



The returns to early-life interventions for very low birth weight children

Eric Chyn^{a,*}, Samantha Gold^b, Justine Hastings^c

^a Dartmouth College, RIPL, NBER, United States

^b RIPL, United States

^c Brown University, RIPL, NBER, United States



ARTICLE INFO

Article history:

Received 26 November 2019

Received in revised form 14 October 2020

Accepted 30 October 2020

Available online 4 December 2020

JEL classification:

I11

I12

I18

I21

J13

Keywords:

Medical care

Early-life interventions

ABSTRACT

We use comprehensive administrative data from Rhode Island to measure the impact of early-life interventions for low birth weight newborns. Our analysis relies on a regression discontinuity design based on the 1,500 g threshold for Very Low Birth Weight (VLBW) status. We find that threshold crossing causes more intense in-hospital care, in line with prior studies. In terms of later-life outcomes, we show that threshold crossing causes a 0.34 standard deviation increase in test scores in elementary and middle school, a 17.1 percentage point increase in the probability of college enrollment, and a \$66,997 decrease in social program expenditures by age 14.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Children born with low birth weight tend to have more health difficulties and worse later-life outcomes relative to their peers born at normal birth weight (Almond and Currie, 2011; Currie, 2011). A key question for public policy is whether interventions delivered during early-life can counteract the negative effects of low weight and poor health at birth. For example, investments in additional medical care could generate substantial social benefits by enhancing later-life economic self-sufficiency and reducing reliance on social programs.

This paper provides new evidence on the effectiveness of early-life interventions for children born with birth weight below the 1,500 g (3 pounds and 5 ounces) threshold for Very Low Birth Weight (VLBW) status. Prior research has studied this question using a regression discontinuity (RD) design that exploits changes in hospital treatment across the VLBW threshold. Almond et al. (2010) find that children in the U.S. born just under 1,500 g have higher hospital costs and lower one-year mortality rates. Using data from Chile and Sweden, Bharadwaj et al. (2013) find similar results and show that children with birth weight just below 1,500 g have 0.13 to 0.22 standard deviations higher test scores measured during elementary and high school grades.

We build on previous studies using comprehensive administrative data from Rhode Island. The data contain over three decades of birth records that we join to a range

* Corresponding author.

E-mail addresses: eric.t.chyn@dartmouth.edu (E. Chyn), sgold@ripl.org (S. Gold), justine.hastings@brown.edu (J. Hastings).

of outcomes for education and social program participation. We study academic achievement throughout primary and secondary public education, as well as college attendance. We also examine enrollment in social programs and total benefits received during childhood. To explore post-hospital investment mechanisms, we use measures of parental behavior such as records on maternal labor market activity and school choice.¹

We find significant and large positive impacts of crossing the 1,500 g VLBW threshold from above on educational outcomes at all levels. Children born just under 1,500 g perform 0.34 standard deviations better on test scores measured during grades 3–8 relative to those born just above the threshold. For each grade throughout childhood, we find similar magnitude impacts, suggesting that factors influencing higher academic performance accrued prior to enrollment in school. These gains in test scores are matched by significant and positive effects on post-secondary educational enrollment. Threshold crossing is associated with a 17.1 percentage point (32 percent) increase in the likelihood of enrolling in any college by age 22.

There are also significant threshold crossing effects on the amount of social program expenditures that a child received. We measure spending using payment records for SNAP and TANF, claims files from Medicaid, and school data on enrollment in special education programs (IEP). By age 14, children born just under the 1,500 g threshold receive \$66,997 less in social program expenditures. This is a large reduction relative to \$145,486 spent on children born above the threshold. As another comparison, Almond et al. (2010) estimate that hospital costs increase by an additional \$4,000 for children just under the 1,500 g threshold.

Given the size of estimated impacts relative to the in-hospital expenditures, we analyze two types of potential post-hospital mechanisms. First, we examine proxies for parental behavior using maternal employment records and responses to the Center for Disease Control's Pregnancy Risk Assessment Monitoring Program System (PRAMS) survey. We find no significant impacts on employment or earnings for mothers within the first two years after birth. Similarly, we find no significant impacts on measures for maternal care and stress based on the PRAMS survey. Although imprecise, the point estimates suggest that, if anything, threshold crossing increases care and reduces stress for mothers.

Second, we study whether threshold crossing affects school choice by creating value-added measures for public schools in Rhode Island. For elementary and middle school grades, we study school value-added for standardized test scores. For high school grades, we study school value-added in terms of enrollment in college. We find small, insignificant impacts of crossing the 1,500 g threshold on measures

¹ As we discuss in Section 3, we study the reduced form impact of VLBW status. Prior studies have noted that crossing the VLBW threshold identifies the combined impact of changes in medical treatment provided in-hospital and changes in post-hospital inputs (Bharadwaj et al., 2013). Post-hospital inputs could change at the threshold because parents may also react to VLBW status and the additional medical care provided to their children.

of school quality. The results suggest impacts on academic achievement are not driven by parents selecting substantially better schools.

The results are robust to standard regression discontinuity checks for balance in baseline characteristics across the 1,500 g threshold and tests of the sensitivity of our estimates to higher-order polynomial RD specifications. In addition, we address concerns about non-random heaping at the VLBW threshold by excluding observations within a small window around the 1,500 g threshold (Barreca et al., 2011). Finally, we show that our results do not appear to be affected by differential survival rates (or other forms of attrition) around the VLBW threshold. In our sample, we find little evidence of significant threshold crossing effects on the likelihood of having a test score or enrolling in high school.²

Overall, our findings add to the literature on the effectiveness of early-life interventions and medical treatments for at-risk children. We extend Almond et al. (2010) by showing positive impacts on long-run measures of human capital and reductions in social program expenditures throughout childhood. The impacts on childhood test scores are consistent with findings for Chile and Norway reported in Bharadwaj et al. (2013).³ The effect sizes that we detect are also similar in magnitude to experimental and quasi-experimental evaluations of the test-score impact of early childhood education programs (Duncan and Magnuson 2013).⁴ Our results more generally contribute to the literature demonstrating that interventions delivered in early-life can generate notable gains for children (Currie and Thomas, 1995; Chetty et al., 2011; Havnes and Mogstad, 2011; Heckman et al., 2013; Chetty et al., 2016; Hoynes et al., 2016; Chyn, 2018; Bald et al., 2019).

