Effects of school quality on health and fertility *

Ozkan Eren, University of California-Riverside

Krzysztof Karbownik, Emory University

Orgul Ozturk, University of South Carolina

May 11, 2026

Abstract

Using administrative data from South Carolina and a regression discontinuity design, we present novel evidence on the effects of school quality on health, fertility, and mortality in early adulthood as well as health of the next generation. Students that attended improving schools have lower likelihood of hospitalization (16%), ED visit (5.0%), dying (25%) and having children (9.8%) at ages 18 to 28. For those with children we do not find differential gains in birth outcomes of the next generation. Candidate mechanisms include improved cognitive ability, school satisfaction, short-run health and reductions in risky behaviors while we can rule out meaningful changes in extensive margin of education, employment, income, and welfare utilization.

JEL: I20, I28, I10, I14

Keywords: School quality, education, health, mortality, fertility, birth outcomes, intergenerational transmission, accountability, regression discontinuity

*Eren: Department of Economics, University of California-Riverside (email: ozkane@ucr.edu); Karbownik: Department of Economics, Emory University, CesIfo, Hoover Institution, IZA, and NBER (email: krzysztof.karbownik@emory.edu); Ozturk: Department of Economics, University of South Carolina (email: odozturk@moore.sc.edu). We thank David Figlio, Maya Rossin-Slater and Heather Royer as well as seminar participants at Catholic University of Milan, Stanford University, and Tel Aviv University for helpful comments. The conclusions expressed in this paper are those of the authors and do not represent the positions of State of South Carolina or local agencies. All errors are our own.

1 Introduction

There is large variation in school quality in the U.S. both across (Card and Krueger, 1992) as well as within (Fahle and Reardon, 2018) states, and even within school districts (Figlio and Karbownik, 2017). Such educational inequality is not unique to the US and has been documented in both developed (see e.g., (Bradley et al., 2004) for Australia or Agasisti and Cordero-Ferrera (2013) for Italy and Spain) and developing (see e.g., Asadullah and Yalonetzky (2012) for India or Lucas and Mbiti (2014) for Kenya) countries. To the extent that school quality causally affects long-run outcomes, these inequities in inputs could translate to meaningful inequality in adult outcomes. An obvious problem establishing such causal relationship is the fact that school quality is not randomly assigned. In contrast, it is correlated with, among others, neighborhood (Wodtke et al., 2023), demographic (Reardon et al., 2024), socioeconomic (Jang and Reardon, 2019), and even health (Jessica and Domingue, 2023) characteristics. At the same time, to the extent that such causal relationship exists, it opens promising avenues for policy interventions and experimentation, which could reduce long standing societal gaps.

Two non-market outcomes that have long been considered as moderated by education are health and fertility. Individuals with less education have worse health consistent with the former playing a crucial role in the demand for health model (Grossman, 1972). This observation has been confirmed in multiple countries, both developed (Marmot, 2015) and developing (Le and Nguyen, 2020), and across subjective (Ross and Wu, 1995) and objective (Meara et al., 2008) measures of health. Similarly, seminal work by Becker (1960) predicts that more educated women will have fewer children, due to increased opportunity cost, a data pattern observed in both developed (Zang et al., 2022) and developing countries (Kebede et al., 2019).

Despite strong theoretical foundations and overwhelming descriptive evidence, causal estimates of the effects of education on health (Galama et al., 2018) and fertility (Vasireddy et al., 2023) remain mixed and illusive. They appear to be sensitive to measurement, specific iden-

tification strategies, and locations. For example, considering health outcomes, Lleras-Muney (2005) shows improvements, Clark and Royer (2013) show no effects, while Barcellos et al. (2023) even show declines in health. When it comes to fertility, Sohn and Lee (2019) find declines in childbearing, McCrary and Royer (2011) find no effects, while Virtanen et al. (2024) find increases in fertility. One common feature of all these studies is that they use exogenous variation in the extensive margin of education e.g., being forced to stay in high school one more year or attending newly accessible college. At the same time, despite the aforementioned huge variation in school quality, there is a dearth of evidence on the effects of changes in the intensive margin of education. Put differently, it is entirely plausible that it is the amount of learning (intensive margin) that matters for long-run outcomes rather than an additional year(s) (extensive margin) one spends in schooling.

Considering the intensive margin, rather than or in addition to the extensive margin, could be important for at least two reasons. First, theoretical models by Becker (1960) and Grossman (1972) require that education increases the opportunity cost and makes people more efficient health producers, respectively. To the extent that schooling quality rather than quantity could affect these factors (more), there is a theoretical reason to investigate this conjecture empirically. Second, some of the null extensive margin results present a conundrum whereby education, commonly thought of as a panacea for improving health (Acemoglu and Johnson, 2007) and reducing fertility (Li and Zhang, 2007) in order to increase economic growth, might not be an actionable tool for policy makers. This naturally extends beyond developing countries as policy makers might likewise be concerned with health inequality (Link and Phelan, 1995) or cost of healthcare (Papanicolas et al., 2018).

In this paper we seek to understand the role of school quality in health production function and fertility decisions. Specifically we ask if improvements in high school quality boost health outcomes in short- and long-run? Using administrative data, we measure the former when individuals are still in high school while the latter at ages 18 to 28. We also ask if these improvements affect fertility and health at birth of the next generation? Given affirmative

evidence for the first two research questions we then turn to studying mechanisms behind the causal effects we uncover. Specifically, we ask if school quality affects extensive margin of education, labor market outcomes, welfare utilization, and criminal activity. Finally, we investigate if these effects differ across demographic, socioeconomic and cognitive dimensions including: gender, race/ethnicity, poverty, and baseline test scores.

Our empirical approach is a regression-discontinuity design based on a targeted policy stemming from South Carolina's school accountability rules. We compare students in schools that barely missed the cutoff for "average rating" (out treated group) to students in schools that barely passed this cutoff (our control group). Under the policy the former set of institutions is pressured to develop improvement plans explicitly designed to boosting educational quality, academic performance, and character development which can proxy for behavioral skills. Thus, we first document that schools that barely missed the cutoff indeed improve student learning compared to schools that barely made the cutoff. Students in the treated schools are 4.8% more likely to pass 10th grade exit exams, 11% more likely to obtain test score that deems them proficients, have 0.42 SD higher subjective perceptions about their educational experiences in school, and have 0.55 SD higher character development score. On the other hand, we find no changes in the extensive margin of education and based on 95% confidence interval can rule out increases in years in high school as small as 2.1%.

Having established that the policy improves schooling quality but does not change schooling quantity, we track students educated in the treated and the control group schools into adulthood where we observe their health and fertility outcomes. Treated students are 16.0% less likely to be hospitalized, 5.0% less likely to visit emergency department, and 25% less likely to die between ages 18 and 28 compared to students educated in control schools. We also observe large reductions in the intensive margin of hospital and ED utilization. When it comes to fertility, the probability of having children by age 28 declines by 9.8%. On the other hand, we find no statistically significant effects on the health at birth of children born to parents educated in treated compared to control group schools.

We consider several mechanisms that could explain our findings and find support for improvements in short-run (ages 15 to 18) health and cognition (measured both as higher test scores on exit exams and lower likelihood as being labeled as learning disabled) as well as reductions in risky behaviors through juvenile criminal activity. At the same time, we can rule out meaningful changes in employment, earnings, and welfare utilizationin adulthood as well as extensive margin of education. Our conservative back-of-the-envelope calculation using average costs of hospital stay and emergency department visit suggests that improving school quality of failing schools in South Carolina saved the state at minimum USD 225M between 2004 and 2019 with relatively minimal educational costs. Additionally, in aggregate, individuals themselves likely saved at minimum USD 11M based on information on Medicare deductibles (CMS, 2014).

Our results are robust to alternative specifications including non-parametric estimation and adding control variables as well as samples including donut RD and changes in bandwidth used. Furthermore, they do not change when we expand the sample to include Covid-19 years or extend observation period to age 30. Statistical significance is unaffected by either alternative standard errors adjustments or randomization inference.

Our primary contribution is to the literatures investigating the effects of education on health and fertility. Although much prior work considers each of these dimensions separately, here we investigate them simultaneously together with exogenous variation in school quality rather than school quantity. To the best of our knowledge there is no prior research on the effects of school quality on both those outcomes for the same population and there is no prior research on the effects of improving quality of failing schools for either of those outcomes. We view this as an important gap in the literature given large variation in school quality in many countries, increasing number of failing schools in the US, and the (low) likelihood that policymakers will be able to change extensive margin of schooling via either increase mandatory schooling age or investments in school/college expansion which have been commonly used as sources of exogenous variation to date. Through our mechanisms

analyses, we also extend the limited literature on the returns to school quality in terms of crime, labor market outcomes, and welfare participation.

First, we contribute to the voluminous body of work on the effects of education on health (see e.g., Zajacova and Lawrence (2018), Xue et al. (2021) or Lleras-Muney (2022) for recent reviews). Despite robust descriptive evidence suggesting that schooling is positively associated with multiple health outcomes in short- and long-run, causal evidence is mixed and limited. Prior work has predominantly relied on regression discontinuity/instrumental variables/difference-in-differences estimation using school leaving age reforms (e.g., Lleras-Muney (2005) for the US, Clark and Royer (2013) for the UK, Jiang et al. (2020) for China, Malamud et al. (2023) for Romania, or Brunello et al. (2013) for multi-country study in Europe), genetic data and Mendelian randomization (e.g., , Ding et al. (2019) for the US, Lehtimäki et al. (2022) for Finland, or Howe et al. (2023) for the UK and Norway), or sibling/twin fixed effects (e.g., Lleras-Muney et al. (2022) for the US, Gensowski and Gørtz (2024) for Denmark, or Amin et al. (2013) for the UK). The findings in these studies range from positive effects (e.g., Lleras-Muney (2005) or Howe et al. (2023)), through zero effects (e.g., Clark and Royer (2013) or Malamud et al. (2023)), to even some negative effects (e.g., Barcellos et al. (2023) for blood pressure or Palme and Simeonova (2015) for breast cancer). Other less commonly used approaches rely on variation in compulsory school openings (Aaronson et al. (2021) for historical US context and Mazumder et al. (2023) for Indonesia), college openings (Fletcher and Noghanibehambari (2024) and González et al. (2024)), the Vietnam Draft lottery (Grimard and Parent (2007) and de Walque (2007)), or abolishing of school fees (Barlow, 2021). Here likewise the results range from null to positive effects.

All of the aforementioned studies examine the extensive margin effects of schooling i.e., these papers seek to utilize credible variation in an additional year of schooling or in access to schooling at a given level. Our approach is fundamentally different in that we ask if increasing *quality* of schooling improves health. To the best of our knowledge this intensive margin question has not been examined in the extant literature while papers most similar

to ours rely on admission cutoffs to elite high schools in Norway (Bütikofer et al., 2023), school tracking reforms in UK and France (Basu et al., 2018; Canaan, 2024) or variation in school financing in the US (Rauscher et al., 2025). Although Rauscher et al. (2025) focuses on U.S. context, they can only investigate mortality at an aggregate level while children are in schooling. Our work presents the first evidence that improving failing schools could lead to long-run gains in health, a positive externality underappreciated when considering cost-benefit analyses of the education reform, and to changes in fertility.

We also measure health differently. The vast majority of papers rely either on administrative mortality/longevity data (e.g., Lleras-Muney (2005) or Lundborg et al. (2016)) or on survey data with mostly self-reported health measures (e.g., Jiang et al. (2020) or Mazumder et al. (2023)). The former can be viewed as an extreme and very long-run outcome while the latter is prone to measurement error (Bharadwaj et al., 2017) and Likert-scale estimation biases (Bond and Lang, 2019). In contrast, we use administrative data on all emergency department visits and hospitalizations in South Carolina between 2005 and 2019. This data does not suffer from measurement error, is more relevant for young adults given their low levels of mortality, and is policy relevant generating financial costs for both patients (Dobkin et al., 2018) and insurers (Cooper et al., 2019). We are not aware of any other work causally linking schooling and hospitalizations in the US while it is plausible that findings based on research on European healthcare systems where the use of administrative data is more common (Leuven et al., 2016; Malamud et al., 2023; Gensowski and Gørtz, 2024) might not apply to the US.

Second, we also contribute to the literature examining the effects of education on fertility. Much of this literature likewise relies on compulsory schooling laws to generate exogenous variation in education levels which is then linked to various measures of childbearing. The findings range from negative (e.g., Cygan-Rehm and Maeder (2013) in Germany or Fort et al. (2016) in England) to null (e.g., McCrary and Royer (2011) in the US and Fort et al. (2016) in multiple continental Europe countries) effects. There is even some evidence – based on

exogenous variation in access to college (Virtanen et al., 2024) – that more education could increase fertility. Other work that is closer to the intensive margin of schooling that we are interested in, studied lengthening of school year in Norway (Abel et al., 2024), shortening of school year in Germany (Koebe and Marcus, 2022), admission to elite high schools in Scotland (Clark and Del Bono, 2016), and private tutoring bans in China (Meng et al., 2025). These studies broadly support inverse relationship between education and fertility although in all cases they operate partially through changes in extensive margin of education. To the best of our knowledge, this is the first paper to directly study the effects of improving school quality on fertility.

Third, we contribute to the literature on the intergenerational effects of education. Here likewise much research relies on variation in the extensive margin of education while the two papers closest to our work rely on variation in school curricula. Cuevas Ruiz et al. (2023) study the effect of mothers educated using more general knowledge curriculum on health of their babies and find reductions in very low births weight and very preterm births. Similarly Dursun et al. (2024) study high school curricular reforms in the US and show that these reforms significantly reduced incidence of low birth weight and prematurity for Black (but not White) mothers. Findings from the curricular reforms which could have also improved learning and school quality – although neither paper documents such effects directly – are generally more positive than results based on variation in extensive margin of education where findings are mixed. On the one hand, Currie and Moretti (2003) finds positive effects on birth outcomes of access to higher educations in the US while Chou et al. (2010), Dursun et al. (2022), Grytten et al. (2014) find positive effects on birth outcomes of compulsory schooling laws in Taiwan, Turkey, and Norway, respectively. On the other hand, Lindeboom et al. (2009), McCrary and Royer (2011), and Arendt et al. (2021) all report null effects of such reforms in UK, US, and Denmark, respectively. We contribute to this literature by offering the first evidence on interngerational effects of school quality rather than quantity.

More broadly, we build on the existing school accountability literature (see e.g., Figlio

and Loeb (2011) or Mittleman and Jennings (2018) for recent reviews). Although ours is not a study evaluating the effects of school accountability on outcomes beyond education, we leverage parts of the South Carolina accountability legislation to obtain exogenous variation in school quality improvement at the bottom of the quality distribution. Prior work showed that school accountability improves student test scores (Hanushek and Raymond, 2004, 2005; Rockoff and Turner, 2010) and that these effects persist into graduation (McElroy, 2023) and earnings (Deming et al., 2016). These papers generally use variation in either the introduction of school accountability or changes to how the accountability system is implemented, and thus answer research questions related to efficacy of the system. Our approach is different, in that we utilize rules within the existing system that force certain set of schools to improve their educational offering while allow other schools (control group) to conduct their business as usual.

