature. For this variable there is significant raw gap between white and non-white children to be around 0.15 SD. We next introduce sequentially some sets of variables on the regression. Column (1) includes socioeconomic status of families, represented by log

[12]Rahmati et al. (2017) and Figueiredo et al. (2018) find evidence that maternal anemia is related to infant low birth weight; gestational diabetes is associated with higher levels of insulin in the blood and thus with higher birth weight (Silverman et al., 1991); high pressure (in the form of preeclampsia) may be related to lower birth weight (Ødegård et al., 2000); and urinary infection may be causally related with birth weight in developing countries, although evidence is not sufficient (Kramer, 1987).

[13]Thus including children born before the interviewed child was born.

family income *per capita*, mother's schooling and age and whether the mother lived with a partner (not necessarily the child's dad). We observe a considerable reduction by 1/3 in the racial gap. Next, in column (2) we introduce variables related to *in uterus* investments (number of prenatal visits, whether the mother smoked during pregnancy, whether she worked during pregnancy, and the gestational weight gain) and other controls that the medical literature have related as determinants of birth weight, as previously discussed. We notice that with the addition of such variables, the covariates representing families' socioeconomic status cease to be significant. This is line with the medical literature that find no effect of socioeconomic variables once the relevant obstetric factors are controlled for (Kramer, 1987). In particular, the results also show that the number of prenatal appointments and smoking during pregnancy are particularly relevant in determining birth weight. An increase of one SD in the number of prenatal appointments (half visit) is related with an increase of 0.05 SD in birth weight, whereas smoking during pregnancy considerably reduces a baby's birth weight in nearly a quarter of SD. Nonetheless, we still find significant racial differences in birth weight. The size of the gap is reduced by more than 50%, to 0.07 SD, around 40 grams. As a matter of comparison, Morisaki et al. (2017) find a black-white gap in birth weight of nearly 150 grams in the United States using similar controls.

Table 3: Regressions for health at birth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Birth Weight		Length at Birth		Low Birth Weight		Premature	
Conditional Racial Gap	-0.10***	-0.07**	-0.08**	-0.08**	-0.00	-0.00	0.01	-0.00
	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Log fam. inc. 1993 pc	0.01	-0.02	0.03	-0.01	0.00	0.00	0.00	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.00)	(0.01)	(0.01)	(0.01)
Mother's schooling (Std)	0.07***	0.02	0.02	-0.01	-0.01***	-0.01**	-0.01	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.00)	(0.00)	(0.01)	(0.01)
Number prenatal (Std)		0.05***		0.03*		-0.01**		-0.02***
		(0.02)		(0.02)		(0.00)		(0.01)
Smoked in Preg.		-0.26***		-0.28***		0.02***		0.00
		(0.03)		(0.03)		(0.01)		(0.01)
Other Socioeconomic variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other investments in pregnancy	No	Yes	No	Yes	No	Yes	No	Yes
Obstetric Factors	No	Yes	No	Yes	No	Yes	No	Yes
Hospital FE	No	Yes	No	Yes	No	Yes	No	Yes
Month of Birth	No	Yes	No	Yes	No	Yes	No	Yes
Raw Racial Gap	-0.15***	-0.15***	-0.12***	-0.12***	0.01	0.01	0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
R-squared	0.023	0.235	0.010	0.186	0.003	0.040	-0.000	0.039
N	3464	3464	3455	3455	3464	3464	3757	3757

Birth weight and length at birth are normalized to have mean zero and standard deviation one (mean ≈ 3150 grams and SD ≈ 550 grams for birth weight and mean ≈ 48.7cm and SD ≈ 2.4 for length)). Low Birth Weight (< 2500 grams) and Premature (< 37 weeks) are dummies. For Premature, we include only singletons, and for the remaining dependent variables we also exclude premature babies. Variables with "(Std)" are also normalized to have mean zero and standard deviation one. Other Socioeconomic variables include mother's age and whether partner lives in the household. Other investments in pregnancy include: Whether mother worked during pregnancy and gestational weight gain. Obstetric factor include: mother's BMI and height prior pregnancy, previous cases of low birth weight and prematurity, infant's gender, parity and dummies for the occurrence of diabetes, high pressure, urinary infections, anemia and risk of abortion during pregnancy. Robust standard errors in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In columns (3)-(8) of Table 3 we conduct a similar analysis for length at birth (also normalized, with mean ≈ 48.7cm and SD ≈ 2.4), chances of being low birth weight (< 2500 grams), and chances of being premature[14] (< 37 weeks). The results for birth weight are again found for length at birth, but we do not find any racial differences in the chances of being low birth weight of being premature. Some of these finding are reported in the literature while other diverge. For instance, Kramer (1987) finds no evidence of racial differences in the chances of preterm births. However, there is evidence of higher prevalence of low birth weights in African-Americans in the United States compared to White Americans (David and Collins Jr, 1997). Among the hypothesis raised to explain these gaps in outcomes are life-course perspectives that highlight the importance of factors such as maternal birth weight, exposure to stress, and different environments. Since black mothers were born with lower birth weights, are more exposed to stress (due to racism, for instance), and live in poorer neighborhood, this could end up affecting the health of their babies (Lu and Halfon, 2003; Collins Jr and David, 2009). However, the reason why this is not reflect into prematurity or LBW is not clear.