2. Background on medical and non-hospital care for VLBW children

The American Academy of Pediatrics (AAP) publishes guidelines for U.S. neonatal care that recommend admitting all VLBW children into a Neonatal Intensive Care Unit (NICU) (American Academic of Pediatrics and Committee on Fetus and Newborn, 2012). At the NICU, a child receives continuous, specialized care from a broad range of specialists including neonatologists, pediatric nurses, and respiratory therapists. The NICU also has specialized equipment (e.g., mechanical ventilators) to provide life support

² Using the census of U.S. births from 1983–2002, Almond et al. (2010) estimate that crossing the VLBW threshold decreases one-year mortality by one percentage point. In our sample of births in Rhode Island, there is no statistically significant discontinuity in the likelihood of having a test score in grades 3 through 8. We present these results in table A.13. At the same time, the standard error for our estimate does not rule out threshold crossing effects on mortality (and survivorship) that are as small as one percentage point.

³ Daysal et al. (2019) study data from Denmark and find that children born just below the VLBW threshold have higher test scores in ninth grade. They also document large positive spillovers on academic outcomes of siblings.

⁴ Duncan and Magnuson (2013) review of 84 evaluations of early childhood education programs and find that the average effect size is 0.35 standard deviations.

for as long as necessary. In 2012, the admission rate to the NICU for VLBW children was 844.1 per 1,000 live births (Harrison and Goodman, 2015).

The VLBW designation also may affect receipt of specific medical treatments and procedures. For example, the 1,500 g threshold is commonly cited as a threshold for using diagnostic ultrasounds.⁵ As noted in Almond et al. (2010), it is likely that VLBW infants also receive differential treatment in terms of health inputs that are often difficult to measure in administrative data. For example, Angert and Adam (2009) recommend that VLBW infants should be “handled gently and not placed in a head down or Trendelenburg position” in order to minimize risk of intraventricular hemorrhage or brain injury.

Finally, children born under 1,500 g often receive additional post-hospital treatment. The AAP guidelines recommend that high-risk infants should receive follow-up services to conduct neurological assessment. Many hospitals have adopted this recommendation and created follow-up programs that specifically use VLBW status as a criterion for inclusion. In our context of Rhode Island, the Women and Infants Hospital routinely schedules follow-up appointments from birth to adolescence for children born less than 1,500 g.⁶ This follow-up program provides medical management for infants with respiratory problems and developmental assessments for cognitive, language, motor skills, behavior, executive functioning, memory, and phonological processing (Vohr et al., 2010).⁷

3. Conceptual framework and empirical strategy

Why might crossing the 1,500 g threshold from above matter for long-run outcomes of children? This section provides a conceptual framework that highlights mechanisms that could drive an impact of crossing the threshold for VLBW status. Our discussion is based on the model introduced in Bharadwaj et al. (2013). We assume that initial health at birth (H) is based on birth weight (BW) and a component unobserved to the econometrician (ε). Physicians and health care workers provide treatments (D) which are determined by a decreasing function of initial health and a random component (ν). The discussion from Section 2 also suggests that medical treatments increase discretely by the amount κ at the 1,500 g threshold because health care providers are responsive to the VLBW designation. Formally, we assume the following two-equation model of health and medical treatments for individual i :

$$H_i = BW_i + \varepsilon_i \tag{1}$$

$$D_i = g(H_i) + \kappa 1(BW_i < 1,500g) + \nu_i \tag{2}$$

⁵ Almond et al. (2010) note that the neonatal manual at the Longwood Medical Area (in Boston, MA) stipulates the use of diagnostic ultrasounds for at-risk newborns.

⁶ Note that we do not have data on follow-up appointments at the Women and Infants Hospital so we cannot examine participation in this program for our main analysis sample.

⁷ Similarly, 90 percent of NICUs in California are funded by the California Children’s Services (CCS) program. All CCS-funded NICUs are required to provide a neonatal follow-up service for high-risk infants (California Perinatal Quality Care Collaborative, 2018).

This framework makes clear that medical treatments D will be correlated with unobserved health status through the decreasing function $g(H)$. This implies that estimating the impact of early-life medical interventions on later-life outcomes can suffer from endogeneity even when conditioning on birth weight.

To address the identification challenge implied by Eqs. (1) and (2), we use the following regression discontinuity (RD) specification proposed by Almond et al. (2010) and Bharadwaj et al. (2013):

$$y_i = \alpha + \gamma d_i + \theta 1(BW_i < 1,500g) + \delta 1(BW_i < 1,500g) d_i + \varepsilon_i \tag{3}$$

where y is an outcome of interest, d is a running variable which is equal to the difference between an infant’s BW and the 1,500 g threshold for VLBW status, $1(BW < 1,500g)$ is an indicator for VLBW status, and ε is a regression error term. The key regression parameter of interest is θ which is the coefficient on the indicator for crossing the VLBW threshold from above. Estimates of this coefficient will reflect the two types of effects. First, the causal effects of medical care will be captured because treatments jump at the VLBW threshold by the amount $\Delta D = \kappa$ and this discrete shift is uncorrelated with unobserved determinants of initial health. Second, the effects of post-hospital investments will also be captured if parental behavior or other post-hospital inputs (e.g., follow-up care appointments for at-risk infants) respond to the VLBW designation.

To formalize this discussion of mechanisms, we assume there is a structural model of outcomes that features both medical treatments and post-hospital investments. For the latter, we assume investments up to period t are a function of post-hospital investments (ξ), initial child health, and medical treatments at birth: $I_t^{post}(\xi, H, D)$. The model for an outcome for individual i at period t can be written as:

$$y_{it} = \phi_t H_i + \psi_t D_i + \tau I_t^{post}(\xi_i, H_i, D_i) + \omega_{it}. \tag{4}$$

This framework makes clear that threshold crossing estimates from the RD model from Eq. 3 will capture both the combined effects of local variation in D and $I_t^{post}(\xi, H, D)$. Formally, Eq. 4 allows us to re-write the threshold crossing estimate as:

$$\hat{\theta} = \psi_t \kappa + \tau \Delta I_t^{post} \tag{5}$$

where ΔI_t^{post} is the difference in the average post-hospital investments that children receive at the threshold for VLBW status. The estimated coefficient $\hat{\theta}$ should be thus interpreted as the total policy relevant effect of the increased medical care at the VLBW threshold, which can include potential changes (positive or negative) in accumulated investments. As discussed further in Section 6, we attempt to gauge the importance of post-hospital investments by studying threshold crossing impacts on several measures of parental behavior and non-hospital inputs.