A separate strand of this research investigates how accountability can change school production function affecting disciplinary actions (Figlio, 2006), instructional practices (Rouse et al., 2013), teacher perceptions (Reback et al., 2014), and teacher labor supply (Dizon-Ross, 2020). We likewise find that students in affected schools have improved educational outcomes and satisfaction and we additionally link these changes to objective short- and long-run health outcomes. Another paper examining externalities, Eren et al. (2023), focuses on the effects of the South Carolina policy on crime. Understanding the non-educational effects of education policies is important to fully understand their welfare implications.[1]

We view our results are having three important policy implications. First, documented improvements in short- and longer-run health suggest that by focusing only on educational (or labor market) outcomes policymakers might underestimate the economic benefits of school improvement. Second, across various dimensions of health - including mental health - we do not find any adverse effects. This is important given prior work suggesting that school

[1]Beyond human capital multiple papers examine the effects of school accountability and more broadly school quality on housing values e.g., Black (1999) and Figlio and Lucas (2004) or more recently Lafortune and Schönholzer (2022).

accountability could increase psychiatric drugs prescriptions (Bokhari and Schneider, 2011; King et al., 2014) while increased instructional time can increase student's stress-related health problems (Marcus et al., 2020). In our setting we do not observe such trade-offs. Third, these positive externalities appear to be particularly relevant now when many states and countries are increasing the accountability pressures on schools.

2 Institutional Setting

In 2000, South Carolina Department of Education (SCDE) implemented a statewide accountability system that assigns each public school to one of five performance categories: Unsatisfactory, Below Average, Average, Good, and Excellent. These categories are determined by a continuous measure known as the accountability performance score. During the period we consider, a high school's performance score is calculated using a weighted sum of four components: (i) the fraction of tenth-grade students meeting exit exam standards (20 percent), (ii) longitudinal exit exam performance (30 percent), (iii) the share of students eligible for merit-based LIFE scholarship at a four-year institution (20 percent), and (iv) the graduation rate (30 percent).[2] The accountability score, which ranges from 1 to 5, determines the category assigned to each school. For example, in the 2000–2001 academic year, schools with scores below 2.2 are classified as Unsatisfactory, while a score of 2.2 corresponds to Below Average category.[3]

The rating formula underwent a number of modifications in the early years of the accountability system. For instance, the graduation rate component was incorporated into the overall score starting in the 2002–2003 academic year.[4] These adjustments to the scoring

[2]From 1986 to 2005, the state used the Basic Skills Assessment Program as its exit exam, requiring students to pass reading, writing, and math. Longitudinal performance is defined as the share of students passing by twelfth grade, while LIFE scholarship eligibility is based on meeting state GPA and SAT/ACT criteria.

[3]The threshold separating the Unsatisfactory and Below Average categories increased over time, rising to 2.3 in 2003, 2.4 in 2004, and 2.5 in 2006.

[4]High school performance in the first two years of the accountability system was based on three components: the share of tenth-grade students meeting exit exam standards (30 percent), longitudinal exit exam

methodology are important, as they reduced schools' ability to precisely manipulate their scores around the rating thresholds. Moreover, the accountability regime in South Carolina constrained strategic behavior by restricting exemptions from high-stakes testing only if students' circumstances related to disabilities or Limited English Proficiency were in line with federal guidelines. The SCDE releases these ratings to parents, as part of school report cards, and the general public by mid-November, roughly two weeks after providing them to schools.

The SCDE linked performance ratings to both incentives and corrective actions. Schools assigned a Below Average or Unsatisfactory rating are required to prepare improvement plans in collaboration with an external review team composed of representatives from the state and selected school districts, as well as retired educators, parents, and community members. These plans are expected to outline strategies to improve academic and behavioral outcomes, incorporate professional development for teachers, and specify a timeline for implementation. Furthermore, school ranked Unsatisfactory – which by default fall under the scrutiny of the state's oversight committee – had to implement a "Character Development Program" which must include instruction in respect, responsibility, fairness, caring, trustworthiness, and citizenship (SC Code of Laws Section 59-18-350).

Based on the review team's recommendations, the state and district may also assign teacher or principal specialists to underperforming schools. These specialists provide targeted support, including the development of research-based instructional practices tailored to student needs, facilitation of teacher learning groups, classroom observation with feedback, and instructional modeling. For example, teacher specialists often serve as instructional coaches, helping teachers strengthen classroom practices, develop new skills, and use assessment data to identify and implement necessary changes in instructional strategies. In addition, the SCDE introduced grant programs targeted at schools rated Unsatisfactory and Below Average. These funds are intended to support necessary retraining of faculty and

performance (30 percent), and the proportion of students eligible for the LIFE scholarship at a four-year institution (40 percent).

administrators to help schools meet established standards and implement effective practices.

Although South Carolina's accountability system provides similar targeted assistance to both schools rated Unsatisfactory and Below Average, which the exception of mandatory "Character Development Program" for the former set of schools, the associated pressures are likely stronger for those rated Unsatisfactory given the consequences attached to that designation. In particular, schools receiving an Unsatisfactory rating, absent sufficient progress, may face leadership changes, restructuring, or state takeover (South Carolina Education Oversight Committee 2000–2003, Article 15). More broadly, the system is designed to exert intense pressure on administrators in these schools to improve student achievement through increased scrutiny, reputational concerns, and the threat of job loss (Rouse et al., 2013). Finally, based on anecdotal evidence SCDE mostly intervenes with Unsatisfactory but not with Below Average schools.[5]

3 Methods

3.1 Data

3.1.1 Data Sources

The data for this study are compiled from several different sources. We use the administrative records from the South Carolina Department of Education, which contain information on student race, gender, age, and free or reduced-price lunch status, as well as test scores for selected grades. Unique student identifiers allow us to track individuals throughout their enrollment in the public school system beginning in Fall 2000.

To study the effects of receiving a lower accountability rating on health and fertility outcomes, using a common individual identifier, we link student records to administrative

[5]On the incentive side, schools with high performance ratings or improvement are eligible for additional funding, although the amounts are modest—for example, average awards were about $4,500 per school in 2002–2003 (South Carolina Education Oversight Committee 2000–2003). Since we are focusing on failing schools in this paper such incentives are not relevant for our sample.

data from the South Carolina Hospital Encounters (SCHE) and the South Carolina Vital Statistics (SCVS). The SCHE data include discharge-level information on all inpatient hospitalizations and emergency department (ED) visits from 2005 to 2022. The SCVS data, available over the same period, provide birth records that allow us to examine fertility among female students. This information is based on maternal and hospital worksheets completed shortly after birth. Students who do not match to any of these sources are coded as zeros for the relevant outcome variables.

To examine mechanisms underlying the observed effects of school quality, we further link these data to administrative information from (i) the South Carolina Department of Juvenile Justice (SCDJJ), (ii) the South Carolina Department of Social Services (SCDSS), and (iii) the South Carolina Department of Employment and Workforce (SCDEW). The juvenile crime data include date of offense and arrests by category of crime from 2002 to 2017. From the SCDSS, we use information on the receipt of food stamps and Temporary Assistance for Needy Families (TANF) to capture long-run dependence on government assistance. We obtained quarterly earnings records (based on the South Carolina Unemployment Insurance) from the SCDEW through 2022. Finally, we rely on publicly available school report cards for information on schools' performance ratings (Unsatisfactory, Below Average, etc.), their overall score, its components, and other school-level attributes such as measures of school climate, character development scores, and graduation rates.

To avoid potential confounding effects from the COVID-19 pandemic, we restrict our analysis sample to administrative records through the end of 2019. However, as discussed in Section 5, the results are robust to including COVID-19 years. Our analysis focuses on first-time ninth graders entering high school between the 2000–2001 and 2005–2006 academic years, roughly corresponding to cohorts born between 1985 and 1991. We assign students to the first high school they attended. Doing so circumvents concerns related to endogenous responses from students and parents such as transferring to another school following a low

performance rating and it gives our results an intent-to-treat interpretation.[6] In addition, this sample restriction ensures that individuals can be followed into mid-adulthood to measure long-run outcomes. For example, in our main analysis sample, fertility is observed through age 34 for those born in 1985, with the observable horizon declining by one year for each successive cohort (to age 28 for the youngest cohort born in 1991). Thus, in the main analysis to maintain balanced panel, we restrict follow-up age to 28 for all cohort.

3.1.2 Outcomes

Using the administrative data, we construct several short- to long-run outcomes for each student. Adult health outcomes include (i) an indicator for ever being hospitalized as an inpatient, (ii) number of hospitalizations, (iii) total number of days hospitalized, (iv) an indicator for any ED visits, (v) number of ED visits, and (vi) mortality–an indicator for patient discharge codes indicating expired status and discharge to hospice. We further classify adult health outcomes by disease categories based on the International Classification of Diseases, Ninth Revision Standard (ICD-9). To examine short-run health effects, we define analogous measures prior to age 18. Finally, we consider several different measures of fertility, including a dummy variable for having a birth, number of births, teen birth, and age at first birth. In order to avoid conditioning on outcome we assign zeros to intensive margin measures of fertility for those who do not have children.

As part of our mechanism analysis, we examine a school's receipt of a lower rating on youth criminal involvement by defining an indicator for ever being arrested as a juvenile. For academic performance, we use a dummy variable for performing at or above the "proficient level" on the state's mandatory exit exam.[7] We further complement these analyses using several school-level aggregate measures such as graduation rate and fraction of students eligible for LIFE scholarship. In order to measure any changes in school climate following the

[6] We also examine school switching and school leaving as educational outcomes.

[7] South Carolina repealed the Common Core Standards in 2015. As part of this policy change, the state discontinued its exit exams. This decision was part of a broader effort to develop state-specific educational standards and implement assessments better suited to identifying areas of academic strength and weakness.

receipt of a lower rating, as in Eren et al. (2023), we take advantage of a survey administered annually. As part of the survey, students were asked about their satisfaction level with the school's learning, social and physical environment, and home/school relationship. Furthermore, we collected information on character development scores assigned to each school on annual basis which range from 0 to 4, with 4 being "Excellent". We derive this information using school report cards.

Using the SCDSS data, we also constructed indicators for whether the student ever received welfare benefits (food stamps or TANF) during adulthood. Finally, we analyze labor market outcomes by defining indicators for employment and average annual earnings, both of which are measured at different ages. Years with no observed earnings are coded as zero. All dollar amounts are adjusted to 2015 dollars.

3.1.3 Descriptive Statistics

Online Appendix Table A1 reports summary statistics for the full sample (Column 1) and separately for schools at the lower end of the ratings distribution (Columns 2 and 3). Panel A shows that Black and White students account for 41 and 55 percent of the all students sample, respectively, with a much higher share of Black students in schools rated Unsatisfactory and Below Average. The share of students eligible for free or reduced-price lunch is also substantially higher in lower-rated schools. Students in Unsatisfactory schools have lower baseline (eighth-grade) English Language Arts (ELA) and math test scores than those in schools rated Below Average, and both groups are at least 0.38 SD below the state average.

Panel B presents long-run outcomes. Adult hospitalization rates are around 16 percent, and the total number of days hospitalized is slightly above one day for students who attended Unsatisfactory and Below Average schools in ninth grade. Emergency Department visits are more common than hospitalizations, with about 80 percent of these students experiencing at least one adult ED visit at ages 19 to 28. The fraction of female students with a biological child is 63 percent among those who attended high schools rated Unsatisfactory in ninth

grade, compared to 61 percent for those in schools designated as Below Average. Both of these values are higher than statewide average in Column 1. Teen birth is also more common among students at the bottom end of the ratings distribution, with the highest rate in schools rated Unsatisfactory (at nearly 21 percent).

3.2 Estimation

We combine the regulation and the administrative data into a reduced-form regression discontinuity design. The estimating equation is

$$Y_{ics} = \alpha + \beta_{RF}\text{Below}_{cs} + f(\text{Score}_{cs}) + \text{Below}_{cs} \times f(\text{Score}_{cs}) + \varepsilon_{ics} \tag{1}$$

where Y_{ics} is an outcome of interest (e.g., hospitalization) observed for individual i from entry cohort c at high school s. Our treatment variable – Below_{cs} – takes value of 1 if school is below the "Unsatisfactory" performance school cutoff and 0 otherwise. This treatment varies at school-by-year level meaning that school can change their position around the cutoff over time.

Given the semi-transitory nature of the shock, we define a cohort (c) in our data as first time 9th graders in any given school year which we then follow over time through high school and into their adulthood up to the age of 28. Thus, in our data, we consider six cohorts entering high school for the first time between school years 2000-01 and 2005-06 whose outcomes we observe for years 2000 to 2019. We balance this sample which means that we follow each entering cohort for 14 years until they are 28 which is 2014 for the first and 2019 for the last cohort. We generally consider observations during high school (ages 14/15 to 17/18) as short-run outcomes while observations after expected on time high school graduation as long-run outcomes. Thus, for most analyses we collapse these periods to a single data point meaning that we measure hospitalization in adulthood as any hospitalization between ages 19 and 28 while the short-run hospitalizations are coded as any

hospitalization between grades ages 14/15 to 18.

Our estimation strategy faces two related data challenges when it comes to the running variable: discreteness and sparsity. Namely our running variable is not only discrete which calls for specific set of estimation tools but it is also sparse meaning that it has small number of data points. Because of this, in the main analysis, we rely on a global linear polynomial fit with a uniform kernel as advised by Cattaneo et al. (2024). We center our running variable around the "Unsatisfactory" vs. "Below average" rating cutoff, which we define as zero, and consider all schools in the range of -1.3 to 0.3 points around this recentered cutoff. The asymmetry in distance to the cutoff on either side is due to the fact that score of "cutoff value" + 0.3 points qualified school for the next best i.e., "Average" designation level. In Section 5 we further present results using local linear estimation, data driven non-parametric estimation (Cattaneo et al., 2018), and local randomization approach. We note that the last approach estimates treatment effect in the neighborhood of the cutoff rather than at the cutoff as the other methods do. Finally, given that our three variables of interest are count variables with wide count range – number of hospitalizations, number of ED visits, and days spend in hospital – we use Poisson Pseudo Maximum Likelihood (PPML) for these estimations. These coefficients can be interpreted as semi-elasticities of healthcare utilization and the estimation method handles well present of zeros in the dependent variable (Silva and Tenreyro, 2011). Another useful feature of PPL is that it allows for a proportional analysis that ensures the comparability of estimates in short- and long-run e.g., the baseline risk of any hospitalization in our adult sample is 16% while it is only 4% during high school years.

3.3 Identification

In order to causally identify parameter of interest, β_{RF}, we need to assume that potential outcomes of interest are smooth across the policy cutoff which for most analyses in this paper falls at the "Unsatisfactory" vs. "Below average" margin. Although this assumption is not testable, as is common in the RDD literature (Lee and Lemieux, 2010; Cattaneo et al.,

2019, 2024) we provide empirical evidence that is consistent with the assumption based on density and balance of covariates tests. Figure 1 presents the distribution of schools across the cutoff with the size of the bubble referencing the density. Two important conclusions steam from this graph. First, the discontinuity is sharp meaning that no schools scoring below the cutoff are rated as "Below average" and no schools above the cutoff are rated as "Unsatisfactory". Second, the density appears to be smooth around the cutoff. Formal tests for "no manipulation" condition confirm this with p-values of 0.47 and 0.11 based on procedures proposed by Cattaneo et al. (2018) and Frandsen (2017), respectively. The latter test is particularly suited for discrete running variables.