3.2 Racial differences in schooling and socioemotional skills

In this section we discuss differences in schooling and socioemotional skills at age 11. Table 4 displays the results of regressing years of schooling at 11 years old on a series of covariates. Column 1 display the raw gap of nearly 0.4 SD, which corresponds to almost half an year of schooling, a quite substantial difference given a mean of less than 4 years. Next, we regress education at 11 years on standardized birth weight. Higher birth weights are correlated with higher levels of schooling, but it only slightly reduces racial differences in education. In column 3 we add differences in mother SRQ-20, a measure of psychological health that has not yet been studied as a possible source of racial inequalities in human capital. An increase of one standard deviation in the score in this measure (nearly 4 points in a scale that ranges from 0 to 20) is related with 0.17 SD less years of schooling, and a reduction in the racial gap of around 10%. Column 4 display the results when socioeconomic status at 11 years are introduced, which are highly correlated with race and psychological health. As expected, there is a considerable decreases in the coefficient of the non-white indicator, together with a significant drop in the mother's SRQ-20 measure. Nonetheless, both coefficients remain significant.

Finally, we introduce variables related to mother behavior during pregnancy, SES at birth, and obstetric factors. These variables further reduce the racial gap indicator, and magnitude of the coefficients on birth weight and mother's SRQ-20. Also, the number of prenatal appointments and whether or not the mother smoked during pregnancy remain significantly correlated with education at 11 years, meaning that *in uterus* investment may have long-lasting effects on schooling. Due to lack of space, we do not display the estimated coefficients for the remaining variables here, but the results confirm the relation between years of schooling and other socioeconomic variables, such as mother's education, number of children, and *per capita* family income. In particular, families' income both at the moment of birth and at 11 years are significantly correlated with educational achievement. The effects are large, with an increase of 1% in the *per capita* family income at birth related with an increase of 0.08 SD in years of schooling, and 0.15 SD when we consider family income at 11 years.

Table 5 reproduces the same estimations for the standardized total score in the SDQ, with its 4 main components aggregated (that is, excluding pro-social behavior), measured using children's answer. We interpret this measure as a proxy for socioemotional skills. Column 1 displays the raw gap in the score, of 1/4 of a standard deviation. At column 2 we observe that birth weight is barely correlated with the dependent variable. Moreover, its inclusion almost no affect on the racial gap. Including mother's SRQ-20 alters slightly the coefficient of the racial dummy, and it has a considerable correlation with SDQ score. Column 4 further adds SES at 11 years, which considerably reduces racial inequalities in the dependent variable, even though it is not able to completely close it. In column 5 covariates measured at birth were included, further reducing the gap. The coefficient of smoking during pregnancy is significant and sizable, whereas the number of

[14] In this case, we include in the regression singletons, but do not exclude premature babies.

Table 4: Racial gaps in schooling at 11 years old

	(1)	(2)	(3)	(4)	(5)
Dependent Variable			Schooling at 11 years old		
Non-white	-0.409***	-0.393***	-0.362***	-0.148***	-0.111***
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
Std. Birth Weight		0.097***	0.095***	0.078***	0.066***
		(0.02)	(0.02)	(0.02)	(0.02)
Mother's SRQ (Std)			-0.170***	-0.068***	-0.050***
			(0.02)	(0.02)	(0.01)
Number prenatal (Std)					0.074***
					(0.02)
Smoked in Preg.					-0.050*
					(0.03)
SES at 11 years	No	No	No	Yes	Yes
SES at birth	No	No	No	No	Yes
Other investments in pregnancy	No	No	No	No	Yes
Obstetric Factors	No	No	No	No	Yes
R-squared	0.15	0.16	0.20	0.34	0.39
N	3382	3382	3382	3382	3382