An additional consideration for the RD analysis is that the composition of the sample for some outcomes could be affected if crossing the VLBW threshold reduces mortality. For example, we would not observe standardized test scores for a child who does not survive to school age. We can test whether concerns over composition bias are war-

ranted by conducting an RD analysis of sample attrition. Specifically, Section 5.7 presents results where we estimate impacts of threshold crossing on indicators for whether we observe children enrolled in Rhode Island public schools.

Finally, note that we estimate the RD model from Eq. 3 using OLS and cluster standard errors at the gram level to address concern about the discreteness of birth weight (Lee and Card, 2008). Our main results are based on using a window of 1,300–1,700 g for birth weight, which follows other recent studies of VLBW (Bharadwaj et al., 2013; Daysal et al., 2019).⁸ All our analysis excludes births where the infant's weight was within three grams of the 1,500 g threshold (Barreca et al., 2011).

4. Data

We use two types of data for our analysis. First, the bulk of our analysis relies on data from anonymized administrative records from a relational database created by researchers at Research Improving People's Lives (RIPL) in partnership with the state of Rhode Island. All personally identifiable information has been removed and replaced with anonymous identifiers. These identifiers allow researchers with approved access to join records across administrative sources (Hastings et al., 2019; Hastings, 2019). Second, we use hospital discharge data from the Healthcare Cost and Utilization Project (HCUP). Note that the HCUP records do not contain the identifying information needed to link these records to our administrative sources. In the following subsections, we provide further details on the samples and outcomes that we construct for our analysis.

4.1. HCUP sample and hospitalization outcomes

To provide evidence on the early-life medical interventions delivered in our context, we rely on data from the Healthcare Cost and Utilization Project (HCUP). The HCUP records for Rhode Island are available for the years 2002–2015. In these records, we focus on two samples of children. First, we study 1,724 children who weigh 1,300–1,700 g at birth and for whom HCUP provides non-missing data on the length of hospital stay and total charges. Second, we study 1,245 children who weigh 1,300–1,700 g at birth and for whom we have non-missing data on the number of days spent in the NICU.⁹

4.2. Main birth records sample

For our main analysis of outcomes in childhood and early adulthood, we construct an analysis sample based on birth records from the state of Rhode Island. We construct this sample as follows. We start with the universe of birth records from 1984 to 2016 ($N = 407,697$) and impose

two restrictions. First, because we identify impacts using an RD design in a window around the 1,500 g cutoff, we remove births where the infant's weight was less than 1300 or greater than 1,700 g. Second, following Barreca et al. (2011), we remove births where the infant's weight was within a “doughnut hole” of three grams around the 1,500 g threshold to exclude observations that could be subject to technological constraints in measurement precision and rounding tendencies by attending physicians (which cause bunching in birthweight around the cutoff). Based on these restrictions, our main analysis sample contains 2,726 children who weigh 1,300–1,700 g at birth.

4.3. Outcomes for the birth records sample

We join the main birth records sample to several administrative data sources to measure outcomes in childhood and early adulthood. As detailed in the following sections, each outcome is only defined for a subset of cohorts in the birth records. This is because data sources for outcomes do not perfectly overlap with the years covered by the birth records. As a result, the sample sizes vary for each outcome that we consider. Note that the samples for many of our main outcomes (i.e., test scores, achievement, and social program expenditures) have a similar size to those used in prior studies examining the effects of VLBW.¹⁰

4.3.1. Supplemental security income (SSI) program participation

Records from the Rhode Island Department of Human Services (RIDHS) provide data on enrollment in the SSI program during the period 1996–2015. The SSI program is a federal program that provides cash assistance to a variety of recipients, including disabled children. For children, eligibility for SSI is determined using birth weight or a “failure to thrive” benchmark (SSA, 2018a, 2018b).¹¹ We measure whether a child enrolled in SSI at any point from birth up to age 3.

4.3.2. Grade repetition, special education services (IEP) and test scores

Records from the Rhode Island Department of Education (RIDE) provide information on several key educational outcomes. We use records from the years 2003–2016 to study whether children repeat a grade or have a written Individualized Education Program (IEP). We study grade

⁸ Section 5.4 examines the sensitivity of results to changing this bandwidth. In addition, Section 5.4 studies robustness by reporting results based on specifications that feature quadratic functions for the running variable terms.

⁹ We have fewer children in the analysis for days spent in the NICU because this variable is only available for the years 2006–2015.

¹⁰ To better understand our prospective ability to detect impacts of crossing the 1,500 g threshold, it is useful to compare our samples sizes to those from Bharadwaj et al. (2013). As we detail below, we pool standardized test scores resulting in a sample of 3,070 person-year observations. For their analysis of test scores in Chile and in Norway, Bharadwaj et al. (2013) use a sample of 2,877 children in Chile and 1,163 in Norway (see Column 2 of Table 3 of their paper).

¹¹ Specifically, the SSA uses two rules to define whether a child has sufficiently low birth weight to qualify for SSI (SSA, 2018a). The first rule defines a child as being eligible based on whether a child has birth weight less than 1,200 g. The second rule defines a child as being eligible based on birth weight and gestational age. Evidence presented in Galdi et al. (2018) suggests that low birth weight children who are eligible for SSI may not participate in the program.

repetition during grades 1–4.¹² Similarly, we study the number of school years that a child has an IEP during the same grade range.¹³ We also use RIDE records from the years 2005–2014 to study performance on elementary and middle school tests. Specifically, we study test scores in grades 3, 5, and 8, in addition to conducting a pooled analysis using all test scores for grades 3–8. Note that all RIDE outcomes are available only for children who enroll in public schools in Rhode Island. As a result, children in the birth records sample who never enroll in public schools are not included in the analysis of grade repetition, IEP outcomes, or academic achievement.¹⁴

4.3.3. High school disciplinary offenses and college preparation

Records from RIDE provide information on disciplinary offenses and college preparation during high school. For the former, we use records from 2003 to 2016 to study the number of offenses during grades 9–12.¹⁵ For the latter, we use SAT, PSAT and AP records from the years 2010–2015. We use this data to construct an index measure for college preparation. For the PSAT and SAT components, we create indicators for whether a child took either exam in grades 9–12. For the AP component, we create an indicator for whether a child took at least one AP class in grades 9–12.¹⁶ As with our other RIDE outcomes (e.g., test scores in grades 3–8), the disciplinary offense and college preparation outcomes are only defined for children who enroll in public schools and are sufficiently old to be observed in high school.