In Tables 1 and A2 we further present multiple tests of covariates balance. The former table executes our preferred RDD analysis (panel A) and local randomization approach analysis (panel B) for two demographic characteristics, a SES proxy, and baseline test scores. Out of the 10 estimates, none is statistically significant at conventional levels and they are all quantitatively small compared to the control group mean. Table A2 then presents balance in predicted adult outcomes of interest likewise using these two different econometric approaches. Here we predict all main outcomes of interest using pre-determined characteristics including gender, race, free and reduced price lunch, test scores, and cohort fixed effects and then estimate the models using predicted rather than true outcomes.[8] Out of 18 estimates none are statistically significant at conventional levels and none imply imbalances larger than 1.6% of the control group mean. These results stand in stark contrast to the estimates presented in Section 4 where we use real outcomes. Overall, both tests are consistent with the identifying assumption and suggest that our estimates are unlikely to be biased.

[8]The results are very similar when we exclude test scores and estimate the regression on a larger sample for which demographic variables but not student assessments are available.

4 Results

4.1 Effects of the Policy on School Quality and Educational Outcomes

Before moving to long-run health and fertility we first present evidence that the policy "had a bite" and indeed led to improvements in school quality. We measure this using multiple educational outcomes at both school and individual level which reflect both intensive and extensive margins of education. This is important because to the extent that the policy does not affect schooling outcomes and experiences we could not claim that it changed school quality and expect it to influence health and fertility, or at least not through the quality dimension. Table 2 presents these results. Panel A presents aggregate school-level findings while Panel B uses individual level data. We note that data for several outcomes are missing for subset of years but the results are substantively unchanged when we redo the analysis for the smallest common denominator. Dependent variables in the top panel represent the three inputs into the performance score (columns 1 to 3), an aggregate index of these components (column 4), students' satisfaction index (column 5) and character development score (column 6). Grade 10 student satisfaction index is not an input into the performance index calculations and reflects self-reported satisfaction of students with leaning environment, social and physical environment as well as home-school relationship. Similarly, the character development score does not enter accountability formula but it is directly targeted by the policy at schools labeled "Unsatisfactory" All six estimates are positive while five of those are statistically significant at conventional levels. Focusing on the summary indexes, the academics improve by 31% of a SD, student subjective perceptions about schooling improve by 42% of a SD, and character development improves by 55% of a SD. Note that although the last outcome is directly targeted by the policy (and does improve), the student subjective perceptions are not targeted by the policy and should be thought of as reflecting broader improvements in school climate and quality as a result of the changes that the policy triggers.

Moving to individual level data we have proxy for achievement which is defined as reaching proficient level on 10th grade exit exams. Despite much smaller sample size we find positive and statistically significant effects on academic achievement among students attending treated compared to control schools. The effect size of 11% compared to control group mean appears meaningful and serves at the best example we can offer using individual level data of the policy improving student ability.

In subsequent columns we consider various measures of the extensive margin of education. First, we consider continuous enrollment as a result of the policy and we do not find differential persistence in schooling. Unsurprisingly then we also do not observe differences in years spend in high school between treated and control students. These coefficients are generally small and often "incorrectly signed". Second, we also considered school switching which could be a result of stigma from attending school labeled as "failing" or differential demand for school quality. We do not find any meaningful effects for these variables either. Based on these analyses we conclude that our results should not be driven by changes in the extensive margin of education that was investigated in almost all prior work but rather they are driven by changes in intensive margin of education via improved achievement, behavioral outcomes, and satisfaction.

One wrinkle in this interpretation could be a seemingly contradictory finding on positive and statistically significant effect on graduating high school in 4 years at school level (column 3) and negative and statistically insignificant effect on years in high school measured at individual level. A simple explanation for this is the fact that in South Carolina it is possible to enroll in the last year of high school but not necessarily graduate from high school.Therefore, increased high school graduation rate paired with no changes in enrollment/years in high school is consistent with increased achievement rather than the extensive margin variation in years of schooling.

4.2 Effects of the Policy on Adult Health Outcomes

In Section 4.1 we established that the policy lead to improvements in school quality and student achievement but it did not affect school quantity. We now move to one of our two main outcomes of interest – long-run health measured at ages 19 to 28 – where we compare students that attended improving vs. control schools. Table 3 presents these results for three types of outcomes: hospitalizations (columns 1 to 3), emergency department visits (column 4 and 5), and probability of dying (column 6). The final column combines all these outcomes into a single standardized index.

Irrespective of the outcome we find negative and statistically significant at conventional levels estimates. Considering the extensive margin students who attended improving schools had 16.0% lower likelihood of hospitalization, 5.0% lower likelihood of emergency department visit, and 25.0% lower likelihood of death prior to age 28. We find even larger improvements on the intensive margin with numbers of hospitalizations and ED visits declining by 19.3% and 15.9%, respectively. Combining all the outcomes together we estimate that long-run health outcomes of students in treated schools improved by 0.076 σ compared with students in control group schools.

Given that most of our outcomes are based on administrative hospitalization and ED visit data, it is natural to ask to what extent fewer events are due to improved health and to what extent it is due to lower utilization (perhaps due to inability to access the services or shifts to primary care). We address this concern in four ways. First, we also document declines in mortality as a results of the treatment and one would expect increases in mortality if declines in hospitalizations/ED visits were driven by avoidance or transition to less appropriate sources of care. Second, in Section 6 we document no changes in employment, income, or welfare participation. This means that our effects are unlikely to be driven by changes in ability to afford healthcare. Third, prior literature documents very low elasticities of demand for hospitalizations and ED visits implying that people are unlikely to underutilize these services even if financially constrained (Aron-Dine et al., 2013). Fourth, in Section 6

we also document improvements in health at ages 15 to 18 when our subjects are still in high school and for the most part covered either by their parents insurance or by Medicaid. Given this evidence we lean towards interpreting our findings as improvements in health, perhaps as results of better health in teenagehood, and as a result of this declines in utilization rather than underutilization or shifting of healthcare services.

Another question that arises relates to magnitudes. We address it in two ways. First, we compute expected savings for the state from lower hospitalizations and welfare gains due declines in mortality. Second, we compare our findings to the extant literature on the effects of extensive margin changes in education on health and mortality. The average charges in South Carolina in 2024 for adults age 18 to 28 were $34,467 per hospitalization (including patients admitted through emergency department) while the average charges per ED admission (excluding patients subsequently admitted to hospital) were $4,297 (South Carolina Revenue and Fiscal Affairs Office, 2024). Results in Table 3 imply 19.3% fewer hospitalizations and 15.9% fewer ED visits of students in the treatment group (N=27,784) which implies reductions in costs of about $51M and $152M, respectively. Thinking about patients we can also consider their average expenses when hospitalized or visiting emergency room. Given that these are almost impossible to get and vary widely by insurance and services providers (Cooper et al., 2019), we consider 2024 Medicare inpatient hospital ($1,632) and ED ($240) deductibles (CMS, 2024) as a lower bound. Repeating the same calculation we arrive at $11M of consumer savings for individuals in our treatment group between 2004 and 2019. Finally, considering mortality we can use value of statistical life (VSL) estimates to get some sense on the policy benefits with respect to mortality. Given that these estimates vary depending on the study we chose to use relatively more conservative estimate of $2.8M proposed by Ashenfelter and Greenstone (2004). During our time period 318 individuals died in the control group sample while the policy reduced the number of deaths by 25% or by 79 deaths which implies welfare gains of approximately $223M.[9]

[9]If we instead chose VSL of $10.3M proposed for the U.S. in a review article by Viscusi (2018), our estimates gains from reducing mortality would be even bigger at about $820M.

4.3 Effects of the Policy on Fertility

Our second set of main outcomes of interest relate to fertility. We present these results in Table 4 where we consider two extensive margin measures (column 1 and 2) and two intensive margin measures (columns 3 and 4). These outcomes are only available for women in our sample. Irrespective of the outcome we find negative estimates on fertility, however, estimate on teen birth is not statistically significant at conventional level. They imply almost 10% decline in likelihood of having any children and 12.4% decline in number of children. We also find reductions in age at first birth but this is an unconditional estimate which is driven by the negative extensive margin estimate. When we condition having at least one child, there is no difference in age at first birth between treated and control group individuals. Given that our population is on the younger side we view the intensive margin estimates as potentially less reliable despite the fact that we do not find any positive effects on age at first birth conditional on having at least one child which would be suggestive of shifting of births

Our findings stand in opposition to prior work by McCrary and Royer (2011) in the US and Beck et al. (2024) in Norway but are consistent with work from Germany (Cygan-Rehm and Maeder, 2013) and South Korea (Sohn and Lee, 2019). Compared to the latter two studies, however, we find much smaller effect sizes. Our extensive margin estimate is less than one-half of the estimates reported in those two studies.

5 Robustness

We presents several robustness checks that are specific to RDD designs as well as more directly to the research questions we are studying. The first set of these results is presented in Tables A3 and A4 for health and fertility outcomes, respectively. We do not present robustness estimates for teen birth since this coefficient did not reach statistical significance in the main analysis (and except of one robustness specification it did not reach statistical

significance for these estimation permutations either). In both tables outcomes are in rows while columns present alternative specifications. First column replicates the main results from Tables 3 and 4 to ease comparisons.

In the main specification we use global linear polynomial approach (Cattaneo et al., 2024) while in column 2 we present local linear estimates using bandwidth of 0.3 performance points (Lee and Lemieux, 2010) and in column 3 local randomization estimation (Cattaneo et al., 2015). In both cases sample sizes decreases meaningfully (over 50% in the latter approach) but nonetheless most estimates are of similar size and remain statistically significant at conventional levels. The only quantitatively meaningful departure is for probability of any ED visit in local linear specification where the estimates is not statistically significant at conventional levels and it is slightly more than a quarter of the preferred estimate. For the remaining coefficients 11 out of 18 remain statistically significant at least at 10% level and 14 out of 18 do not shrink in magnitude towards zero from the preferred estimate by more than a third (and two out of the four coefficients that move towards zero by more than a third nonetheless remain negative and statistically significant at least at 10% level). We also ensure that our local linear estimates are not driven by specific bandwidth that we chose. Figures A1 and A2 present local linear estimates where we expand the left side of the distribution from 0.3 performance index points to 1.3 points. As noted before we cannot expand the right side of the distribution beyond 0.3 points because this reaches threshold for the next policy cutoff. Except of the very small bandwidths of 0.3 to 0.4 points all the estimates are negative, stable, and mostly statistically significant. We thus view our findings as broadly robust to alternative estimation strategies.

Even though Tables 1 and A2 presented balance of covariates and predicted outcomes while tests performed for density of running variable distribution (Figure 1) suggested no discontinuity there is always a concern in RDD designs about manipulation of running variable close to the cutoff. In our application this concern could be heightened by the fact that Frandsen (2017) test p-value is on the lower end at 0.11. Thus, to address this potential issue

in column 4 we present donut RDD estimates where we exclude observations that are immediately below and above the discontinuity. The estimates are largely similar to our preferred specification – in fact in most cases they become more negative – and remain statistically significant at conventional level for 16 out of 18 outcomes.

Third, there nonetheless could be a concern about sorting of students based on their achievement or about independent effect of baseline ability on health and fertility outcomes beyond the policy intervention. Therefore, in column 5, for the subsample of students for whom we can observe grade 8 test scores we control for this input in regressions. The estimates barely move and remain statistically significant at conventional levels for 17 out of 18 outcomes. This exercise addresses balance in test scores, robustness to controlling for powerful determinant of outcomes we consider,and different sample (we only observe grade 8 test scores for cohorts 2001-02 to 2005-06). It could also be the case that schools to the left and to the right of the cutoff differ somewhat in their sociodemographic composition – especially that although insignificant our imbalances in Table 1 reached 6.2% of the control group mean – which could confound our findings. Thus, to be on the safe side in column 6 we control for school-level covariates: fraction of students who are female and Black, total school enrollment, attendance and grade retention rates. Our estimates remain virtually unchanged.

Finally, we test robustness of our results to expanding the end year of the sample from 2019 to 2022. In the main specification we excluded years 2020 to 2022 to avoid noise in the outcomes related to Covid-19 pandemic. This concern applies most directly to health outcomes but we now know that during the pandemic fertility trends likewise changed and did not necessarily reflect the steady state. Extending the sample represents a trade off. On one hand, we potentially introduce noise due to pandemic events but on the other hand we extend age horizon of individuals in our sample from 28 to 31 years. These results are presented in Column 7 and irrespective of outcome our estimates remain similar both in quantitative terms as well as when it comes to statistical significance.

Our final robustness exercise is a placebo check. Specifically we create "fake cutoffs" by adding 0.1 index points to the running variable in the range or -1.0 to 0.2 points and then re-estimating our equation of interest. Note that this placebo cutoffs could still pick up some treatment effects since many schools will nonetheless be correctly assigned if we move the cutoff by just 0.1 points either way. Figures A3 and A4 present these results. For all fertility outcomes as well as three out of six health outcomes our true estimates are clearly away from any of the placebo estimates. This is less true for number of hospitalizations, number of ED visits, and likelihood of dying although even in this case the true estimate is always one of the most negative estimates. We have verified that the placebo estimates close to the true estimates in these cases all come from small movements of the cutoff (of 0.1 to 0.2 points) and thus these are less reliable as true placebos. Once we exclude them even for these three outcomes our true estimates are clearly to the left of the placebo distribution.

6 Mechanisms

We have already established in Section 4.1 that the policy led to improvements in quality of treated schools compared to their unaffected counterparts. This has manifested by both improved student satisfaction from schooling (subjective measure) as well as academic (objective) outcomes but it did not lead to changes in years of schooling. Thus, our focus is on the effects of school quality rather than quantity and in this section we explore some of the possible consequences of such improvements that could moderate our long-run health and fertility effects.

We propose and test multiple mechanisms that can be broadly classified as leading to relaxed budget constraint and/or to behavioral changes. When it comes to the former we consider income and employment, utilization of welfare benefits, and special education classification when still in high school. For the latter set, we analyze health and criminal activity under the age of 18. All of these outcomes could plausibly be affected by changes in

school quality and lead to improved health and declines in fertility through either increased productivity or improved outside options (be it monetary or not).

We begin by presenting effects on employment and quarterly earnings at ages 23 to 28 in Table 7.[10] Recall that we study a disadvantaged population where only approximately 70% of students graduate on time and less than 7% are eligible for LIFE scholarship to attend college. Therefore, we believe that life-cycle bias in earnings is less likely to play an important role in our estimates (Bhuller et al., 2017) or at least it should be similar in treatment and control group. Nonetheless, to be on the safe size, we present results separately for ages 23-25 (Columns 1 and 3) and 26-28 (Columns 2 and 4).

We find inconsistently signed, never statistically significant at conventional levels, and quantitatively small effects on both employment and earnings. The former coefficients are negative while the latter are positive while for both outcomes effect sizes are small at -2.0 to -3.7% for employment and 0.1 to 0.2% for earnings. Based on these results we conclude that our findings are unlikely to be driven by long-run changes in labor market opportunities.

Lack of meaningful changes in employment and income is confirmed when considering welfare enrollment. In Table A6 we present RDD estimates for the probability of being enrolled in any welfare program (column 1), TANF (column 2), and SNAP (column 3) at ages 19 to 28. This information is based on administrative data from the South Carolina Department of Social Services. Irrespective of the outcome, we do not find any statistically significant effects while the relative magnitude of the coefficients does not exceed 1%. Using 95% confidence intervals and any welfare participation, we can rule out reductions larger than 5.1% and increases larger than 3.6% (relative to control group mean). Small, negative, and statistically insignificant effects on welfare participation are consistent with small, mixed singed, and statistically insignificant effects on employment and earnings.