All regression include intercept and fixed effects for month of birth and interview. Years of schooling are normalized to have mean zero and standard deviation one (mean ≈ 3.5 years and SD ≈ 1.15 years). Covariates with "(Std)" are also normalized. Other investments in pregnancy include: Whether mother worked during pregnancy and gestational weight gain. Obstetric factor include: mother's BMI and height prior pregnancy, previous cases of low birth weight and prematurity, infant's gender, parity and dummies for the occurrence of diabetes, high pressure, urinary infections, anemia and risk of abortion during pregnancy. SES at birth include: log of per capita family income, mother's education, age, and indicator if she was living with husband. SES at 11 include log of per capita family income, number of children, and if mother was living husband. Robust standard errors in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

prenatal visit is not. Although we do not show the estimated coefficients for all variables, we find that *per capita* family income is not correlated with SDQ scores, even though other socioeconomic indicators such as number of children, presence of a partner at the household, and mother's age and education are all significantly correlated with coefficients in the expected direction.

3.3 The evolution of the racial in schooling from 15 to 22 years old

Based on the theoretical framework previously described, we now address the dynamics of racial gaps in years schooling at ages 15, 18, and 22 years old. To estimate the conditional gap on these outcomes, we condition the regression on previous educational achievement and SDQ total scores at age 11. Table 6 shows the results. At age 15, black children are lagging behind nearly 0.9 year of schooling, or half SD (one SD ≈ 1.75 years of schooling). Nonetheless, when we control for previous schooling and SDQ score the racial gap is reduced considerably to 0.18 SD. Further controlling for family socioeconomic status, mother SRQ-20 by the time children were 11 years old, and controls from a child's birth reduces marginally the size of the racial gap. Birth weight and maternal SRQ-20 are not significant once we control for all other variables, meaning that their effects are possibly restricted to the first years of schooling. However, number of prenatal appointments and smoking during pregnancy seem to have lasting effects.

The results for years of schooling at ages 18 and 22 are distinct. First, controlling for previous schooling and SDQ total score reduces the racial gap significantly, to nearly 1/10 of its original size in the case of education at 18 years. When we include other variables related to SES at various ages and factors that affect birth weights the racial gap ceases to be significant at both ages. Hence,

Table 5: Racial gaps in SDQ total score at 11 years old

	(1)	(2)	(3)	(4)	(5)
Dependent Variable			SDQ total score at 11 years old		
Non-white	-0.262***	-0.256***	-0.224***	-0.139***	-0.118***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Std. Birth Weight		0.035*	0.032*	0.025	0.018
		(0.02)	(0.02)	(0.02)	(0.02)
Mother's SRQ (Std)			-0.177***	-0.137***	-0.124***
			(0.02)	(0.02)	(0.02)
Number prenatal (Std)					0.010
					(0.02)
Smoked in Preg.					-0.140***
					(0.04)
SES at 11 years	No	No	No	Yes	Yes
SES at birth	No	No	No	No	Yes
Other investments in pregnancy	No	No	No	No	Yes
Obstetric Factors	No	No	No	No	Yes
R-squared	0.01	0.02	0.05	0.07	0.08
N	3293	3293	3293	3293	3293

All regression include intercept and fixed effects for month of birth and interview. Other investments in pregnancy include: Whether mother worked during pregnancy and gestational weight gain. Obstetric factor include: mother's BMI and height prior pregnancy, previous cases of low birth weight and prematurity, infant's gender, parity and dummies for the occurrence of diabetes, high pressure, urinary infections, anemia and risk of abortion during pregnancy. SES at birth include: log of per capita family income, mother's education, age, and indicator if she was living with husband. SES at 11 include log of per capita family income, number of children, and if mother was living husbund. Robust standard errors in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

we conclude that although the normalized unconditional gap is somewhat stable across childhood and early adulthood, the conditional gap is determined up to adolescence. This is in line with finds by Cameron and Heckman (2001) showing that the racial gap in college attendance is determined by early home environment and parental background rather than by financial constraints when children choose to go or not to college. Finally, *in uterus* investments such as number of prenatal appointments and smoking in pregnancy, as well as birth weight, are not significantly related with schooling at age 18, although it is for schooling at age 22. This may be a sign of long lasting effects of these factors, although the reason why the only manifest at 22 years but not 18 is not clear. Another possibility is that early home environments and cognitive skills are more important to educational achievements past age 18 than for schooling between 15 and 18 years old, what is captured by the coefficients on birth weights and *in uterus* investments, since these variables are correlated with both inputs, corroborating again the finds of Cameron and Heckman (2001).