4.3.4. Post-secondary enrollment

Records from the National Student Clearinghouse (NSC) provide information on post-secondary enrollment during 2003–2016. We study two enrollment outcomes: whether a child ever enrolls in a two- or four-year college by age 22 and whether a child ever enrolls in a four-year college by age 22. Unlike our other measures of academic achievement, we can measure college-going outcomes for children

regardless of whether they enroll in Rhode Island public schools.

4.3.5. Social program expenditures

Records from the Rhode Island Department of Human Services allow us to construct social program expenditure measures. We include payments from the following social programs: Supplemental Nutrition Assistance Program (SNAP), Medicaid, and Temporary Assistance for Needy Families (TANF). Anonymized program enrollment and payment records are available from 1997–2016.¹⁷ In addition, we include program costs for special education participation (i.e., having a written IEP), which we set to an average expenditure of \$34,000.¹⁸ We construct measures of expenditure received during childhood by adding up program payments received and IEP costs incurred by age 10, 12, or 14. Note that we use the per-capita payment value for family-level program payments.

4.3.6. Maternal care and stress

PRAMS survey data provide information on maternal care and stress outcomes.¹⁹ This survey collects information on maternal attitudes and experiences before, during, and after pregnancy. All mothers with low birth weight children (less than 2,500 g) receive the PRAMS survey, though only a fraction respond. We construct two indices from these data. The first is an index of maternal care based on survey questions on childcare practices such as breastfeeding or knowledge of proper sleeping position. This care index is available from 2002–2014. The second is an index of maternal stress based on responses to separate questions on stressful events, postpartum depression, and degree of child difficulty. The stress index is available from 2004–2014. Note that each index is set to missing if a respondent fails to answer all the questions associated with the measure.

4.3.7. Maternal labor market activity and subsequent fertility

We also study maternal labor market outcomes and subsequent fertility. For the former, we use records from the Rhode Island Department of Labor and Training cover-

¹² A child is indicated as repeating a grade if they are observed in the same grade in different school years.

¹³ Note that an IEP can be given in preschool, and children are assessed each year.

¹⁴ One potential concern is that enrollment in public schools may change at the threshold for VLBW status. This would change the composition of the sample of children on either side of the cutoff. The analysis that we conduct in Section 5.7 addresses this concern by estimating threshold crossing effects on the likelihood of enrolling in public school. We generally do not find any significant effects.

¹⁵ There are 44 distinct types of offenses recorded in RIDE records. These include the following: alcohol, arson, assault/battery of another student, assault/battery of a teacher, skipping class, skipping detention, skipping school, leaving school grounds, tardiness, truant, making bomb threats, cheating/plagiarism, prohibited use of communication devices, use of controlled substance, disorderly conduct, fighting, fire regulation violation, forgery, gambling, gang activity, harassment-stalking, harassment-sexual, harassment-prejudice, hazing, insubordination, kidnapping, larceny, obscene language toward student, obscene language toward teacher, robbery, sexual assault, sexual misconduct, unauthorized use of computers, threat, tobacco, trespassing, vandalism, and weapon possession.

¹⁶ We define the index for children who are enrolled in a Rhode Island public school for at least one high school grade

¹⁷ Note that the anonymized payment records do not have uniform coverage during the period 1997–2016. For example, SNAP payment records do not begin until 2004, while Medicaid claim records begin in 2000. Thus, children born prior to the latest source start date (2004) will be missing some payment records for the first years of their life. We address censoring issues by imputing program payments for children born before 2004 using the annual average payments during the years we observe them. For example, suppose a child was born in 2003. This child will be missing one year of social program participation since complete coverage for all children starts in 2004. To address this, we impute social program spending in 2003 using the average spending level observed in the years we do observe them (2004 - X), where X is the year in which the child turns 10, 12, or 14.

¹⁸ We calculate IEP per pupil costs using expenditure statistics from the Annual Per Pupil Expenditure Report from RIDE for the years 2012–2013 and IEP pupil data from the U.S. Department of Education ([Rhode Island Department of Education, 2013](#); [U.S. Department of Education, 2015](#)).

¹⁹ Note that the birth records from Rhode Island contain information on mothers. As mentioned, researchers at RIPL created anonymous identifiers that can be used to link birth records with PRAMS respondents.

ing the period 1991–2016. The outcomes that we measure are employment and average earnings during the first two years after birth. For the latter, we use Rhode Island birth records from 1984 to 2016 to measure subsequent maternal fertility. Specifically, we create a measure of whether a mother has had an additional child within three years following the birth of a child in our analysis sample.

4.3.8. Child medicaid enrollment and expenditures by age 2

Records from the RIDHS allow us to measure a child's enrollment and total expenditures from Medicaid. The data on enrollment and expenditures are available for the periods 1984–2016 and 1999–2015, respectively. The measure of total expenditures is based on claims for emergency, institutional, physician, and pharmacy services. In our analysis, we focus on enrollment and total expenditures during the first two years of the child's life.

4.3.9. School value-added measures

Records from RIDE and NSC allow us to measure school-level value-added for the public elementary, middle, and high schools that children attend in Rhode Island. For elementary and middle schools, we measure school-level value-added in terms of standardized math and reading test scores in RIDE test score records. For high schools, we measure school-level value-added in terms of college enrollment observed in the NSC data. Value-added for each school is estimated using all years available, excluding the students in our RD sample. To construct value-added, we estimate models which regress individual test scores or college enrollment on prior achievement or background characteristics (Kane et al., 2008; Chetty et al., 2014).²⁰ We use the school-level mean of the resulting residuals as the measure of value-added.

4.4. Descriptive statistics

Appendix Table A.1 presents summary statistics for all births in Rhode Island during our sample period (Column 1) and the RD sample of births (Column 2). Column 3 reports the *p*-value from a test of the equality of each statistic between the RD sample and all other births. The RD sample contains children who differ in terms of demographic and economic background characteristics. For example, mothers of children in the RD sample are less likely to be married or have a post-secondary degree (college or higher). Based on PRAMS survey measures, these mothers are also about 7 percentage points more likely to smoke.^{21,22} Finally, children in the RD sample come from households with greater rates of pre-birth enrollment in social assistance programs.