[10]We have also investigated employment and earnings at ages 18 to 22. We do not find any treatment effects on this outcome. First, such result alligns with later ages evidence presented in Table 7. Second, it suggests that treated students are unlikely to differentially enroll in college compared to control group students. Should this be the case, we'd expect lower employment and earnings in the treated group but we do not find such result.

Two prior analyses considered additional resources available to treated vs. control group children in adulthood, plausibly as a result of improved educational outcomes documented in Section 4.1, however we can also consider if increased resources could be available to these individual in high school. We proxy for this using special education classification, which in South Carolina triggers additional resources to schools. Specifically, the state provides $2.6 for every student with an Individual Education Plan (IEP) per $1 spend on a student without an IEP. It is thus plausible that schools under accountability pressure will designate more students as special education and use at least some of these resources to improve student outcomes (Jacob, 2005). Note that fraction of students classified as "learning disabled" does not affect student testing or school designation under South Carolina accountability ratings.

Table A5 presents these results for grades 9 to 12 separately (Columns 1 to 4) as well as for any high school grade (Column 5). Note that school designation happens after start of the academic year, and thus results in the first grade of high school (Column 1) could be treated as another balance check. Indeed, in grade 9, we do not see any differential labeling of students in treated vs. control group schools. In contrast for subsequent grades we observe statistically significant and quantitatively meaningful reduction in number of students designated as "learning disabled". This is inconsistent with the story of schools seeking additional resources for their students via "disability labeling", but it aligns with our improved human capital and learning evidence presented in Section 4.1. Therefore, we conclude that learning improvements are the likely cause of decline in share of students with learning disability.

Having established that primary drivers of our long-run findings are unlikely to be due to relaxed budget constraint, we now consider behavioral adjustments as another explanation. Prior work by Eren et al. (2023) using the same source variation as we use in this paper documented declines in crime in adulthood. Here we attempt to understand if these effects manifest already in high school and could thus lead to improved health in both short- and long-run. Table 6 presents these results for any arrests at ages 15-18 (column 1) as well

as category-specific arrests (column 2 to 5). We find negative and sizable reductions in all forms of crime that are particularly meaningful for drugs and alcohol arrests. Overall, as a result of the policy juvenile criminal activity declines by almost 30% compared to control group mean. We are not aware of any prior work that would document effects of this type of policy on juvenile crime, but it is reasonable to assume that behavioral adjustments leading to reductions in arrests – particularly those related to substance abuse – could translate into better health short-run and ultimately in the long-run.

Following this train of thought we also investigate short-run health effects. Here we replicate the results from Table 3, excluding mortality which is extremely rare event among teens, but for ages 15 to 18 when most individuals are still in school. Table 5 presents these results. We find pattern of results that is largely consistent with long-run health estimates. All coefficients are negative but four out of six are estimated more noisily, perhaps because incidence of the health outcomes we consider is 2 to 9 times smaller in teenagehood than in adulthood. The effect sizes are somewhat smaller for hospitalizations, and particularly so for number of days hospitalized, but are modestly larger for emergency department visits. This is consistent with the type of healthcare received in teenagehood compared to adulthood. Furthermore, fewer risky behaviors related to criminal activity documented in Table 6 could also lead to fewer ED visits. When we combine all the outcomes into a single health utilization index (Column 6) we find negative and statistically significant reductions in utilization of health services which is consistent with improvements in short-run health. The coefficient of -0.075 σ is very similar to -0.076 σ estimated in adulthood (Column 7 of Table 3). This means that a non-trivial share of long-run health gains can likely be attributed to short-run health gains with the caveat that hospitalizations effects are larger in adulthood while ED visits effects are larger in teenagehood.

Overall, we conclude that the likely mechanisms behind our long-run findings include direct improvements in cognitive ability and behavior as a result of the policy that translate into, or act in addition to, fewer risky behaviors and improved short-run health. On the

other hand, we can rule out major role of changes in extensive margin of education (e.g., years of schooling or college going), labor market effects operating through either extensive (employment) or intensive (earnings) margins, and increased utilization of welfare services.

7 Heterogeneity

In this section we present heterogeneity analyses along sociodemographic and baseline ability dimensions. These are important in order to understand distributional effects of the policy and to ensure that our positive average gains do not represent a combination of very positive effects for one set of individuals paired with negative effects for another set of individuals.

We being with sociodemographic splits presented in Tables 8 and 9 for health and fertility outcomes, respectively. In Table 8, columns 1 and 2 split the sample by sex, columns 3 and 4 by race, columns 5 and 6 by socioeconomic status in grade 9, and columns 7 and 8 by the median of grade 8 test scores (our proxy for baseline ability). Since fertility outcomes are limited to females in our sample, columns 1 and 2 represent split by race, columns 3 and 4 represent split by socioeconomic status, and columns 5 and 6 respect split by baseline ability in Table 9.

We find relatively similar effects for males and females except perhaps for total days hospitalized and mortality which appear to be meaningfully larger for males. Nonetheless given relatively large 95% confidence intervals we cannot rule out that these estimates are statistically identical. Splitting the sample by race presents a more nuanced story. Effects appear to be very similar for probability of any hospitalization and mortality but they are meaningfully larger for Black students when it comes to intensive margin of hospitalizations and either ED utilization measure. In a stark contrast for fertility outcomes we observe larger reductions in White compared to Black fertility.

When it comes to socioeconomic status – proxied by free and reduced price lunch eligibility in grade 9 – except for mortality where the percentage point estimates are similar we

find meaningfully larger estimates for those who are in deeper poverty. Specifically, although still negative there is not a single statistically significant estimate for students who were not eligible for free and reduced price lunch. Lining up closely with the racial divide, however, we find larger reduction in fertility for students less likely to be in poverty.

Although sample of individuals for whom we can observe baseline test scores is smaller than what we use in the preferred specification, it is important to investigate effects by baseline ability. This is because prior research suggests that teachers and schools might be particularly saavy in terms of targeting specific students in response to strong incentives. However, we note that in our population students who have lower test scores are also more likely to be Black and low SES. Thus, consistent with the aforementioned results, we find larger reductions in hospitalizations and mortality for those with below-median baseline test scores. On the other hand, the effects for emergency department visits are somewhat larger for those with above-median baseline test scores. When it comes to fertility outcomes the findings are much more clear cut: students with above-median baseline ability have larger reductions in fertility. This last result is consistent with larger fertility effects for White students and those not on free or reduced price lunch.

Overall, the heterogeneity analysis presents interesting divergent findings. Long-run health gains are concentrated among individuals with markers of disadvantage, irrespective of how exactly we measure it. On the other hand, fertility reductions are concentrated among the more advantaged students.

8 Intergenerational Effects

The final analysis in the paper pertains to intergenerational effects. Namely we investigate health at birth of children born to adults who attended either the improving schools (treatment group) or the status quo schools (control group). Unlike work by McCrary and Royer (2011) who do not find any fertility effects of compulsory schooling laws, our analysis is

complicated by the fact that as a result of the policy females in the treatment group are less likely to have children compared to females educated in the control group schools. With that in mind, Table 10 presents the results for two indicators: low birth weight (columns 1 to 3) and prematurity (columns 4 to 6). We present effects for both the subset of women who give births in our sample (columns 1 and 4) as well as estimates a'la Lee bounds (Lee, 2009) where we impute extreme positive (columns 2 and 5) or extreme negative (columns 3 and 6) outcomes for those women for whom we do not observe births in the sample.

Our findings in columns 1 and 4 imply no statistically significant effects on health at birth of the next generation. Although the coefficients are positive, they are small in percentage points terms or even considering the relatively low means in terms of effect sizes. Average rates of low birth weight and prematurity incidence in our sample are 10.4% and 6.4%, respectively. Both numbers correspond closely with what is reported for the full population of births in South Carolina. Given that our population is disadvantaged this means that relatively better off individuals are more likely to have births in our sample and thus our estimates could be downward biased.

We try to address this issue in the remaining columns of the table where we impute either the negative or the positive outcome for all mother that do not give birth in our sample. This is an extreme exercise which should be obvious given the changes in control group means that become unbelievably low (when we assume the most positive outcome) or unbelievably high (when we assume the most negative outcome). Nonetheless, the lower bound coefficients remain relatively small and not statistically significant while the upper bound coefficients are positive and statistically significant. We caution readers about putting too much emphasis on the estimated increases in low birth weight and prematurity as these coefficients are based on a sample with implausibly large baseline means.

9 Conclusions

There is an extensive and inconclusive literature on the effects of education on health and fertility. Almost all of the causal research on this topic uses variation in extensive margin of schooling e.g., years of schooling or access to particular educational level. In this paper we propose that extensive margin variation might not be most relevant given groundbreaking theoretical work by Grossman (1972) and Becker (1960). Rather it could be the intensive margin of schooling – its quality and associated learning – that should be of central interest to economists and policy makers. This dimension appears particularly relevant from policy perspective given that it's unlikely that many countries could massively scale up their supply of educational institutions or increase mandatory schooling age.

We present evidence that school quality affects both long-run health and fertility. Using exogenous variation in accountability pressure and a regression discontinuity design, we compare long-run outcomes of students from schools that were incentivized to improve vs. schools that did not face such pressures. We first document that both subjective and objective measures of school quality improved in the former compared with the latter group of schools. We then document that students educated in improving schools have lower likelihood of hospitalization, emergency department visit, and death at ages 18 to 28. They are also less likely to have children. On the other hand, we do not observe differential birth outcomes for the children of treated vs. control group students. Candidate mechanisms include improvements in cognitive and behavioral skills, reductions in risky behaviors, and better health in high school. At the same time, we can rule out meaningful changes in educational attainment, employment, income, and welfare utilization.

Our findings present a policy conundrum. On one hand, it is clear that motivating failing schools to improve could work and generate positive health externalities beyond direct educational gains. Thus, reductions in school accountability or other mechanisms forcing school to improve their quality – especially at the bottom of the distribution – might have been premature. On the other hand, our fertility findings suggest that better schools could

deepen the reproductive crisis that many developed countries face with their fertility rates well below replacement.

References

Aaronson, Daniel, Bhashkar Mazumder, Seth G. Sanders, and Evan J. Taylor (2021) "Estimating the Effect of School Quality on Mortality in the Presence of Migration: Evidence from the Jim Crow South," *Journal of Labor Economics*, Vol. 39, No. 2, pp. 527–558.

Abel, Eirik Berger, Aline Butikofer, and Kjell G Salvanes (2024) "Fertility, Partner Choice, and Human Capital."

Acemoglu, Daron and Simon Johnson (2007) "Disease and Development: The Effect of Life Expectancy on Economic Growth," *Journal of Political Economy*, Vol. 115, No. 6, pp. 925–985.

Agasisti, Tommaso and Jose Cordero-Ferrera (2013) "Educational Disparities Across Regions: A Multilevel Analysis for Italy and Spain," *Journal of Policy Modeling*, Vol. 35, No. 6, pp. 1079–1102.

Amin, Vikesh, Jere R. Behrman, and Tim D. Spector (2013) "Does More Schooling Improve Health Outcomes and Health-Related Behaviors? Evidence from U.K. Twins," *Economics of Education Review*, Vol. 35, pp. 134–148.

Arendt, Jacob Nielsen, Mads Lybech Christensen, and Anders Hjorth-Trolle (2021) "Maternal education and child health: Causal evidence from Denmark," *Journal of health economics*, Vol. 80, p. 102552.

Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein (2013) "The RAND health insurance experiment, three decades later," *Journal of Economic Perspectives*, Vol. 27, No. 1, pp. 197–222.

Asadullah, Niaz and Gaston Yalonetzky (2012) "Inequality of Educational Opportunity in India: Changes Over Time and Across States," *World Development*, Vol. 40, No. 6, pp. 1151–1163.

Ashenfelter, Orley and Michael Greenstone (2004) "Using mandated speed limits to measure the value of a statistical life," *Journal of political Economy*, Vol. 112, No. S1, pp. S226–S267.

Barcellos, Silvia H., Leandro S. Carvalho, and Patrick Turley (2023) "Distributional Effects of Education on Health," *Journal of Human Resources*, Vol. 58, No. 4, pp. 1273–1306.

Barlow, Pepita (2021) "The effect of schooling on women's overweight and obesity: A natural experiment in Nigeria," *Demography*, Vol. 58, No. 2, pp. 685–710.

Basu, Anirban, Andrew M. Jones, and Pedro Rosa Dias (2018) "Heterogeneity in the impact of type of schooling on adult health and lifestyle," *Journal of Health Economics*, Vol. 57, pp. 1–14.

Beck, Kathryn Christine, Rannveig Kaldager Hart, and Martin Flatø (2024) "School Starting Age, Fertility, and Family Formation: Evidence From the School Entry Cutoff Using Exact Date of Birth," *Demography*, Vol. 61, No. 6, pp. 1999–2026.

Becker, Gary (1960) "An Economic Analysis of Fertility," in *Demographic and Economic Change in Developed Countries*, pp. 209–231.

Bharadwaj, Prashant, Mallesh M. Pai, and Agne Suziedelyte (2017) "Mental Health Stigma," *Economics Letters*, Vol. 159, pp. 57–60.

Bhuller, Manudeep, Magne Mogstad, and Kjell G Salvanes (2017) "Life-cycle earnings, education premiums, and internal rates of return," *Journal of Labor Economics*, Vol. 35, No. 4, pp. 993–1030.

Black, Sandra E. (1999) "Do Better Schools Matter? Parental Valuation of Elementary Education," *The Quarterly Journal of Economics*, Vol. 114, No. 2, pp. 577–599.

Bokhari, Farasat A. S. and Helen Schneider (2011) "School accountability laws and the consumption of psychostimulants," *Journal of Health Economics*, Vol. 30, No. 2, pp. 355–372.

Bond, Timothy N. and Kevin Lang (2019) "The Sad Truth about Happiness Scales," *Journal of Political Economy*, Vol. 127, No. 4, pp. 1629–1640.

Bradley, Steve, Mirko Draca, and Colin Green (2004) "School Performance in Australia: Is There a Role for Quasi-Markets?" *Australian Economic Review*, Vol. 37, No. 3, pp. 271–286.

Brunello, Giorgio, Daniele Fabbri, and Margherita Fort (2013) "The Causal Effect of Education on Body Mass: Evidence from Europe," *Journal of Labor Economics*, Vol. 31, No. 1, pp. 195–223.

Bütikofer, Aline, Rita Ginja, Katrine L, and Fanny Landaud (2023) "Higher-Achievement Schools, Peers and Mental Health," *The Economic Journal*, Vol. 133, No. 655, pp. 2580–2613.

Canaan, Serena (2024) "Does Reducing Early School Tracking Affect Health Behaviors?" *American Economic Association Papers and Proceedings*, Vol. 114, pp. 375–380.

Card, David and Alan Krueger (1992) "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, Vol. 100, No. 1, pp. 1–40.

Cattaneo, Matias D, Brigham R Frandsen, and Rocio Titiunik (2015) "Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate," *Journal of Causal Inference*, Vol. 3, No. 1, pp. 1–24.

Cattaneo, Matias D, Nicolás Idrobo, and Rocío Titiunik (2019) *A practical introduction to regression discontinuity designs: Foundations*: Cambridge University Press.