4 Heterogeneity Checks

4.1 Differences in dynamics by family socioeconomic status

Here, we investigate whether the racial gap in years of schooling has different dynamics within different socioeconomic groups. We consider two cutoffs for SES: one based on real family income at the time children were born and another based on maternal schooling by the time of delivery. Families are split in two halves of the distribution of family *per capita* income at the median, whereas mothers are split in three educational groups: up to 7th grade (incomplete primary school), between 8th and 11th grade (complete primary but incomplete secondary school), and those with at least

Table 6: Racial gaps in schooling at 15, 18, and 22 years old

Dependent Variable	(1)	(2) Schooling at 15	(3)	(4)	(5) Schooling at 18	(6)	(7)	(8) Schooling at 22	(9)
Racial gap	-0.511***	-0.184***	-0.123***	-0.477***	-0.049**	-0.005	-0.435***	-0.071***	-0.017
	(0.037)	(0.026)	(0.026)	(0.039)	(0.025)	(0.025)	(0.039)	(0.025)	(0.025)
Std. SDQ Total score		0.113***	0.083***		0.065***	0.049***		0.032**	0.022*
		(0.012)	(0.012)		(0.013)	(0.012)		(0.013)	(0.012)
Mother's SRQ (Std)			-0.019			-0.014			0.006
			(0.012)			(0.011)			(0.012)
Std. Birth Weight			-0.005			0.017			0.046***
			(0.012)			(0.011)			(0.014)
Number prenatal (Std)			0.025**			0.012			0.032**
			(0.012)			(0.012)			(0.014)
Smoked in Preg.			-0.077***			-0.036			-0.049**
			(0.024)			(0.022)			(0.024)
Previous schooling	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Obstetric Factor	No	No	Yes	No	No	Yes	No	No	Yes
Other *in uterus* investments	No	No	Yes	No	No	Yes	No	No	Yes
SES at birth	No	No	Yes	No	No	Yes	No	No	Yes
SES at 11 years	No	No	Yes	No	No	Yes	No	No	Yes
SES at 18 years	No	No	No	No	No	Yes	No	No	Yes
R-squared	0.08	0.56	0.60	0.06	0.69	0.71	0.04	0.63	0.65
N	3159	3159	3159	2786	2786	2786	2487	2487	2487

All regression include intercept and fixed effects for month of birth. Years of schooling are normalized to have mean zero and standard deviation one within each year (at 15: mean ≈ 6.4 years and SD ≈ 1.75 years; at 18: mean ≈ 8.6 years and SD ≈ 2.25 years; at 22: mean ≈ 10 years and SD ≈ 2.92 years). Covariates with "(Std)" are also normalized.Obstetric factor include: mother's BMI and height prior pregnancy, previous cases of low birth weight and prematurity, infant's gender, parity and dummies for the occurrence of diabetes, high pressure, urinary infections, anemia and risk of abortion during pregnancy. *In uterus* investments include: Mother worked during pregnancy, number of prenatal appointments, mother smoked during pregnancy, and gestational weight gain. SES at birth include: log of per capita family income, mother's education, age, and indicator if she was living with husband. SES at 11 include log of per capita family income, number of children, and if mother was living with husband. SES at 18 years include log of per capita family income. Robust standard errors in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

complete high school. Table 7 summarizes some information concerning these socioeconomic groups regarding average family income *per capita*, mother's schooling, the sample size of white and non-white children, and racial gaps in family income *per capita* and maternal schooling. As we move up in the socioeconomic latter, the number of non-white children decreases, an expect fact given Brazil's racial inequality. Moreover, we notice that the differences between the average white and non-white family is larger among privileged subgroups of the population. For instance, whereas the racial gap in family income *per capita* is only 15 Brazilian Reals within the bottom half of the distribution and differences in maternal education are only 0.7 years of schooling, in the top half these differences are respectively almost 300 Brazilian Reals and 1.8 years of schooling.

Table 7: Descriptive statistics by socioeconomic groups

	Avg. family income pc	Avg. mother schooling	Number of Whites	Number of Non-whites	Racial Gap in family income pc	Racial Gap in maternal schooling
Bottom half	176,46	5,66	800	444	15,43	0,68
Top half	878,60	8,58	964	279	296,96	1,80
Mother up to 7th	287,41	4,43	1682	941	63,60	0,31
Mother 8th-11th	494,66	8,58	665	209	97,98	0,10
Mother w/ HS+	1186,81	12,53	601	107	545,23	0,79

Table 8 display racial gaps in years of schooling together with standard errors for different models and subgroups. In panel (a), we consider the sample split according to the family income criterion. For ages 11 through 18 the behavior is quite similar for both groups, with significant conditional gap at 11 and 15 years and no conditional gaps when children are 18 years old. The point estimates for all gaps are larger within the richer group, although they are always quite similar. We observe that the conditional racial gap in only significantly different from zero at age 22 for the richer group, around 0.09 SD, whereas the point estimate for the poorer group is even positive, although not significant. This suggests that black children from more advantaged background may face additional difficulties in getting into college.