²⁰ Details on the specifications are provided in the notes for Appendix Table A.16.

²¹ The PRAMS survey response data is available only for a subsample of mothers.

²² Prior research suggests maternal smoking is the leading cause of low birth weight in the U.S. (Kramer, 1987).

Household earnings (based on UI records in Rhode Island) are also lower in the RD sample.²³

5. Results

5.1. Evidence on medical spending and health interventions

As discussed in Section 2, medical practitioners may use the VLBW classification for considering whether an infant receives additional medical care. Almond et al. (2010) examine the impact of crossing the 1,500 g threshold on measures of hospital care from discharge records. Their analysis finds that hospital costs increase by \$4,000 and the length of hospital stay is extended by 2 days for children just under the VLBW threshold.

Table 1 replicates the analysis of medical treatments for VLBW children using the HCUP samples described in Section 4. Specifically, we estimate the impact of crossing the 1,500 g threshold on the number of days spent in the NICU, length of stay in the hospital, and total charges. We find a statistically significant 3.4 day increase in the number of days spent in the NICU for children born just under the 1,500 g threshold. The effects on total days spent in the hospital and charges are not statistically significant, but they are positive. While the point estimates are sometimes imprecisely estimated, our results are broadly consistent with results from prior studies that find evidence of an increase in the intensity of medical care increases for children born just under the 1,500 g threshold (Almond et al., 2010; Bharadwaj et al., 2013; Daysal et al., 2019).

5.2. Evidence on identifying assumptions

Our research design relies on the assumption that all pre-birth factors vary continuously with birth weight across the 1,500 threshold for VLBW status. Appendix Table A.2 shows results from testing the validity of this assumption where we estimate threshold crossing effects for various demographic or pre-birth characteristics on the independent variables from Eq. 3. Column 2 reports estimates for the coefficient of the VLBW indicator. The results generally show no detectable discontinuities in baseline measures.

We also test our identification assumption by examining the density of births around the threshold for VLBW status. Panel B of Appendix Figure A.1 is a histogram of births between 1,300 and 1,700 g. Note that this figure includes births from the RD sample and births where the infant's weight was within three grams of the 1,500 g threshold (which we exclude from our main analysis). As in prior studies, there is pronounced heaping at the "round" gram numbers (such as multiples of 100) and at the gram equivalent of ounce intervals (Almond et al., 2010;

²³ We define household income as earnings for mothers and fathers who work in Rhode Island in the four quarters prior to birth. Earnings outside of Rhode Island are not captured in this measure. If a woman is not married, the father's earnings will also not be captured in our definition of household income.

Table 1
Impacts on Hospitalization Outcomes (HCUP Analysis)

	Mean for BW > 1,500 grams (1)	RD Est. BW < 1,500 grams (2)	Observations (3)	Birth Cohort (4)
Days in the NICU	9.925	3.427** (1.508)	1,245	2006–2015
Length of Stay	24.25	0.192 (1.818)	1,724	2002–2015
Total Charges (\$)	116,965	3,470.24 (10,793.38)	1,724	2002–2015

Notes: Column 1 reports the mean of the dependent variable for children born above 1,500 g. Column 2 provides estimates of the impact of crossing the 1,500 g threshold using Eq. 3. Standard errors clustered at the gram level are presented in parentheses. Columns 3 and 4 report the total number of observations and describe the birth cohorts included in the analysis. The sample for all results includes children born with birth weight between 1,300 and 1,700 g (excluding children born within 3 g of the 1,500 g threshold). The HCUP data for Rhode Island is available from 2002–2015. Statistical significance is denoted by *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Bharadwaj et al., 2013). There is no apparent irregular pattern of heaping near the 1,500 g threshold. We conduct a McCrary (2008) test and fail to reject the null hypothesis of continuity at the threshold (p -value = 0.924).

5.3. Main results for child outcomes

Tables 2–4 report regression results based on Eq. 3. To supplement this analysis, Fig. 1 and Appendix Figures A.2–A.4 plot means of outcome variables within 20 g bins of birth weight. For all results, we use the sample of 1,300–1,700 g births (excluding observations within three grams of the 1,500 g threshold).

5.3.1. Development and childhood education

Table 2 reports results from Eq. 3 for development and childhood education outcomes. Children below the 1,500 g threshold have 0.444 higher average (math and reading) standardized test scores measured in third grade. There are similar impacts on standardized test scores in fifth and eighth grade, suggesting that the permanent increases in child ability accrued by third grade persists through middle school. To give a clearer sense of our results, Fig. 1 (Panel A) shows means for test scores in the pooled sample (all grades 3–8 scores) in 20 g bins of birth weight. This figure shows a clear discontinuity at the 1,500 g threshold.

While we find no significant effects of threshold crossing for enrollment in SSI or special education services (IEP), the point estimates are consistent with children just under the threshold having improved outcomes. For example, the point estimates suggest that children who marginally qualify for VLBW status are 2.4 percentage points (96 percent) less likely to participate in SSI from ages 0–3. Similarly, the point estimate suggests that there is a 20.5 percentage point reduction in the years that a child has an IEP in grades 1–4.²⁴

²⁴ Appendix Table A.3 provides additional analysis of IEP outcomes. This analysis estimates impacts on an indicator for whether the child ever has an IEP for a learning disability or for a non-learning disability. We focus on IEPs associated with learning disabilities because this is the largest category of IEP that we observe in our sample of children who weigh 1,300–1,700 g at birth. We find that crossing the 1,500 g threshold from above reduces the likelihood that a child has an IEP due to a learning disability, and this result is statistically significant at the 10 percent level.

Overall, these results are similar to previous findings of significant threshold crossing impacts on standardized test scores in Chile and Norway. Bharadwaj et al. (2013) find that children in Chile who are born just under the 1,500 g threshold have 0.13 standard deviations higher test scores measured in first to eighth grade in Chile. Similarly, they find that threshold crossing has a positive impact of 0.22 standard deviations on test scores in tenth grade for children in Norway. An additional comparison for these results comes from studies of early childhood education programs such as Head Start. Duncan and Magnuson (2013) review 84 experimental and quasi-experimental evaluations of early childhood education programs and report that the average impact is 0.35 standard deviations.