Cattaneo, Matias D., Nicolas Idrobo, and Rocío Titiunik (2024) *A Practical Introduction to Regression Discontinuity Designs: Extensions*, Vol. 2398 of Quantitative and Computational Methods for the Social Sciences, Cambridge, UK: Cambridge University Press.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma (2018) "Manipulation Testing Based on Density Discontinuity," *Stata Journal*, Vol. 18, No. 1, pp. 234–261.

Chou, Shin-Yi, Jin-Tan Liu, Michael Grossman, and Ted Joyce (2010) "Parental education and child health: evidence from a natural experiment in Taiwan," *American Economic Journal: Applied Economics*, Vol. 2, No. 1, pp. 33–61.

Clark, Damon and Emilia Del Bono (2016) "The long-run effects of attending an elite school: Evidence from the United Kingdom," *American Economic Journal: Applied Economics*, Vol. 8, No. 1, pp. 150–176.

Clark, Damon and Heather Royer (2013) "The Effect of Education on Adult Mortality and Health: Evidence from Britain," *American Economic Review*, Vol. 103, No. 6, pp. 2087–2120.

Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen (2019) "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured," *Quarterly Journal of Economics*, Vol. 134, No. 1, pp. 51–107.

Cuevas Ruiz, Pilar, Cristina Borra, and Almudena Sevilla (2023) "The causal impact of maternal educational curricula on infant health at birth."

Currie, Janet and Enrico Moretti (2003) "Mother's education and the intergenerational transmission of human capital: Evidence from college openings," *The Quarterly journal of economics*, Vol. 118, No. 4, pp. 1495–1532.

Cygan-Rehm, Kamila and Miriam Maeder (2013) "The effect of education on fertility: Evidence from a compulsory schooling reform," *Labour Economics*, Vol. 25, pp. 35–48.

Deming, David J., Sarah Cohodes, Jennifer Jennings, and Christopher Jencks (2016) "School Accountability, Postsecondary Attainment, and Earnings," *Review of Economics and Statistics*, Vol. 98, No. 5, pp. 848–862.

Ding, Xuejie, Nicola Barban, and Melinda C. Mills (2019) "Educational Attainment and Allostatic Load in Later Life: Evidence Using Genetic Markers," *Preventive Medicine*, Vol. 129, p. 105866.

Dizon-Ross, Rebecca (2020) "How Does School Accountability Affect Teachers? Evidence from New York City," *Journal of Human Resources*, Vol. 55, No. 1, pp. 76–118.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo (2018) "The Economic Consequences of Hospital Admissions," *American Economic Review*, Vol. 108, No. 2, pp. 308–352.

Dursun, Bahadır, Resul Cesur, and Inas R Kelly (2022) "Mandatory Schooling of Girls Improved Their Children's Health: Evidence from Turkey's 1997 Education Reform," *Journal of Policy Analysis and Management*, Vol. 41, No. 3, pp. 824–858.

Dursun, Bahadir, Ozkan Eren, and My Nguyen (2024) "Curriculum reforms and infant health," *Review of Economics and Statistics*, Vol. 106, No. 2, pp. 394–408.

Eren, Ozkan, David N. Figlio, Naci H. Mocan, and Orgul D. Ozturk (2023) "School Accountability, Long-Run Criminal Activity, and Self-Sufficiency," NBER Working Paper 31556, National Bureau of Economic Research.

Fahle, Erin and Sean Reardon (2018) "How Much Do Test Scores Vary Among School Districts? New Estimates Using Population Data, 2009-2015," *Educational Researcher*, Vol. 47, No. 4, pp. 221–234.

Figlio, David and Krzysztof Karbownik (2017) "Some Schools Much Better than Others at Closing Achievmeent Gaps Between Their Advantaged and Disadvantaged Students," *Economic Studies at Brookings: Evidence Speaks Reports*, Vol. 2, No. 19, pp. 1–8.

Figlio, David N. (2006) "Testing, Crime and Punishment," *Journal of Public Economics*, Vol. 90, No. 4-5, pp. 837–851.

Figlio, David N. and Susanna Loeb (2011) "School Accountability," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann eds. *Handbook of the Economics of Education*, Vol. 3, pp. 383–423.

Figlio, David N. and Maurice E. Lucas (2004) "What's in a Grade? School Report Cards and the Housing Market," *American Economic Review*, Vol. 94, No. 3, pp. 591–604.

Fletcher, Jason and Hamid Noghanibehambari (2024) "The Effects of Education on Mortality: Evidence Using College Expansions," *Health Economics*, Vol. 33, No. 3, pp. 541–575.

Fort, Margherita, Nicole Schneeweis, and Rudolf Winter-Ebmer (2016) "Is education always reducing fertility? Evidence from compulsory schooling reforms," *The Economic Journal*, Vol. 126, No. 595, pp. 1823–1855.

Frandsen, Brigham R. (2017) "Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete," in Matias D. Cattaneo and Javier C. Escanciano eds. *Regression Discontinuity Designs (Advances in Econometrics, Vol. 38)*, Leeds, UK: Emerald Publishing Limited, pp. 281–315.

Galama, Titus J., Adriana Lleras-Muney, and Hans van Kippersluis (2018) "The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence," in Jonathan Hamilton and Andrew Jones eds. *Oxford Research Encyclopedia of Economics and Finance*, USA: Oxford University Press.

Gensowski, Miriam and Mette Gørtz (2024) "The Education–Health Gradient: Revisiting the Role of Socio-Emotional Skills," *Journal of Health Economics*, Vol. 97, p. 102911.

González, Felipe, Luis R. Martínez, Pablo Muñoz, and Mounu Prem (2024) "Higher Education and Mortality: Legacies of an Authoritarian College Contraction," *Journal of the European Economic Association*, Vol. 22, No. 4, pp. 1762–1797.

Grimard, Franque and Daniel Parent (2007) "Education and Smoking: Were Vietnam War Draft Avoiders Also More Likely to Avoid Smoking?" *Journal of Health Economics*, Vol. 26, No. 5, pp. 896–926.

Grossman, Michael (1972) "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, Vol. 80, No. 2, pp. 223–255.

Grytten, Jostein, Irene Skau, and Rune J Sørensen (2014) "Educated mothers, healthy infants. The impact of a school reform on the birth weight of Norwegian infants 1967–2005," *Social Science & Medicine*, Vol. 105, pp. 84–92.

Hanushek, Eric A. and Margaret E. Raymond (2004) "The Effect of School Accountability Systems on the Level and Distribution of Student Achievement," *Journal of the European Economic Association*, Vol. 2, No. 2-3, pp. 406–415.

——— (2005) "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management*, Vol. 24, No. 2, pp. 297–327.

Howe, Laurence J., Humaira Rasheed, Paul R. Jones, Dorret I. Boomsma, David M. Evans, Alexandros Giannelis, Caroline Hayward, John L. Hopper, Amanda Hughes, Hannu Lahtinen, Shuai Li, Penelope A. Lind, Nicholas G. Martin, Pekka Martikainen, Sarah E. Medland, Tim T. Morris, Michel G. Nivard, Jean-Baptiste Pingault, Karri Silventoinen, Jennifer A. Smith, Emily A. Willoughby, James F. Wilson, Bjørn Olav Åsvold, Øyvind E. Næss, George Davey Smith, Jaakko Kaprio, Ben Brumpton, Neil M. Davies, and Within Family Consortium (2023) "Educational Attainment, Health Outcomes and Mortality: A Within-Sibship Mendelian Randomization Study," *International Journal of Epidemiology*, Vol. 52, No. 5, pp. 1579–1591.

Jacob, Brian A (2005) "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools," *Journal of public Economics*, Vol. 89, No. 5-6, pp. 761–796.

Jang, Heewon and Sean Reardon (2019) "Stests as Sites of Educational (In)equality: State Contexts and the Socioeconomic Achievement Gradient," *AERA Open*, Vol. 5, No. 3, p. 2332858419872459.

Jessica, Drescher and Benjamin Domingue (2023) "The Distribution of Child Physicians and Early Academic Achievement," *Health Services Research*, Vol. 58, pp. 165–174.

Jiang, Wei, Yi Lu, and Huihua Xie (2020) "Education and Mental Health: Evidence and Mechanisms," *Journal of Economic Behavior & Organization*, Vol. 180, pp. 407–437.

Kebede, Endale, Anne Goujon, and Wolfgang Lutz (2019) "Stalls in Africa's Fertility Decline Partly Result from Disruptions in Female Education," *Proceedings of the National Academy of Sciences*, Vol. 116, No. 8, pp. 2891–2896.

King, Marissa D., Jennifer Jennings, and Jason M. Fletcher (2014) "Medical Adaptation to Academic Pressure: Schooling, Stimulant Use, and Socioeconomic Status," *American Sociological Review*, Vol. 79, No. 5, pp. 831–856.

Koebe, Josefine and Jan Marcus (2022) "The length of schooling and the timing of family formation," *CESifo Economic Studies*, Vol. 68, No. 1, pp. 1–45.

Lafortune, Julien and David Schönholzer (2022) "The Impact of School Facility Investments on Students and Homeowners: Evidence from Los Angeles," *American Economic Journal: Applied Economics*, Vol. 14, No. 3, pp. 254–289.

Le, Kien and My Nguyen (2020) "Shedding Light on Maternal Education and Child Health in Developing Countries," *World Development*, Vol. 133, p. 105005.

Lee, David S (2009) "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, Vol. 76, No. 3, pp. 1071–1102.

Lee, David S and Thomas Lemieux (2010) "Regression discontinuity designs in economics," *Journal of economic literature*, Vol. 48, No. 2, pp. 281–355.

Lehtimäki, Terho, Jaana Viinikainen, Alex Bryson, Petri Böckerman, Marko Elovainio, Nina Pitkänen, and Laura Pulkki-Råback (2022) "Does Better Education Mitigate Risky Health Behavior? A Mendelian Randomization Study," *Economics and Human Biology*, Vol. 44, p. 101134.

Leuven, Edwin, Erik Plug, and Marte Rønning (2016) "Education and Cancer Risk," *Labour Economics*, Vol. 43, pp. 106–121.

Li, Hongbin and Junsen Zhang (2007) "Do High Birth Rates Hamper Economic Growth?" *Review of Economics and Statistics*, Vol. 89, No. 1, pp. 110–117.

Lindeboom, Maarten, Ana Llena-Nozal, and Bas van Der Klaauw (2009) "Parental education and child health: Evidence from a schooling reform," *Journal of health Economics*, Vol. 28, No. 1, pp. 109–131.

Link, Bruce G. and Jo C. Phelan (1995) "Social Conditions as Fundamental Causes of Disease," *Journal of Health and Social Behavior*, Vol. 35, No. Extra Issue, pp. 80–94.

Lleras-Muney, Adriana (2005) "The Relationships Between Education and Adult Mortality in the United States," *Review of Economic Studies*, Vol. 72, No. 1, pp. 189–221.

——— (2022) "Education and Income Gradients in Longevity: The Role of Policy," *Canadian Journal of Economics*, Vol. 55, No. 1, pp. 5–37.

Lleras-Muney, Adriana, Joseph Price, and Dahai Yue (2022) "The Association Between Educational Attainment and Longevity Using Individual-Level Data from the 1940 Census," *Journal of Health Economics*, Vol. 84, p. 102534.

Lucas, Adrienne and Isaac Mbiti (2014) "Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya," *American Economic Journal: Applied Economics*, Vol. 6, No. 3, pp. 234–263.

Lundborg, Petter, Carl Hampus Lyttkens, and Paul Nystedt (2016) "The Effect of Schooling on Mortality: New Evidence From 50,000 Swedish Twins," *Demography*, Vol. 53, No. 4, pp. 1135–1168.

Malamud, Ofer, Andreea Mitrut, and Cristian Pop-Eleches (2023) "The Effect of Education on Mortality and Health: Evidence from a Schooling Expansion in Romania," *Journal of Human Resources*, Vol. 58, No. 2, pp. 561–592.

Marcus, Jan, Simon Reif, Amélie C. Wuppermann, and Amélie Rouche (2020) "Increased Instruction Time and Stress-Related Health Problems among School Children," *Journal of Health Economics*, Vol. 70, p. 102256.

Marmot, Michael (2015) *The Health Gap: The Challenge of an Unequal World*, London and New York: Bloomsbury Publishing.

Mazumder, Bhashkar, Maria Fernanda Rosales-Rueda, and Margaret Triyana (2023) "Social Interventions, Health, and Well-Being: The Long-Term and Intergenerational Effects of a School Construction Program," *Journal of Human Resources*, Vol. 58, No. 4, pp. 1097–1140.

McCrary, Justin and Heather Royer (2011) "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth," *American Economic Review*, Vol. 101, No. 1, pp. 158–195.

McElroy, Katherine (2023) "Does Test-Based Accountability Improve More Than Just Test Scores?" *Economics of Education Review*, Vol. 94, No. C, p. 102381.

Meara, Ellen, Seth Richards, and David Cutler (2008) "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, by Education, 1981-2000," *Health Affairs*, Vol. 27, No. 2, pp. 350–360.

Meng, Juanjuan, Hui Wang, Yu Yang, and Mingshan Zhang (2025) "Education competition and fertility intention: Evidence from China's private tutoring ban," *The Economic Journal*, p. ueaf096.

Mittleman, Joel and Jennifer L. Jennings (2018) "Accountability, Achievement, and Inequality in American Public Schools: A Review of the Literature," in Kate E. Pickett and Richard Wilkinson eds. *Handbooks of Sociology and Social Research*: Springer Nature, pp. 475–492.

Palme, Mårten and Emilia Simeonova (2015) "Does Women's Education Affect Breast Cancer Risk and Survival? Evidence from a Population-Based Social Experiment in Education," *Journal of Health Economics*, Vol. 42, No. C, pp. 115–124.

Papanicolas, Irene, Liana Woskie, and Ashish Jha (2018) "Health Care Spending in the United States and Other High-Income Countries," *JAMA*, Vol. 319, No. 10, pp. 1024–1039.

Rauscher, Emily, Greer Mellon, and Susanna Loeb (2025) "Priceless Benefits: Effects of School Spending on Child Mortality."

Reardon, Sean, Ericka Weathers, Erin Fahle, Heewon Jang, and Demetra Kalogrides (2024) "Is Separate Still Unequal? New Evidence on School Segregation and Racial Academic Achievement Gaps," *American Sociological Review*, Vol. 89, No. 6, pp. 971–1010.

Reback, Randall, Jonah Rockoff, and Heather L. Schwartz (2014) "Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind," *American Economic Journal: Economic Policy*, Vol. 6, No. 3, pp. 207–241.

Rockoff, Jonah and Lesley J. Turner (2010) "Short-Run Impacts of Accountability on School Quality," *American Economic Journal: Economic Policy*, Vol. 2, No. 4, pp. 119–147.

Ross, Catherine E. and Chien-Lin Wu (1995) "The Links Between Education and Health," *American Sociological Review*, Vol. 60, No. 5, pp. 719–745.

Rouse, Cecilia E., Jane Hannaway, Dan Goldhaber, and David Figlio (2013) "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure," *American Economic Journal: Economic Policy*, Vol. 5, No. 2, pp. 251–281.

Silva, JMC Santos and Silvana Tenreyro (2011) "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator," *Economics Letters*, Vol. 112, No. 2, pp. 220–222.

Sohn, Hosung and Suk-Won Lee (2019) "Causal Impact of Health a College Degree on Women's Fertility: Evidence from Regression Kin Designs," *Demography*, Vol. 56, No. 3, pp. 969–990.