Table 8: Racial Gaps in schooling by family income per capita and mother schooling

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Schooling at 11		Schooling at 15		Schooling at 18		Schooling at 22	
Panel (a): Family income per capita								
Bottom half								
Racial gap	-0.273***	-0.084*	-0.370***	-0.134***	-0.312***	0.016	-0.232***	0.022
	(0.052)	(0.048)	(0.054)	(0.036)	(0.054)	(0.033)	(0.052)	(0.033)
R-squared	0.09	0.33	0.06	0.62	0.03	0.71	0.01	0.65
N	1244	1244	1229	1229	1244	1229	1244	1244
Top Half								
Racial gap	-0.303***	-0.154***	-0.432***	-0.146***	-0.412***	-0.021	-0.466***	-0.087**
	(0.053)	(0.048)	(0.060)	(0.045)	(0.057)	(0.040)	(0.057)	(0.040)
R-squared	0.27	0.39	0.08	0.48	0.06	0.65	0.06	0.56
N	1243	1243	1218	1218	1243	1218	1243	1243
Panel (b): Mother's schooling								
Up to 7th Grade								
Racial gap	-0.329***	-0.136***	-0.401***	-0.142***	-0.355***	0.007	-0.314***	-0.020
	(0.049)	(0.046)	(0.051)	(0.036)	(0.051)	(0.032)	(0.047)	(0.031)
R-squared	0.10	0.30	0.06	0.57	0.03	0.69	0.02	0.65
N	1418	1418	1397	1397	1418	1397	1418	1418
Between 8th and 11th Grade								
Racial gap	-0.156**	-0.108*	-0.240***	-0.099*	-0.148*	0.071	-0.079	0.028
	(0.067)	(0.063)	(0.075)	(0.056)	(0.078)	(0.055)	(0.079)	(0.057)
R-squared	0.23	0.33	0.03	0.46	0.00	0.60	-0.01	0.48
N	587	587	580	580	587	580	587	587
Complete High School or Higher								
Racial gap	0.002	0.113*	-0.280***	-0.182***	-0.375***	-0.131**	-0.477***	-0.145*
	(0.072)	(0.067)	(0.087)	(0.062)	(0.095)	(0.060)	(0.091)	(0.076)
R-squared	0.41	0.48	0.06	0.41	0.05	0.62	0.05	0.40
N	482	482	470	470	482	470	482	482

All dependent variables are standardized to have mean zero and variance one. Odd columns refer to unconditional gaps, whereas even columns refer to conditional ones. Unconditional gaps include as dependent variable only dummies for month of birth and month of interview (the latter only for schooling at 11 years old). Conditional gaps include all controls specified for models described in Tables 4 and 6.

In panel (b), we evaluate the heterogeneity in racial gaps based on mother's schooling and reach similar conclusions. In the first period, when children are 11 years old, the unconditional racial gap in larger among lower socioeconomic groups, and there are no significant differences among children of mother with at least complete high school. The conditional gap favors white children among the two lower strata, but favors black children within the highest. Hence, at least in the first period, the behavior of the racial gap across subgroups seems quite different from that of cutoff based on income, with more smaller gaps within more privileged groups. When children are 15 years old, the results for all three groups are quite similar, with both unconditional and conditional gaps favoring white children. Finally, the conditional racial gap at age 18 and 22 is only significant among children born in the highest socioeconomic group. These results again suggest that among the highest social groups, black children may struggle to continue their studies into higher levels of schooling, even though they seemed to benefit from a better background in the first period.

Why is the conditional racial gap in schooling only significant for the groups of higher socioeconomic status? A possible explanation is that children born in lower SES families simply cannot afford continuing their studies after 18 years old or after they complete high school. Hence, if only the richest children can keep studying after a certain age, and if mechanisms that prevent black students to perform as well as their white counterparts still exist during early adulthood, then it is expected that racial inequalities in schooling grow only among those children.

4.2 Gender

There is some evidence that differences in outcomes between black and white children can be distinct between genders. For instance, Chetty et al. (2018) find that the racial gap in intergenerational mobility in the United States is driven mainly by differences between black and white males, with no racial differences among females. In Brazil, Rangel (2015) and Marteleto and Dondero (2016) find that different rates of school enrollment between siblings of distinct skin complexion is stronger for boys. Therefore, we investigate the possibility that the black-white gap in schooling has distinct dynamics between boys and girls.