5.3.2. High school and post-secondary education

Table 3 shows that crossing the threshold for VLBW status has significant and beneficial impacts on several outcomes measured in high school and early adulthood. For high school outcomes, we find benefits in terms of non-cognitive outcomes. Threshold crossing is associated with a 13-percentage point reduction in the likelihood of having a disciplinary offense in high school. This estimate represents a 72-percent decrease relative to the mean offense rate of 17.9 percent. While this result is statistically significant at the five percent level, it is worth noting that the estimated standard errors are large and imply that the true impact may be substantively smaller in magnitude.

After high school, we find notable improvements in post-secondary education enrollment. Table 3 and Panel B of Fig. 1 shows that children just under the 1,500 g threshold are 20 percentage points more likely to enroll in a 4-year college by age 22. This is a large impact given the mean enrollment rate of 30.5 percent for children born above the birth weight threshold. We find similarly large positive impacts for enrollment specifically at four-year institutions.²⁵

²⁵ While these estimates are statistically significant at conventional levels, it is worth noting that the confidence intervals associated with the impacts on college enrollment are substantively large. For example, the lower bound of the 95-percent confidence interval for any college enrollment is a 3.96 percentage point effect, an effect that is 7.4 percent of the mean of college enrollment for children born just above the 1,500 g threshold. We further discuss the precision of our estimates and implications for policy in Section 7.

Table 2
Impacts for Development and Education Outcomes

	Mean for BW > 1,500 grams (1)	RD Est. BW < 1,500 grams (2)	Observations (3)	Birth Cohort (4)
<i>SSI, IEP, and Grade Rep.</i>				
SSI Age 0	0.017	-0.008 (0.015)	1,790	1996–2015
SSI Ages 0 - 3	0.025	-0.024 (0.019)	1,577	1996–2012
Years on IEP (1–4)	1.068	-0.205 (0.259)	605	1996–2006
Grade Repetition (1–4)	0.105	-0.131*** (0.039)	608	1996–2006
<i>Avg. Std. Test Score</i>				
3rd Grade	-0.284	0.444*** (0.162)	544	1996–2005
5th Grade	-0.244	0.299* (0.159)	527	1994–2003
8th Grade	-0.176	0.314* (0.165)	485	1990–2000
All (3–8)	-0.220	0.338*** (0.128)	3,070	1990–2005

Notes: Column 1 reports the mean of the dependent variable for children born above 1,500 g. Column 2 provides estimates of the impact of crossing the 1,500 g threshold using Equation 3. Standard errors clustered at the gram level are presented in parentheses. Columns 3 and 4 report the total number of observations and describe the birth cohorts included in the analysis. The sample for all results includes children born with birth weight between 1,300 and 1,700 g (excluding children born within 3 g of the 1,500 g threshold). Statistical significance is denoted by *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table 3
Impacts for High School and Higher Education Outcomes

	Mean for BW > 1,500 grams (1)	RD Est. BW < 1,500 grams (2)	Observations (3)	Birth Cohort (4)
Disciplinary Offenses (9–12)	0.179	-0.131** (0.061)	393	1988–1998
College Preparation Index	-0.207	0.059 (0.379)	493	1993–2001
Any College Enrollment by Age 22	0.536	0.171** (0.067)	416	1984–1994
4-Year College Enrollment by Age 22	0.305	0.200** (0.077)	416	1984–1994

Notes: Column 1 reports the mean of the dependent variable for children born above 1,500 g. Column 2 provides estimates of the impact of crossing the 1,500 g threshold using Equation 3. Standard errors clustered at the gram level are presented in parentheses. Columns 3 and 4 report the total number of observations and describe the birth cohorts included in the analysis. The sample for all results includes children born with birth weight between 1,300 and 1,700 g (excluding children born within 3 g of the 1,500 g threshold). Statistical significance is denoted by *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Our estimates suggest that VLBW status can ameliorate previously documented negative impacts of low birth weight on human capital outcomes. For example, [Currie and Hyson \(1999\)](#) find that low birth weight children are more than 25 percent less likely to pass high school exit exams in Britain even after controlling for a range of background characteristics. Using data on twins from Norway and a fixed effects approach, [Black et al. \(2007\)](#) find that lower birth weight significantly reduces the likelihood of graduating from high school. Similarly, [Oreopoulos \(2008\)](#) and [Royer \(2009\)](#) provide estimates based on within family variation that show reduced weight has negative impacts on schooling.

5.3.3. Social program expenditures

We provide direct evidence on the social impact of VLBW threshold crossing by studying social program spending for children in our sample. Any realized impact on social spending can be compared to the costs associated with additional hospital care provided to children born just below the 1,500 g threshold. As noted in Section 5.1, we

find a statistically insignificant (positive) threshold crossing effect on hospital expenditures using data for Rhode Island. Given the imprecision in this estimate, it may be helpful to refer to estimates from [Almond et al. \(2010\)](#). In a larger sample of states, they find a significant \$4,000 increase in hospital costs for children born just below 1,500 g.

As outlined in Section 4, we calculate total social program expenditures (in 2015 dollars) that accrue to the child by the time they reach age 10, 12, and 14. [Table 4](#) presents estimates of the impact of crossing the VLBW threshold on the extensive (Panel A) and intensive (Panel B) margin of these measures of social expenditures. Appendix Figure A.4 provides corresponding visual results for the impact of threshold crossing for the age 10 measures. Panel A in [Table 4](#) shows no significant impacts on the likelihood of having any social program expenditures, although the point estimates are consistently negative. At the same time, Panel B shows that there are marginally significant reductions on total spending for all measures. For example, we find that social safety net expenditures by age 10

Table 4
Impacts for Social Program Expenditure by Age

	Mean for BW > 1500 grams (1)	RD Est. BW < 1500 grams (2)	Observations (3)	Birth Cohort (4)
<i>Panel A: Any Expenditures (= 1)</i>				
By Age 10	0.607	-0.057 (0.071)	951	1997–2006
By Age 12	0.587	-0.077 (0.079)	740	1997–2004
By Age 14	0.588	-0.117 (0.089)	529	1997–2002
<i>Panel B: Total Expenditures (\$)</i>				
By Age 10	91,296	-27,291* (15,736)	951	1997–2006
By Age 12	113,844	-44,067* (25,034)	740	1997–2004
By Age 14	145,486	-66,997** (33,622)	529	1997–2002
<i>Panel C: Total Expenditures (\$) Any Expenditures (= 1)</i>				
By Age 10	150,334	-29,149 (22,832)	572	1997–2006
By Age 12	194,052	-53,444 (33,486)	434	1997–2004
By Age 14	247,515	-74,075* (37,596)	310	1997–2002