South Carolina Revenue and Fiscal Affairs Office (2024) "Healthcare Data and Research," `https://rfa.sc.gov/data-research/healthcare`, Accessed: February 5, 2026.

Vasireddy, Sindhu, Ann Berrington, Bernice Kuang, and Hill Kulu (2023) "Education and Fertility: A Review of Recent Research in Europe," *Comparative Population Studies*, Vol. 48.

Virtanen, Hanna, Tiina Kuuppelomaki, Mikko Silliman, and Kristiina Huttunen (2024) "Education, Gender, and Family Formation," IZA Discussion Paper 17122, IZA.

Viscusi, W Kip (2018) "Best estimate selection bias in the value of a statistical life," *Journal of Benefit-Cost Analysis*, Vol. 9, No. 2, pp. 205–246.

de Walque, Damien (2007) "Does Education Affect Smoking Behaviors? Evidence Using the Vietnam Draft as an Instrument for College Education," *Journal of Health Economics*, Vol. 26, No. 5, pp. 877–895.

Wodtke, Geoffrey, Ugur Yildirim, David Harding, and Felix Elwert (2023) "Are Neighborhood Effects Explained by Differences in School Quality," *American Journal of Sociology*, Vol. 128, No. 5, pp. 1472–1528.

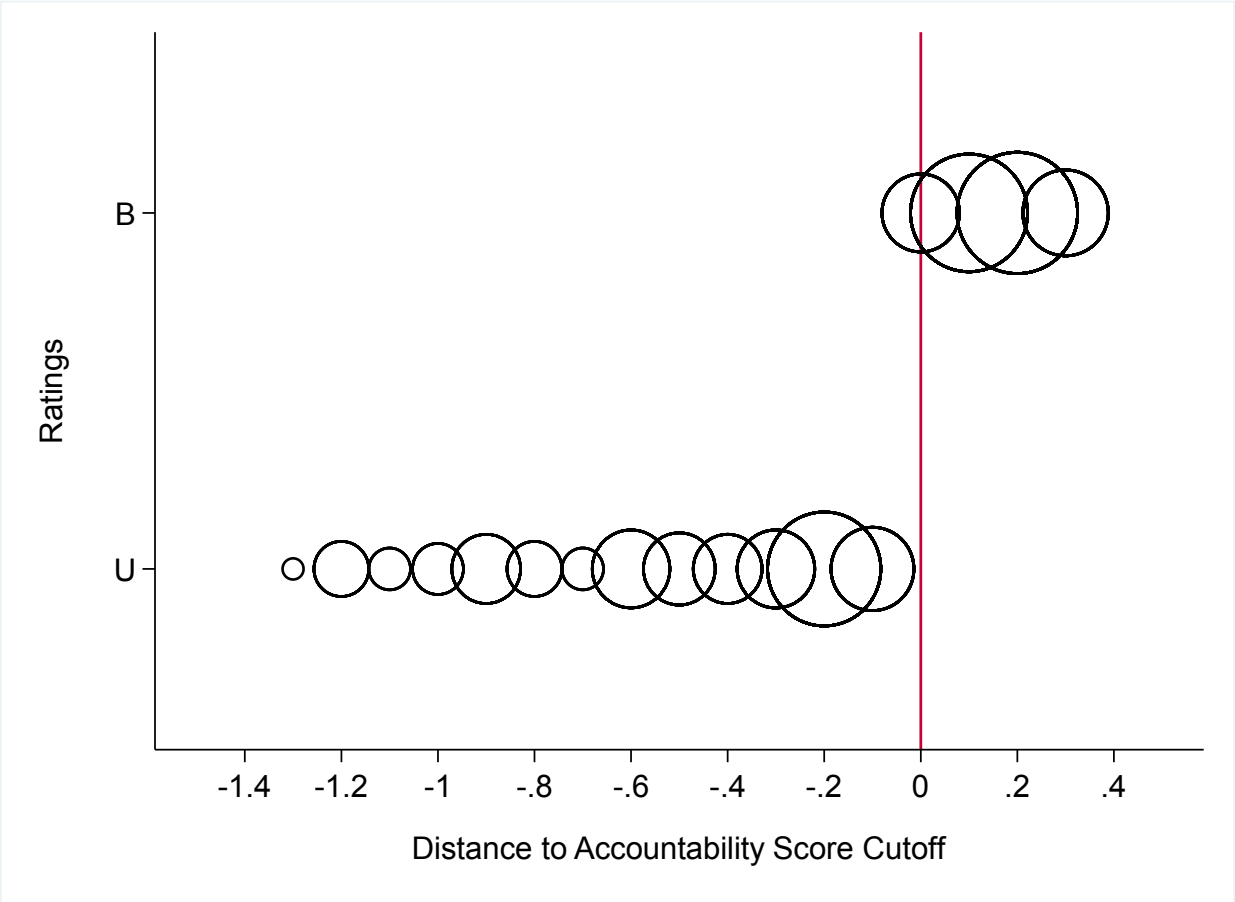
Xue, Xindong, Mingmei Cheng, and Wangyongxin Zhang (2021) "Does Education Really Improve Health? A Meta-Analysis," *Journal of Economic Surveys*, Vol. 35, No. 1, pp. 71–105.

Zajacova, Anna and Elizabeth M. Lawrence (2018) "The Relationship between Education and Health: Reducing Disparities Through a Contextual Approach," *Annual Review of Public Health*, Vol. 39, pp. 273–289.

Zang, Emma, Chloe Sariego, and Anirudh Krishnan (2022) "The Interplay of Race/Ethnicity and Education in Fertility Patterns," *Population Studies*, Vol. 76, No. 3, pp. 363–385.

10 Figures

Figure 1: Distribution of Schools



Notes: The figure displays the distribution of scores relative to the school accountability cutoff for schools rated as Unsatisfactory and Approaching Basic over the sample period. The number of schools at each point is proportional to the size of the bubbles. The vertical line denotes the accountability cutoff centered at 0.

11 Tables

Table 1: Regression Discontinuity and Local Randomization Balancing Tests

	(1)	(2)	(3)	(4)	(5)
	Female	White	Free or reduced price lunch	Baseline test scores	
				Math	Reading
			Panel A: Local linear regression		
Receipt of lower rating	0.003	0.015	0.039	0.013	0.015
	(0.007)	(0.050)	(0.035)	(0.044)	(0.044)
Control group mean	0.472	0.321	0.629	-0.402	-0.379
Implied imbalance (%)	0.6	4.7	6.2	N/A	N/A
Sample Size	54,336	54,336	54,336	35,322	35,322
			Panel B: Local Randomization		
Receipt of lower rating	0.010	-0.007	0.040	-0.015	-0.018
	(0.007)	(0.045)	(0.038)	(0.045)	(0.040)
Control group mean	0.476	0.292	0.653	-0.426	-0.401
Implied imbalance (%)	2.1	-2.4	6.1	N/A	N/A
Sample Size	22,609	22,609	22,609	14,382	14,382

Note: All specifications in Panel A use a linear polynomial with a uniform kernel. The analysis sample in Panel B is restricted to the interval of the running variable that contains only the index points, two on each side, that are immediately consecutive to the cutoff value. Panel B reports the point estimates from regressions of observable student characteristics on an indicator for the receipt of a lower accountability rating inside the interval. Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 2: Effects of the Policy on School Academics and Satisfaction

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. School-level analysis					
		Rate of students		Standardized index		
	in G10 passing exit exams	in G10 eligible for LIFE scholarship	graduating HS in 4 years	Academic	G10 student satisfaction	Character development score
Receipt of lower rating	2.985**	2.265	4.942*	0.311**	0.421**	0.547**
	(1.407)	(1.762)	(2.663)	(0.150)	(0.199)	(0.266)
Control group mean	61.69	6.43	71.10	0.311	0.018	-0.123
Implied effect size (%)	4.8	35.2	7.0	N/A	N/A	N/A
Sample Size	226	122	220	226	175	137

	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Panel B. Individual-level analysis							
	G10 exit exam proficient	School persistence: Enrollend in			Years in high school	School changes: Same school in		
		10th grade	11th grade	12th grade		10th grade	11th grade	12th grade
Receipt of lower rating	0.054*	-0.029	0.003	-0.016	-0.039	-0.005	-0.012	-0.010
	(0.028)	(0.024)	(0.026)	(0.027)	(0.046)	(0.016)	(0.017)	(0.018)
Control group mean	0.493	0.723	0.593	0.591	2.310	0.931	0.932	0.927
Implied effect size (%)	11.0	-4.0	0.5	-2.7	-1.7	-0.5	-1.3	-1.1
Sample Size	13,824	54,336	54,336	54,336	54,336	54,336	54,336	54,336

Note: The outcomes in Columns (1), (2), (5) and (6) are measured at the school-year level using information from (t+1), while Column (3) measures high school graduation using information from (t+3) . Eligibility for LIFE scholarship to a four year institution is based on meeting both the grade point average and SAT/ACT criteria established by the state. Academic index includes the outcomes from Columns (1)to (3). The index is constructed by averaging z-scores over all nonmissing components. The students' satisfaction index, reported in column (5), includes percent of students satisfied with (i) learning environment, (ii) social and physical environment, and (iii) homeschool relationship. The index is constructed by averaging z-scores of each component. Character development score, reported in column (6), is a state-constructed index capturing non-academic aspects of school performance based on administrative indicators such as discipline, attendance and school climate. It is reported on a 0-4 scale where higher values indicate better outcomes. Regressions in panel A are weighted by the total school enrollment at time t. Covariates include fraction of students at the school who are female and black and academic year fixed effects. The outcomes in columns (7) to (14) are measured at individual level. Dependent variable in column (7) is an indicator for performing at or above the proficient level on the state's exit exam. Dependent variables in column (8) to (10) are indicators for being observed enrolled in grades 10, 11, and 12, respectively. Dependent variable in is the number of years observed in high school after grade 9. Dependent variable in columns (12) to (14) are indicators for being enrolled in the same school as in grade 9 in grades 10, 11, and 12, respectively. Covariates include indicators for gender, race, FRPL, and cohort fixed effects. Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 3: Effects of the Policy on Adult Health Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Hospitalizations			ED visits		Mortality	Health index
	Any	Number	Days	Any	Number		
Receipt of lower rating	-0.026**	-0.214***	-0.215**	-0.040**	-0.173*	-0.003*	-0.076***
	(0.011)	(0.080)	(0.097)	(0.020)	(0.093)	(0.001)	(0.026)
Control group mean	0.163	0.276	1.095	0.806	8.022	0.012	-0.001
Implied effect size (%)	-16.0	-19.3	-19.4	-5.0	-15.9	-25.0	N/A
Sample Size	54,336	54,336	54,336	54,336	54,336	54,336	54,336

Note: All specifications use a linear polynomial with a uniform kernel. Adulthood is defined up to age 28. Covariates include indicators for gender, race and free/reduced lunch and cohort fixed effects. The dependent variables are: an indicator for ever being hospitalized as an inpatient during adulthood (Column 1), number of hospitalizations during adulthood (Column 2), number of days hospitalized during adulthood (Column 3), an indicator for ever having an emergency department visit in adulthood (Column 4), number of emergency department visits in adulthood (Column 5), dying during our sample window (Column 6), and omnibus health index (Column 7). Health care utilization index is constructed by averaging the z-scores of variables in columns 1 to 6. OLS estimates in columns (1), (4), (6), and (7). Poisson pseudo maximum likelihood (PPML) estimates in columns (2), (3), and (5). Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 4: Effects of the Policy on Fertility

	(1) Has a birth	(2) Has a teen birth	(3) Number of births	(4) Age at first birth
Receipt of lower rating	-0.060**	-0.012	-0.156**	-1.310**
	(0.026)	(0.014)	(0.066)	(0.569)
Control group mean	0.610	0.194	1.263	21.64
Implied effect size (%)	-9.8	-6.2	-12.4	-6.1
Sample Size	25,918	25,918	25,918	25,918

Note: All specifications use a linear polynomial with a uniform kernel. Covariates include indicators for gender, race and free/reduced lunch and cohort fixed effects. The dependent variables are: an indicator for having birth by age 28 (Column 1), an indicator for having birth before the age of 20 (Column 2), number of births by age 28 (Column 3), and age at first birth (Column 4). Individuals without birth by age 28 are imputed zero as an outcome in dependent variable in column (4). Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 5: Effects of the Policy on Health Outcomes in Teenagehood

	(1)	(2)	(3)	(4)	(5)	(6)
	Hospitalizations			ED visits		Health
	Any	Number	Days	Any	Number	index
Receipt of lower rating	-0.005	-0.113	-0.044	-0.073	-0.334**	-0.075**
	(0.005)	(0.120)	(0.134)	(0.047)	(0.168)	(0.036)
Control group mean	0.039	0.039	0.128	0.387	0.846	0.014
Implied effect size (%)	-12.8	-10.7	-4.3	-18.9	-28.4	N/A
Sample Size	29,911	29,911	29,911	29,911	29,911	29,911

Note: All specifications use a linear polynomial with a uniform kernel. Teenagehood is defined as being in high school in grades 9 to 12. Covariates include indicators for gender, race and free/reduced lunch and cohort fixed effects. The dependent variables are: an indicator for ever being hospitalized as an inpatient (Column 1), number of hospitalizations (Column 2), number of days hospitalized (Column 3), an indicator for ever having an emergency department visit (Column 4), number of emergency department visits (Column 5), and omnibus health index (Column 6). Health care utilization index is constructed by averaging the z-scores of variables in columns 1 to 5. OLS estimates in columns (1), (4), and (6). Poisson pseudo maximum likelihood (PPML) estimates in columns (2), (3), and (5). Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 6: Effects of the Policy on Juvenile Crime

	(1)	(2)	(3)	(4)	(5)
		Category-specific juvenile arrests			
	Any juvenile arrest	Drugs or alcohol	Property	Violent	Other
Receipt of lower rating	-0.024**	-0.006**	-0.005	-0.009	-0.023**
	(0.011)	(0.023)	(0.004)	(0.006)	(0.010)
Control group mean	0.084	0.012	0.026	0.028	0.055
Implied effect size (%)	-28.6	-50.0	-19.2	-32.1	-41.8
Sample Size	47,580	47,580	47,580	47,580	47,580

Note: All specifications use a linear polynomial with a uniform kernel. Covariates include indicators for gender, race and free/reduced lunch and cohort fixed effects. The dependent variables are: an indicator for any juvenile arrest (Column 1), an indicator for having juvenile arrest related to drugs or alcohol (Column 2), an indicator for having juvenile arrest related to property crime (Column 3), an indicator for having juvenile arrest related to violent crime (Column 4), and an indicator for having juvenile arrest related to other crimes not enumerated in columns 2 to 4 (Column 5). Juvenile arrest is defined as an arrest before age 18. Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 7: Effects of the Policy on Employment and Earnings in Adulthood

	(1)	(2)	(3)	(4)
	P(working)		Annual earnings	
	Ages 23-25	Ages 26-28	Ages 23-25	Ages 26-28
Receipt of lower rating	-0.023	-0.012	5.987	20.785
	(0.028)	(0.024)	(492.82)	(648.91)
Control group mean	0.623	0.611	8,222.79	10,936.20
Implied effect size (%)	-3.7	-2.0	0.1	0.2
Sample size	54,336	54,336	54,336	54,336