Table 9 displays the estimated black-white gap in schooling across years for males and females. The normalized unconditional racial gap behaves differently between boys and girls. For boys, it grows considerably fast between 11 and 15 years old, jumping from 0.47 SD to 0.64 SD. After that, the gap decreases at 18 and 22 years old, when it equals the level when children were 11 years old. For girls, however, is starts at a much lower level of 0.31 SD, then increasing to 0.42 at age 15. At 18 years old, the normalized unconditional gap reaches 0.45 and settles at this value. Although the trend is much different than that observed among boys, both groups end up with similar levels of racial inequality. When we test the equality of the racial gap among these two groups, we confirm that they are statistically distinct at 11 and 15 years old, but not after 18 years.

Table 9: Racial gaps in schooling by gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	Schooling at 11		Schooling at 15		Schooling at 18		Schooling at 22	
Panel (a) - Males								
Racial gap	-0.472***	-0.185***	-0.640***	-0.210***	-0.537***	0.028	-0.476***	0.007
	(0.057)	(0.052)	(0.061)	(0.043)	(0.062)	(0.039)	(0.062)	(0.038)
R-squared	0.15	0.38	0.10	0.61	0.06	0.70	0.04	0.68
N	1156	1156	1127	1127	1156	1127	1156	1156
Panel (b) - Females								
Racial gap	-0.310***	-0.051	-0.421***	-0.089***	-0.453***	-0.026	-0.455***	-0.040
	(0.041)	(0.039)	(0.047)	(0.034)	(0.049)	(0.030)	(0.050)	(0.036)
R-squared	0.20	0.41	0.09	0.59	0.07	0.72	0.06	0.60
N	1331	1331	1320	1320	1331	1320	1331	1331
F-statistic	4.377	3.747	7.319	4.725	1.062	1.133	.066	.893
P-value	.036	.053	.007	.03	.303	.287	.797	.345

Odd column display gaps conditional only on month of birth and month of interview (the latter only for schooling at 11 years old). Even columns present conditional gaps based on the full models estimated for each age, as in the last column of tables 4 and 6. Robust standard errors in parenthesis. F-statistic refers to the test of equality between the race dummies estimated for each gender separately.
*p<.10, **p<.05, ***p<.01

Regarding the conditional gap, we again observe differences between gender. First, the conditional racial gap at 11 years is only significantly different from zero among boys, while at 15 years it is different for both groups, but smaller among girls. At 18 and 22 years, there are no conditional racial gaps for neither of these groups.

Therefore, it is not possible to characterize the black-white gap in schooling as being a phenomenon exclusive to one gender, since within both groups black individuals lag behind their peers. Nonetheless, the dynamics seem to be quite different for black boys and girls. For the first group, the first years seem to be the biggest challenge, as the gap increases substantially up to 15 years old. For girls, the gap between black and white students is much smaller, and contrary to boys, we observe a considerable increase in the unconditional black-white gap from 18 to 22 years old. Future research should investigate deeply the reasons behind these distinctive patterns across gender.

5 Summary and Policy Discussion

The degree to which racial inequalities in labor markets or human capital are due to "objective" factors or the results of racial discrimination has been an open topic in economics and other social sciences for decades. In Brazil, the idea of "racial democracy" has been challenged in academia and public discourse in the last decades, and some evidence has pointed out to the existence of racial inequalities in labor market and human capital accumulation that cannot be fully explained by observable variables. Here, we shed some light on the dynamics of human capital accumulation from birth up to early childhood using a unique dataset that allow us to observe family characteristics during pregnancy and childhood, as well as birth outcomes and schooling achievement from 11 to 22 years old. We find evidence that black children are lighter and smaller when they are born, but this does not reflects in prematurity or chances of being low birth weight. We also report unexplained differences between black and white children at age 11 both in years of schooling and SDQ total scores. Although there exists a significant relationship between birth weight and years of schooling, its size is too small when compared to socioeconomic status and can explain only a small difference in racial differences in schooling. However, *in uterus* investments and maternal mental health are other mechanisms that affect racial gap and had received little attention in the past. Future research should investigate this channels further.

We document increasing gaps in years of schooling between black and white children that are already quite substantial at age 11, at around half an year, and grow to surpass one years of education in early adulthood. However, the normalized unconditional racial gap in somewhat stable across years, being of 0.4 SD at age 11, reaching a peak of 0.5 SD at 15, and then decreasing to 0.43 SD at 22 years old. When we include controls for previous education and other covariates in the regressions, the conditional racial gaps in years of schooling cease to be significant after 15 years old. This finding supports the idea that racial inequalities arise early in life and that public policy should focus its attention in providing adequate support for disadvantaged children even before they start school.