Notes: Column 1 reports the mean of the dependent variable for children born above 1,500 g. Column 2 provides estimates of the impact of crossing the 1,500 g threshold using Eq. 3. Standard errors clustered at the gram level are presented in parentheses. Columns 3 and 4 report the total number of observations and describe the birth cohorts included in the analysis. The sample for all results includes children born with birth weight between 1,300 and 1,700 g (excluding children born within 3 g of the 1,500 g threshold). Statistical significance is denoted by *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

are \$27,291 lower (p -value <0.10) for children who are born just to the left of the 1,500 g threshold, and \$66,997 lower (p -value <0.05) by the age of 14. This suggests a sizeable direct public return relative to the \$4,000 spent on additional hospital costs.²⁶

5.4. Robustness tests

We conduct several robustness tests to examine whether our results are sensitive to different choices for the RD analysis. Appendix Tables A.5–A.7 report the results associated with our main test, which is focused on examining sensitivity for all outcomes by using quadratic RD specifications and changing the bandwidth for births included in the analysis. For the analysis of bandwidth sensitivity, we use the optimal bandwidth selection method from Calonico et al. (2014) for each outcome. These results are generally robust to these changes: coefficient signs and magnitudes remain similar, although estimates sometimes lose statistical significance.²⁷

Next, we examine whether our findings are sensitive to varying the window used to exclude observations around the 1,500 g threshold or including additional fixed effects. Both approaches address concerns over non-random heaping. The second and third rows of Appendix Table A.9 show

that we find similar results when excluding observations that are within 5 and 10 g of the threshold.²⁸ The last row reports estimates based on an augmented version of Eq. 3 that includes fixed effects for observations at each 10 g interval of birth weight. The point estimates remain statistically significant and are generally similar in magnitude when we include 10 g fixed effects.²⁹

Finally, we investigate the sensitivity of our results to different choices in clustering. As discussed in Section 3, our main specification clusters at the gram-level. Appendix Table A.10 shows results when we cluster all observations within 10 g, 20 g, and 50 g bins. The results show that clustering at higher levels generally does not affect the standard errors or the statistical significance of our results.

5.5. Alternative birth weight thresholds

We also examine whether we observe discontinuities in outcomes at other intervals of 100 g. A pattern of significant threshold crossing effects at 100 g intervals of birth weight that exceed 1,500 g would generate concern that heaping is driving our main results. The top row of Appendix Table A.12 reproduces the 1,500 g threshold crossing estimates on selected outcomes. The remaining rows provide esti-

²⁶ Appendix Table A.4 reports threshold crossing impacts for participation and expenditures for each program used to construct the measure of total expenditures up to age 14. For each program, the estimated impacts of crossing the 1,500 g threshold are generally not statistically significant, but the point estimates are consistently negative.

²⁷ Appendix Table A.8 further tests for sensitivity based on the choice of bandwidth. Specifically, we provide additional results based using a sample of births from 1,200–1,800 g and 1,100–1,900 g. These results are similar to the results from our preferred bandwidth.

²⁸ One result in Appendix Table A.9 differs from our main estimates. In Column 5, we see that the threshold crossing estimate for total social expenditures by age 14 becomes slightly less negative and is no longer statistically significant when we exclude the observations that are within 10 g of the 1,500 g threshold.

²⁹ An additional robustness test assesses whether we obtain similar impacts when we focus only on the cohorts for the test score analysis that best correspond to the period for which we have HCUP data from Rhode Island. Appendix Table A.11 reports results from this analysis. The estimates are similar to those in Table 2.

mates for every 100 g cutoff between 1,600 and 3,000 g. There is no consistent pattern of threshold crossing effects on test scores or other selected outcomes for any cutoff above 1,500 g.

5.6. Multiple hypothesis testing

Another potential concern is that our findings could be an artifact of multiple hypothesis testing. To assess this, we have followed the recommended practice of adjusting per comparison p -values (Anderson, 2008). Specifically, we use the two-step procedure from Benjamini et al. (2006) to calculate “ q -values” that control for the false discovery rate (FDR), which is the proportion of rejections that are false positives (Type I errors). When we compute the adjusted p -values, we find that our results for test scores, disciplinary offenses, college enrollment and total social expenditures are significant at the 10-percent level or lower after adjusting for the fact that we analyzed multiple outcomes in Tables 2 and 3 and 4.³⁰ Note that we compute adjusted p -values accounting for the fact that we estimate effects on the following 10 outcomes: SSI enrollment at age 0, SSI enrollment at ages 0–3, IEP participation in grades 1–4, repetition in grades 1–4, test scores in grades 3–8, disciplinary offenses, the college preparation index, enrollment in any college, enrollment in a 4-year college, and total social expenditures by age 14.

5.7. Attrition analysis

A final concern is that we will not observe outcomes in the test score and high-school enrollment samples if children do not attend a public school in Rhode Island.³¹ We would be concerned if attrition from our analysis samples is related to the 1,500 g threshold for VLBW status. Appendix Table A.13 tests for attrition by estimating threshold crossing impacts on the likelihood that students have test score records or were enrolled in a public high school. The only statistically significant results is a negative point estimate for eighth grade (p -value<0.10). The fact that we generally find insignificant impacts on attrition in high school suggests differential attrition is not the primary factor driving the results on educational attainment or social program expenditures.

5.8. Heterogeneity

Appendix Table A.14 reports results for education and social expenditure outcomes after dividing the RD sample based on child sex (as reported in birth records). Panels A and B compare results for girls and boys, respectively. In most cases, the subgroup-specific effects are statistically significant, and the point estimates are similar to those

³⁰ Based on the adjusted p -values, the result for grade repetition in grades 1–4 is significant at the one percent level, the results for test scores in grades 3–8, disciplinary offenses, any college enrollment are significant at the five percent level, and the results for enrollment in a 4-year college and total social expenditures are significant at the 10 percent level.