Note: All specifications use a linear polynomial with a uniform kernel. Covariates include indicators for gender, race and free/reduced lunch and cohort fixed effects. The dependent variables are: indicators for being employed on unemployment insurance covered contract at ages 23 to 25 (Column 1) or ages 26 to 28 (Column 2) and annual earnings at ages 23 to 25 (Column 3) or ages 26 to 28 (Column 4). Earnings data are based on unemployment insurance records. Earnings are defined as average annual earnings deflated to 2015 Dollars with zeros imputed to individuals not observed on unemployment insurance covered contracts. Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 8: Heterogeneity Analysis: Adult Health Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Sex		Race		Free or reduced price lunch		Baseline achievement (p50)	
	Female	Male	Black	White	Eligible	Not eligible	Below	Above
				Panel A. P(any hospitalization)				
Receipt of lower rating	-0.031**	-0.022*	-0.022**	-0.029*	-0.031***	-0.015	-0.046***	-0.019*
	(0.012)	(0.013)	(0.010)	(0.018)	(0.011)	(0.014)	(0.015)	(0.010)
Control group mean	0.172	0.154	0.158	0.179	0.170	0.150	0.181	0.138
Implied effect size (%)	-18.0	-14.3	-13.9	-16.2	-18.2	-10.0	-25.4	-13.8
				Panel B. Total number of hospitalizations				
Receipt of lower rating	-0.253***	-0.177	-0.229***	-0.133	-0.250***	-0.128	-0.318***	-0.152
	(0.071)	(0.108)	(0.086)	(0.126)	(0.082)	(0.112)	(0.108)	(0.099)
Control group mean	0.292	0.260	0.272	0.296	0.289	0.253	0.310	0.216
Implied effect size (%)	-22.4	-16.2	-20.5	-12.5	-22.1	-12.0	-27.2	-14.1
				Panel C. Total days hospitalized				
Receipt of lower rating	-0.359***	-0.129	-0.267**	-0.100	-0.296**	-0.110	-0.268**	-0.131
	(0.102)	(0.141)	(0.115)	(0.191)	(0.117)	(0.138)	(0.125)	(0.148)
Control group mean	1.122	1.070	1.140	1.059	1.161	0.982	1.240	0.820
Implied effect size (%)	-30.2	-12.1	-23.4	-9.5	-25.6	-10.4	-23.5	-12.3
				Panel D. P(any emergency department visit)				
Receipt of lower rating	-0.040**	-0.039*	-0.045**	-0.019	-0.048***	-0.023	-0.024	-0.043**
	(0.019)	(0.023)	(0.021)	(0.030)	(0.018)	(0.029)	(0.022)	(0.022)
Control group mean	0.832	0.782	0.842	0.766	0.846	0.737	0.884	0.755
Implied effect size (%)	-4.8	-5.0	-5.3	-2.5	-5.7	-3.1	-2.7	-5.7
				Panel E. Total number of emergency department visits				
Receipt of lower rating	-0.167*	-0.179**	-0.207**	-0.038	-0.196**	-0.100	-0.133	-0.176*
	(0.100)	(0.090)	(0.098)	(0.135)	(0.092)	(0.123)	(0.099)	(0.105)
Control group mean	9.960	6.290	8.722	7.176	9.125	6.152	10.050	5.736
Implied effect size (%)	-15.4	-16.4	-18.7	-3.7	-17.8	-9.5	-12.5	-16.1
				Panel F. Mortality				
Receipt of lower rating	-0.004*	-0.002	-0.004*	-0.003	-0.003	-0.004	-0.001	0.000
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
Control group mean	0.009	0.014	0.013	0.009	0.013	0.009	0.013	0.007
Implied effect size (%)	-42.2	-14.3	-30.8	-33.3	-23.1	-44.4	-7.7	0.0
Sample size	25,918	28,418	39,110	13,647	36,456	17,880	17,667	17,655

Note: All specifications use a linear polynomial with a uniform kernel. There are no additional covariates beyond cohort fixed effects in these regressions. Outcome variables are in panels: an indicator for ever being hospitalized as an inpatient during adulthood (panel A), number of hospitalizations during adulthood (panel B), number of days hospitalized during adulthood (panel C), an indicator for ever having an emergency department visit in adulthood (panel D), number of emergency department visits in adulthood (panel E), dying during our sample window (panel F). Columns represent heterogeneity dimensions with splits by sex (Columns 1 and 2), race (Columns 3 and 4), poverty proxy (Columns 5 and 6), and baseline achievement in grade 8 test scores (Columns 7 and 8). Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 9: Heterogeneity Analysis: Fertility

	(1)	(2)	(3)	(4)	(5)	(6)
	Race		Free or reduced price lunch		Baseline achievement (p50)	
	Black	White	Eligible	Not eligible	Below	Above
			Panel A. P(has a birth)			
Receipt of lower rating	-0.054**	-0.100*	-0.049*	-0.085**	-0.022	-0.076*
	(0.023)	(0.053)	(0.025)	(0.035)	(0.024)	(0.039)
Control group mean	0.620	0.604	0.644	0.550	0.686	0.545
Implied effect size (%)	-8.7	-16.6	-7.6	-15.5	-3.2	-13.9
			Panel B. Total number of births			
Receipt of lower rating	-0.156***	-0.220	-0.138**	-0.194**	-0.099	-0.169
	(0.055)	(0.138)	(0.066)	(0.085)	(0.069)	(0.103)
Control group mean	1.315	1.191	1.373	1.074	1.500	1.058
Implied effect size (%)	-11.9	-18.5	-10.1	-18.1	-6.6	-16.0
			Panel C. Age at first birth			
Receipt of lower rating	-1.113**	-2.168**	-1.082*	-1.859**	-0.462	-1.582*
	(0.503)	(1.083)	(0.560)	(0.747)	(0.539)	(0.819)
Control group mean	21.48	21.98	21.30	22.31	20.79	21.76
Implied effect size (%)	-5.2	-9.9	-5.1	-8.3	-2.2	-7.3
Sample size	18,847	6,317	17,622	8,296	8,682	8,680

Note: All specifications use a linear polynomial with a uniform kernel. There are no additional covariates beyond cohort fixed effects in these regressions. Outcome variables are in panels: an indicator for having birth by age 28 (panel A), number of births by age 28 (panel B), and age at first birth (panel C). Individuals without birth by age 28 are imputed zero as an outcome in panel C. Columns represent heterogeneity dimensions with splits by race (Columns 1 and 2), poverty proxy (Columns 3 and 4), and baseline achievement in grade 8 test scores (Columns 5 and 6). Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table 10: Effects of the Policy on Health at Birth of the Next Generation

	(1)	(2)	(3)	(4)	(5)	(6)
	P(birth weight < 2500g)			P(gestation weeks < 36)		
	Conditional	Missing = 0	Missing = 1	Conditional	Missing = 0	Missing = 1
Receipt of lower rating	0.002	-0.002	0.045**	0.003	-0.001	0.047**
	(0.007)	(0.002)	(0.021)	(0.006)	(0.005)	(0.022)
Control group mean	0.104	0.011	0.324	0.064	0.048	0.294
Implied effect size (%)	1.9	-18.2	13.9	4.7	-2.1	16.0
Sample Size	32,261	42,311	42,311	32,261	42,311	42,311

Note: Sample is limited to female students observed in schools around the cutoff whom we also observe in adulthood with or without births. All specifications use a linear polynomial with a uniform kernel. Covariates include indicators for gender, race and free/reduced lunch and cohort fixed effects. The dependent variables are an indicator for birth weight below 2500g (Columns 1 to 3) and an indicator for having gestational age below 36 weeks (Columns 4 to 6). Columns 1 and 4 present conditional estimates meaning that the sample is limited to mothers whom we observe giving at least one birth. Columns 2, 3, 5, and 6 include all high school females from the baseline sample and present extreme imputation bounds. Specifically in columns 2 and 5 we impute zero to all females without childbirth while in columns 3 and 6 we impute one to all females without childbirth. Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Appendix: For Online Publication

Effects of school quality on health and fertility

Ozkan Eren, Krzysztof Karbownik, and Orgul Ozturk

A Appendix Tables

Table A1: Summary Statistics by Accountability Ratings

	(1)	(2)	(3)
		School Ratings	
	All	Unsatisfactory	Below average
	Panel A: Student Characteristics		
Female	0.487	0.482	0.472
Black	0.410	0.791	0.646
White	0.549	0.185	0.321
Free or reduced price lunch	0.436	0.711	0.629
Math test score (z-score) in 8th grade	-0.003	-0.543	-0.402
ELA test score (z-score) in 8th grade	-0.002	-0.511	-0.379
	Panel B: Adult Outcomes		
Any adult hospitalization	0.139	0.161	0.163
Total number of days hospitalized	0.904	1.131	1.095
Any adult ED visit	0.692	0.810	0.806
Any birth	0.526	0.626	0.610
Total number of births	1.040	1.324	1.263
Teen birth	0.154	0.208	0.194
Sample size	309,890	27,784	26,552
School-year observations	1,111	136	97

Note: The tabulations reflect the research sample which comprises first time ninth graders between the 2000-2001 and 2005-2006 academic years. Column 1 presents all students and schools while columns 2 and 3 present our empirical sample. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A2: Regression Discontinuity and Local Randomization Validation Tests: Predicted outcomes

	(1)	(2) Hospitalizations	(3)	(4) ED visits	(5)	(6)	(7)	(8)	(9)
	P(any)	Number	Number of days	P(any)	Number	Mortality	P(any birth)	Number of births	Age at first birth
Panel A: Local linear regression									
Receipt of lower rating	0.001	0.001	-0.001	-0.001	-0.018	-0.000	0.004	0.011	0.087
	(0.003)	(0.005)	(0.014)	(0.006)	(0.141)	(0.000)	(0.005)	(0.015)	(0.090)
Control group mean	0.156	0.259	1.027	0.808	7.580	0.010	0.613	1.270	26.620
Implied imbalance (%)	0.4	0.3	-0.1	-0.1	-0.2	0.0	0.7	0.8	0.3
Sample Size	35,326	35,326	35,326	35,326	35,326	35,326	17,365	17,365	17,365
Panel B: Local Randomization									
Receipt of lower rating	0.002	0.003	0.011	0.004	0.109	0.000	0.007	0.020	0.131
	(0.002)	(0.004)	(0.012)	(0.006)	(0.131)	(0.000)	(0.005)	(0.014)	(0.090)
Control group mean	0.157	0.260	1.003	0.804	7.325	0.010	0.608	1.246	26.320
Implied imbalance (%)	1.2	1.2	1.1	0.4	1.5	1.0	1.2	1.6	0.5
Sample Size	14,382	14,382	14,382	14,382	14,382	14,382	7,101	7,101	7,101

Note: All specifications in Panel A use a linear polynomial with a uniform kernel. The analysis sample in Panel B is restricted to the interval of the running variable that contains only the index points, two on each side, that are immediately consecutive to the cutoff value. Panel B reports the point estimates from regressions of observable student characteristics on an indicator for the receipt of a lower accountability rating inside the interval. Standard errors are clustered at the school level. Outcomes variables are our predicted values for our main health outcomes in adulthood (Columns 1 to 6) and our main fertility outcomes in adulthood (Columns 7 to 9) with predictions based on observed pre-determined covariates. Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A3: Robustness Checks: Adult Health Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Nonparametric	Local randomization	Donut RD	Control for 8th grade test scores	Include school-grade characteristics	2005-2022 sample
P(any hospitalization)	-0.026**	-0.030**	-0.018*	-0.029**	-0.032***	-0.025**	-0.031**
	(0.011)	(0.014)	(0.010)	(0.013)	(0.010)	(0.010)	(0.013)
Sample size	54,336	31,851	22,609	47,983	35,322	53,191	54,336
Number of hospitalizations	-0.214***	-0.063**	-0.131*	-0.198*	-0.250***	-0.208***	-0.215***
	(0.080)	(0.030)	(0.076)	(0.102)	(0.087)	(0.078)	(0.079)
Sample size	54,336	31,851	22,609	47,983	35,322	53,191	54,336
Number of days hospitalized	-0.215**	-0.279**	-0.177*	-0.183	-0.213*	-0.211**	-0.200**
	(0.097)	(0.130)	(0.093)	(0.127)	(0.113)	(0.094)	(0.091)
Sample size	54,336	39,876	22,609	47,983	35,322	53,191	54,336
P(any ED visit)	-0.040**	-0.012	-0.031	-0.049**	-0.033*	-0.039**	-0.034*
	(0.020)	(0.015)	(0.020)	(0.024)	(0.019)	(0.019)	(0.018)
Sample size	54,336	31,851	22,609	47,983	35,322	53,191	54,336
Number of ED visits	-0.173*	-0.229	-0.127	-0.250**	-0.150	-0.186**	-0.175**
	(0.093)	(0.394)	(0.093)	(0.102)	(0.097)	(0.085)	(0.088)
Sample size	54,336	31,851	22,609	47,983	35,322	53,191	54,336
Mortality	-0.003*	-0.004	-0.002	-0.004	-0.005**	-0.003*	-0.003*
	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)
Sample size	54,336	39,876	22,609	47,893	35,322	53,191	54,336

Note: This table resents a series of robustness checks for our primary adulthood health outcomes of interest. Each coefficient comes from a separate regression with outcomes indicated in rows. Column 1 replicates the baseline results for comparison. Column 2 presents non-parametric estimation, column 3 presents local randomization estimation, column 4 presents donut RD where we exclude observations close to the cutoff, column 5 controls for 8th grade test scores, column 6 includes school-by-grade covariates in high school, and column 7 includes Covid-19 years from the sample. Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A4: Robustness Checks: Fertility

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Nonparametric	Local randomization	Donut RD	Control for 8th grade test scores	Include school-grade characteristics	2005-2022 sample
Has a birth	-0.060**	-0.059***	-0.043*	-0.064**	-0.049*	-0.063***	-0.057**
	(0.026)	(0.022)	(0.025)	(0.028)	(0.028)	(0.021)	(0.026)
Sample size	25,918	19,003	10,861	22,842	17,362	25,356	25,918
Number of births	-0.156**	-0.096	-0.107*	-0.194***	-0.135**	-0.166***	-0.175**
	(0.066)	(0.064)	(0.062)	(0.067)	(0.066)	(0.050)	(0.072)
Sample size	25,918	19,003	10,861	22,842	17,362	25,356	25,918
Age at first birth	-1.310**	-2.511***	-1.012*	-1.259**	-1.040*	-1.374***	-1.216**
	(0.569)	(0.604)	(0.543)	(0.593)	(0.617)	(0.470)	(0.571)
Sample size	25,918	15,118	10,861	22,842	17,362	25,356	25,918

Note: This table resents a series of robustness checks for our primary adulthood fertility outcomes of interest. Each coefficient comes from a separate regression with outcomes indicated in rows. Column 1 replicates the baseline results for comparison. Column 2 presents non-parametric estimation, column 3 presents local randomization estimation, column 4 presents donut RD where we exclude observations close to the cutoff, column 5 controls for 8th grade test scores, column 6 includes school-by-grade covariates in high school, and column 7 includes Covid-19 years from the sample. Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

Table A5: Effects of the Policy on Learning Disability Designation in High School

	(1)	(2)	(3)	(4)	(5)
	Classified as learning disabled in grade				Any grade
	9	10	11	12	
Receipt of lower rating	-0.000	-0.017**	-0.013***	-0.011*	-0.017**
	(0.012)	(0.008)	(0.005)	(0.006)	(0.008)
Control group mean	0.075	0.056	0.045	0.041	0.071
Implied effect size (%)	0.0	-30.4	-28.9	-26.8	-23.9
Sample Size	54,336	54,336	54,336	54,336	54,336