Finally, there seems to be some differences in the dynamics of the racial gap according to socioeconomic status and gender. Conditional racial gaps in years of schooling are significant at ages 18 and 22 for children of mothers with completed high school and in the top half of the income distribution, contrary to what happen in the whole sample or among families of lower SES. These results may be interpreted as a cautionary warning with respect to the stability of the racial gap after age 15, as is the case for the whole sample, since black children with better background - precisely the ones more likely to carry on their studies past high school - may still face difficulties upon entering college. Hence, policies such as affirmative action in universities may be an important instrument in fighting racial inequalities in education. Regarding differences between between boys and girls, we find that the racial gap grows much faster at a young age for boys, while it grows at a steady pace for girls. Conditional racial gaps among girls are only significant at age 15, whereas for boys it is significant both at 11 and 15 years. The reasons behind this distinct pattern across gender in racial gaps should be further investigated in future research.

References

D. Almond and J. Currie. Killing me softly: The fetal origins hypothesis. *Journal of economic perspectives*, 25(3):153–72, 2011.

R. d. A. e. Arraes, F. L. S. Menezes, and A. G. Simonassi. Earning differentials by occupational categories: Gender, race and regions. *EconomiA*, 15(3):363–386, 2014.

R. P. d. Barros, R. Mendonça, D. D. d. Santos, and G. Quintaes. Determinantes do desempenho educacional no brasil. 2001.

F. Botelho, R. A. Madeira, and M. A. Rangel. Racial discrimination in grading: Evidence from brazil. *American Economic Journal: Applied Economics*, 7(4):37–52, 2015.

S. V. Cameron and J. J. Heckman. The dynamics of educational attainment for black, hispanic, and white males. *Journal of political Economy*, 109(3):455–499, 2001.

P. Carneiro, J. J. Heckman, and D. V. Masterov. Labor market discrimination and racial differences in premarket factors. *The Journal of Law and Economics*, 48(1):1–39, 2005.

A. Case and C. Paxson. Causes and consequences of early-life health. *Demography*, 47(1):S65–S85, 2010.

K. Y. Chay, J. Guryan, and B. Mazumder. Birth cohort and the black-white achievement gap: The roles of access and health soon after birth, 2009.

K. Y. Chay, J. Guryan, and B. Mazumder. Early life environment and racial inequality in education and earnings in the united states, 2014.

R. Chetty, N. Hendren, M. R. Jones, and S. R. Porter. Race and economic opportunity in the united states: An intergenerational perspective, 2018.

J. W. Collins Jr and R. J. David. Racial disparity in low birth weight and infant mortality. *Clinics in perinatology*, 36(1):63–73, 2009.

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

F. Cunha and J. J. Heckman. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources*, 43(4):738–782, 2008.

F. Cunha, J. J. Heckman, L. Lochner, and D. V. Masterov. Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:697–812, 2006.

A. Z. Curi and N. Menezes-Filho. The relationship between school performance and future wages in brazil. *Economia*, 15(3):261–274, 2014.

J. Currie. Inequality at birth: Some causes and consequences. *American Economic Review*, 101(3): 1–22, 2011.

J. Currie and D. Almond. Human capital development before age five. In *Handbook of labor economics*, volume 4, pages 1315–1486. Elsevier, Amsterdam, 2011.

J. Currie and E. Moretti. Biology as destiny? short-and long-run determinants of intergenerational transmission of birth weight. *Journal of Labor economics*, 25(2):231–264, 2007.

R. J. David and J. W. Collins Jr. Differing birth weight among infants of us-born blacks, african-born blacks, and us-born whites. *New England Journal of Medicine*, 337(17):1209–1214, 1997.

A. Figueiredo, I. Gomes-Filho, R. Silva, P. Pereira, F. Mata, A. Lyrio, E. Souza, S. Cruz, and M. Pereira. Maternal anemia and low birth weight: a systematic review and meta-analysis. *Nutrients*, 10(5):601, 2018.

R. M. V. Flores and L. G. D. d. S. Scorzafave. Effect of racial segregation on proficiency of brazilian elementary school students. *EconomiA*, 15(1):20–29, 2014.

R. G. Fryer. Racial inequality in the 21st century: The declining significance of discrimination. In *Handbook of labor economics*, volume 4, pages 855–971. Elsevier, Amsterdam, 2011.

R. G. Fryer and S. D. Levitt. Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2):447–464, 2004.

R. G. Fryer and S. D. Levitt. Testing for racial differences in the mental ability of young children. *American Economic Review*, 103(2):981–1005, 2013.

P. Glewwe and E. Miguel. The impact of child health and nutrition on education in less developed countries. forthcoming in the handbook of development economics, vol. 4, edited by t. paul schultz and john strauss eds. *Paul Schultz and John Strauss eds*, 2008.

A. Goodman, D. L. Lamping, and G. B. Ploubidis. When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the strengths and difficulties questionnaire (sdq): Data from british parents, teachers and children. *Journal of Abnormal Child Psychology*, 38(8):1179–1191, Nov 2010. ISSN 1573-2835. doi: 10.1007/s10802-010-9434-x. URL https://doi.org/10.1007/s10802-010-9434-x.

J. Guryan and K. K. Charles. Taste-based or statistical discrimination: The economics of discrimination returns to its roots. *The Economic Journal*, 123(572), 2013.

T. W. Harding, V. De Arango, J. Baltazar, C. E. Climent, H. Ibrahim, L. Ladrido-Ignacio, and N. Wig. Mental disorders in primary health care: a study of their frequency and diagnosis in four developing countries. *Psychological medicine*, 10(2):231–241, 1980.

J. J. Heckman. Detecting discrimination. *The Journal of Economic Perspectives*, 12(2):101–116, 1998.

J. J. Heckman. The american family in black & white: A post-racial strategy for improving skills to promote equality. *Daedalus*, 140(2):70–89, 2011.

G. Hirata and R. Soares. Competition and the racial wage gap: Testing becker's model of employer discrimination, 2016.

M. S. Kramer. Determinants of low birth weight: methodological assessment and meta-analysis. *Bulletin of the world health organization*, 65(5):663, 1987.

M. C. Lu and N. Halfon. Racial and ethnic disparities in birth outcomes: a life-course perspective. *Maternal and child health journal*, 7(1):13–30, 2003.

L. J. Marteleto and M. Dondero. Racial inequality in education in brazil: a twins fixed-effects approach. *Demography*, 53(4):1185–1205, 2016.

R. S. Matos and A. F. Machado. Diferencial de rendimento por cor e sexo no brasil (1987-2001). *Revista Econômica*, 8(1), 2006.

N. A. Menezes-Filho. *Os determinantes do desempenho escolar do Brasil.* IFB, São Paulo, 2007.

N. Morisaki, I. Kawachi, E. Oken, and T. Fujiwara. Social and anthropometric factors explaining racial/ethnical differences in birth weight in the united states. *Scientific reports*, 7:46657, 2017.

R. A. Ødegård, L. J. Vatten, S. T. Nilsen, K. Å. Salvesen, and R. Austgulen. Preeclampsia and fetal growth. *Obstetrics & Gynecology*, 96(6):950–955, 2000.

A. M. H. C. d. Oliveira and E. L. G. Rios-Neto. Tendências da desigualdade salarial para coortes de mulheres brancas e negras no brasil. *Estudos Econômicos*, 36(2):205–236, 2006.

A. L. Portella, T. B. Bussmann, and A. M. H. d. Oliveira. A relação de fatores individuais, familiares e escolares com a distorção idade-série no ensino público brasileiro. *Nova Economia*, 27(3):477–509, 2017.

S. Rahmati, A. Delpishe, M. Azami, M. R. H. Ahmadi, and K. Sayehmiri. Maternal anemia during pregnancy and infant low birth weight: A systematic review and meta-analysis. *International Journal of Reproductive BioMedicine*, 15(3):125, 2017.

M. A. Rangel. Is parental love colorblind? human capital accumulation within mixed families. *The Review of Black Political Economy*, 42(1-2):57–86, 2015.

B. L. Silverman, T. Rizzo, O. C. Green, N. H. Cho, R. J. Winter, E. S. Ogata, G. E. Richards, and B. E. Metzger. Long-term prospective evaluation of offspring of diabetic mothers. *Diabetes*, 40 (Supplement 2):121–125, 1991.

M. d. F. Simões and M. E. Ferrão. Competência percebida e desempenho escolar em matemática. *Estudos em Avaliação Educacional*, 16(32):25–42, 2005.

J. F. Soares. Measuring cognitive achievement gaps and inequalities: The case of brazil. *International Journal of Educational Research*, 45(3):176–187, 2006.

J. F. Soares and M. T. G. Alves. Desigualdades raciais no sistema brasileiro de educação básica. *Educação e Pesquisa*, 29(1):147–165, 2003.

E. Telles. *Race in another America.* Princeton University, Princeton, NJ, 2004.

P. E. Todd and K. I. Wolpin. On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):F3–F33, 2003.

P. E. Todd and K. I. Wolpin. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*, 1(1):91–136, 2007.

C. G. Victora, C. L. P. Araújo, A. M. B. Menezes, P. C. Hallal, M. d. F. Vieira, M. B. Neutzling, H. Gonçalves, N. C. Valle, R. C. Lima, L. Anselmi, et al. Methodological aspects of the 1993 pelotas (brazil) birth cohort study. *Revista de saude publica*, 40(1):39–46, 2006.