Note: All specifications use a linear polynomial with a uniform kernel. Covariates include indicators for gender, race and free/reduced lunch and cohort fixed effects. The dependent variables are indicators for being classified as learning disabled in grades 9 to 12 (Columns 1 to 4) or in any of those grades (Column 5). Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

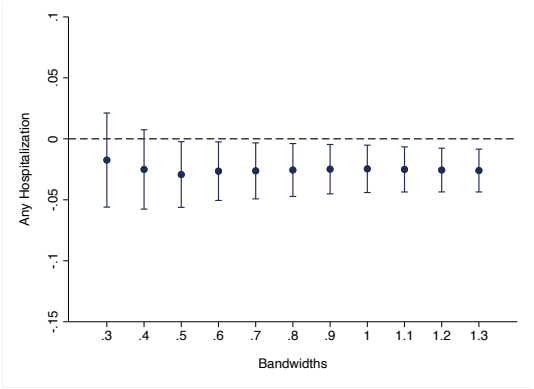
Table A6: Effects of the Policy on Welfare Participation in Adulthood

	(1)	(2)	(3)
	Welfare enrollment		
	Any	TANF	SNAP
Receipt of lower rating	-0.005	-0.001	-0.006
	(0.015)	(0.009)	(0.015)
Control group mean	0.674	0.159	0.673
Implied effect size (%)	-0.7	-0.6	-0.9
Sample Size	54,510	54,510	54,510

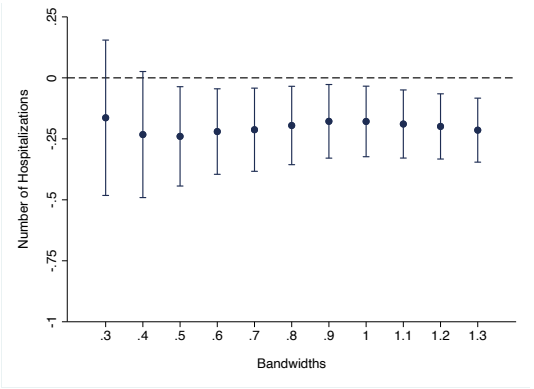
Note: All specifications use a linear polynomial with a uniform kernel. Covariates include indicators for gender, race and free/reduced lunch and cohort fixed effects. The dependent variables are indicators for being on any welfare assistance in adulthood (Column 1), receiving TANF in adulthood (Column 2), and receiving SNAP (Column 3). Adulthood is defined up to age 28. Standard errors are clustered at the school level. ***, ** and * mark estimates statistically different from zero at the 99, 95 and 90 percent confidence level.

B Appendix Figures

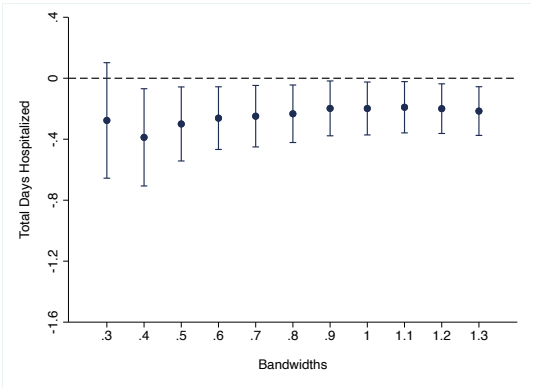
Figure A1: Robustness of Adult Health Results: Local Linear Estimates with Varying Bandwidth



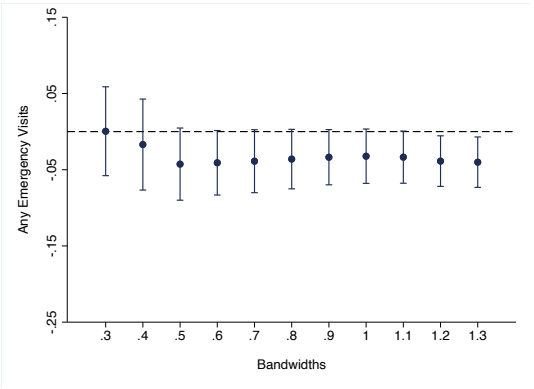
(a) P(any hospitalization)



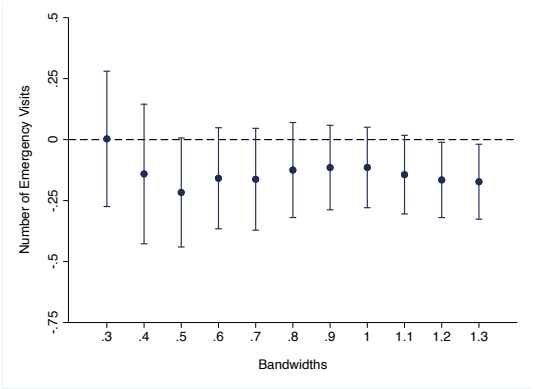
(b) Number of hospitalizations



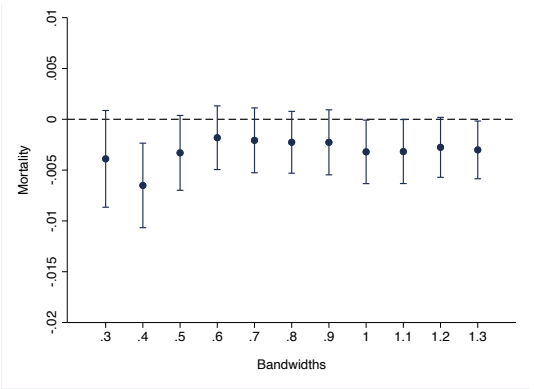
(c) Total days hospitalized



(d) P(any ED visit)



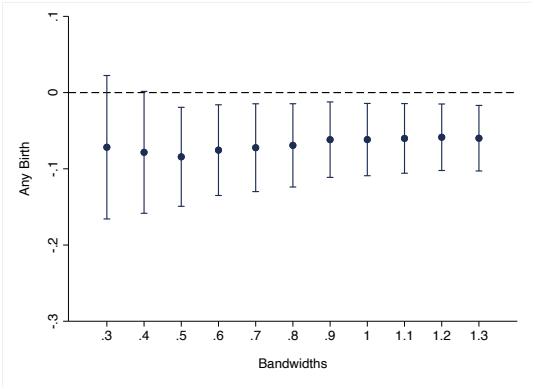
(e) Number of ED visits



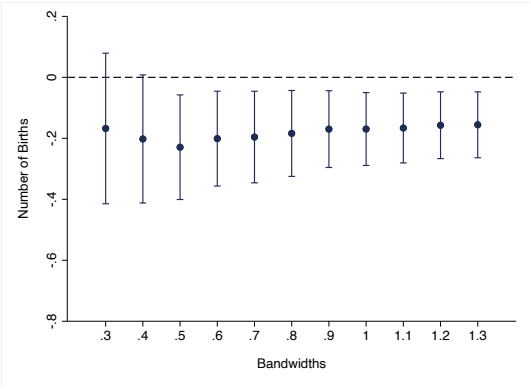
(f) P(dead)

Notes: Bandwidth robustness analysis for our main adulthood health outcomes of interest. All estimates are obtained from a linear polynomial using alternative bandwidths, over a range from 0.3 to 1.3 index points by incrementally adding 0.1 points, with a uniform kernel. The height of the bars from each point represents the bounds of the 95% confidence interval. Standard errors are clustered at the school level.

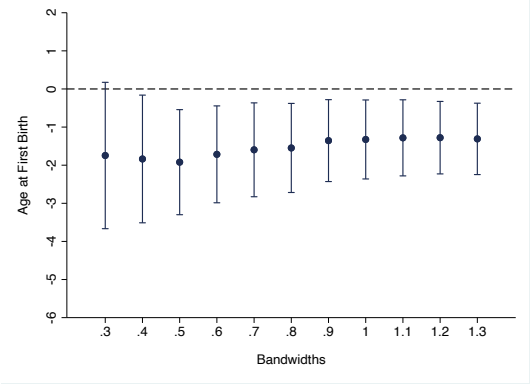
Figure A2: Robustness of Fertility Results: Local Linear Estimates with Varying Bandwidth



(a) P(any hospitalization)



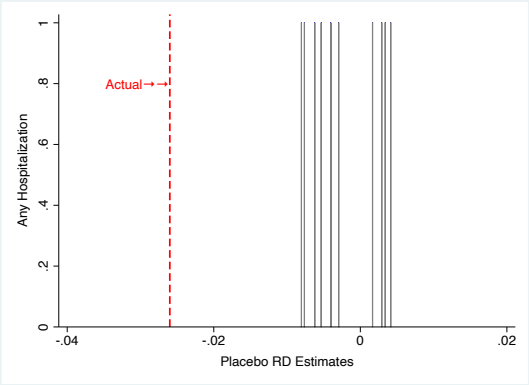
(b) Number of hospitalizations



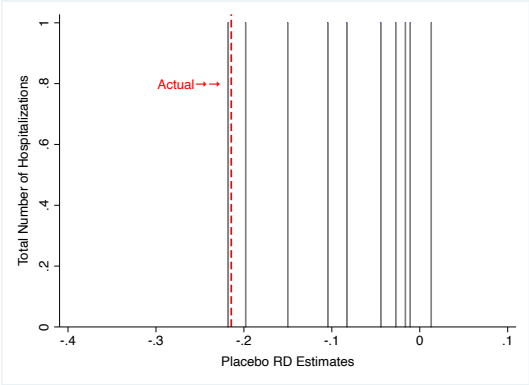
(c) Total days hospitalized

Notes: Bandwidth robustness analysis for our main adulthood fertility outcomes of interest. All estimates are obtained from a linear polynomial using alternative bandwidths, over a range from 0.3 to 1.3 index points by incrementally adding 0.1 points, with a uniform kernel. The height of the bars from each point represents the bounds of the 95% confidence interval. Standard errors are clustered at the school level.

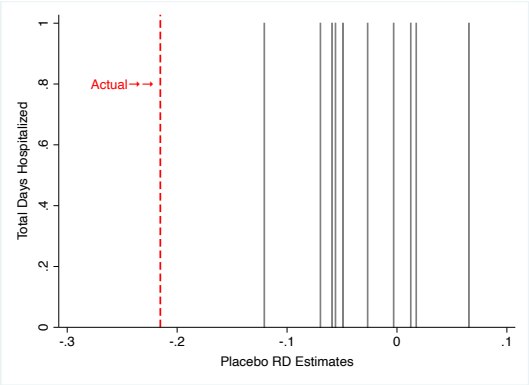
Figure A3: Robustness of Adult Health Results: Placebo Estimates



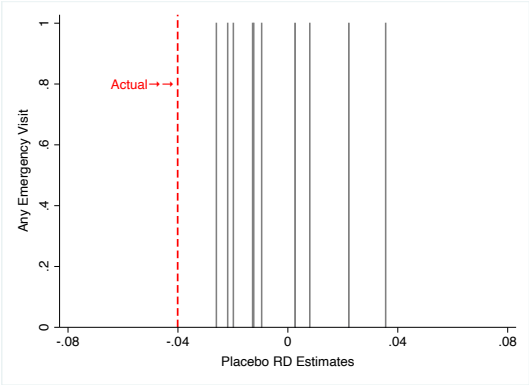
(a) P(any hospitalization)



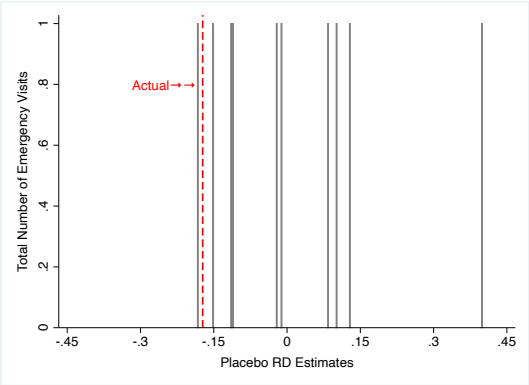
(b) Number of hospitalizations



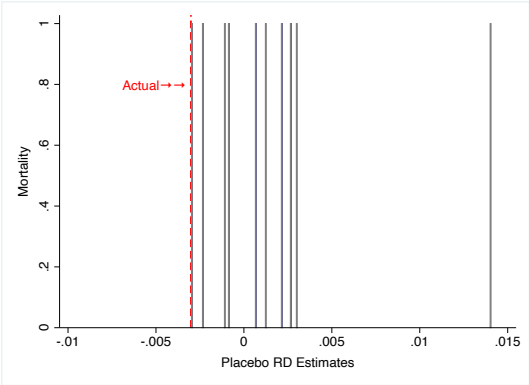
(c) Total days hospitalized



(d) P(any ED visit)



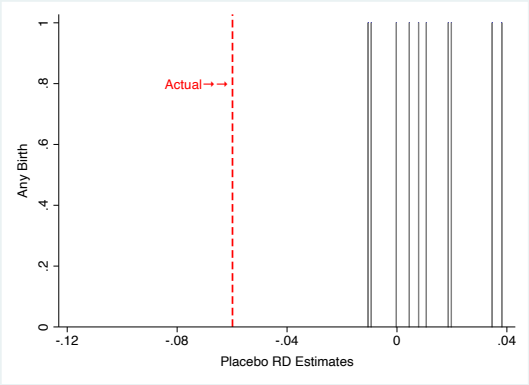
(e) Number of ED visits



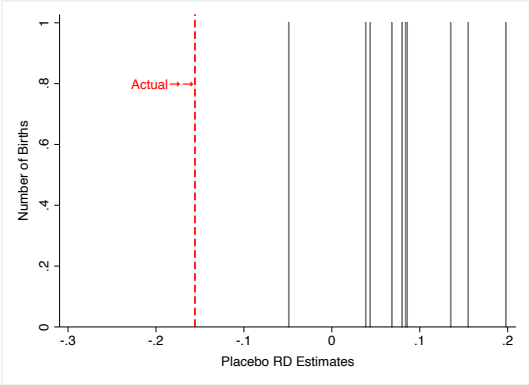
(f) P(dead)

Notes: Placebo estimates for our main adulthood health outcomes of interest. Each placebo estimate assigns a false rating cutoff by incrementally adding 0.1 index points to the actual accountability score cutoff (Unsatisfactory vs. Below Average) over a range of [-1.0, 0.2]. All specifications use a linear polynomial with a uniform kernel.

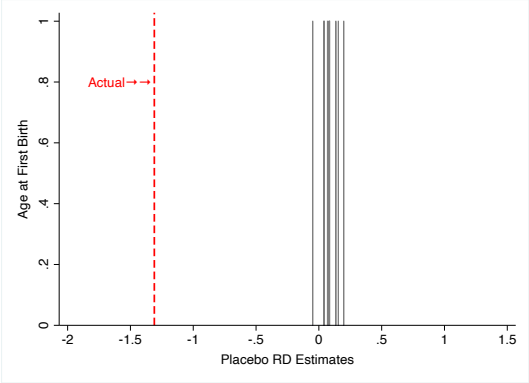
Figure A4: Robustness of Fertility Results: Placebo Estimates



(a) P(any birth)



(b) Number of births



(c) Age at first birth

Notes: Placebo estimates for our main adulthood fertility outcomes of interest. Each placebo estimate assigns a false rating cutoff by incrementally adding 0.1 index points to the actual accountability score cutoff (Unsatisfactory vs. Below Average) over a range of [-1.0, 0.2]. All specifications use a linear polynomial with a uniform kernel.