³¹ Recall that once a child reaches high school, we will see them in the NSC data regardless of whether the child moves from Rhode Island.

from Tables 2–4. We cannot reject the hypothesis that estimates for each outcome are equal for girls and boys.

6. Mechanisms

As discussed in Section 3, the potential for parental responses to child health complicates the interpretation of the reduced form estimates of the impact of crossing the 1,500 g threshold. Parents, teachers, and the broader community may respond to a child’s health status, thereby amplifying or reducing the direct impact of increased early-life medical interventions for VLBW children. Existing empirical studies provide mixed evidence as to whether we should expect to find evidence of parental responses in our setting. Datar et al. (2010) use a family fixed effects approach and find that parents report investing more resources into children that have heavier birth weight. In contrast to these results, Bharadwaj et al. (2013) examine data from Chile and find that time spent reading with a child does not change as a function of crossing the VLBW threshold. They also study whether parents respond to child health through school choice, finding no evidence that children just under the 1,500 g threshold attended elementary schools with higher or lower average test scores.³²

Table 5 shows threshold crossing impacts on direct measures of parental responses and other factors that could plausibly mediate the impacts in our sample.³³ We first study indices for maternal care and stress based on PRAMS responses.³⁴ The care index is the number of correct responses to questions about infant care such as sleeping position. The stress index is the number of affirmative responses to questions about depression and anxiety. The results show that we find statistically insignificant impacts, which is perhaps not surprising given the relatively small sample sizes available for the PRAMS measures. That said, the point estimates for the care and stress indices are positive (indicating better care) and negative (indicating less stress), respectively.

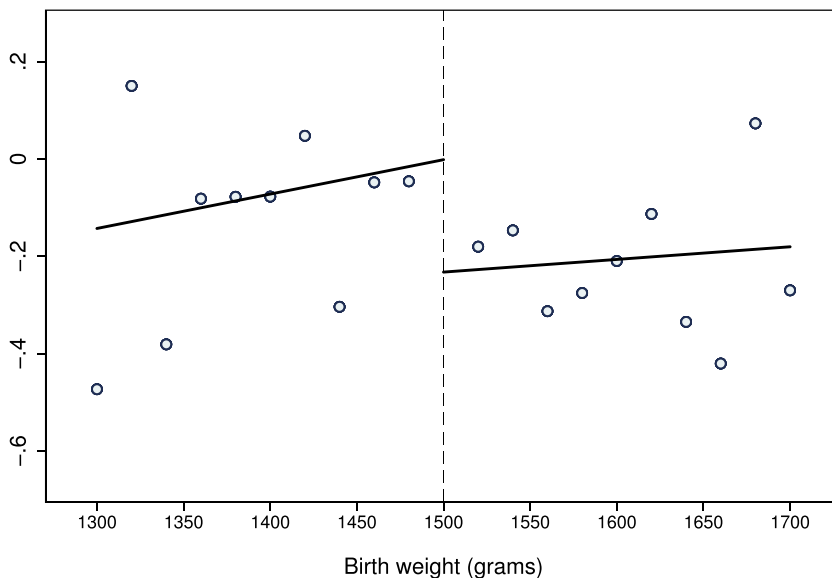
Next, we study maternal labor market outcomes using earnings records. The effect of crossing the 1,500 g threshold could be positive if mothers work more due to the impact of additional in-hospital care on child health. Yet, there could be negative impacts if mothers respond to health improvements by investing more time with their child. Table 5 shows no significant impacts on employment or mean annual earnings measured during the first two years postpartum. The standard errors are sufficiently

³² Note that the results from Datar et al. (2010) may be less relevant for understanding effects in our context since they focus on comparing children born with higher birth weight. In contrast, the analysis by Bharadwaj et al. (2013) may be more relevant since they also focus on VLBW children.

³³ Appendix Figure A.5 plot means of the early childhood investment outcomes within 20 g bins of birth weight. Appendix Table A.15 provides results from robustness tests to test whether the findings from Table 5 are sensitive to different choices for the RD analysis.

³⁴ While all mothers with low birth weight (less than 2,500 grams) receive the PRAMS survey, only a fraction respond. The PRAMS response rate in our main analysis sample of children who weigh 1,300–1,700 grams at birth is 70 percent. Note that we find no evidence that crossing the 1,500 g threshold has any impact on the likelihood of participating in the PRAMS survey.

A. Test Scores (3-8)



B. 4-Year College Enrollment by 22

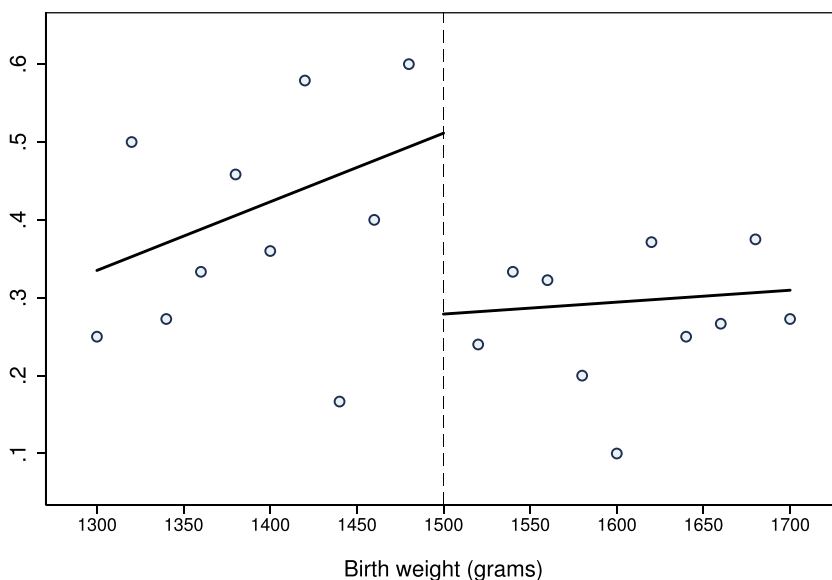


Fig. 1. Impacts for Selected Education Outcomes.

Notes: Each panel shows the relationship between birth weight and a selected education outcome. Dots represent means within 20 g bins of the running variable. The dark lines are predictions from a linear model using the individual-level data.

large that the confidence intervals include substantially large positive and negative effects.

As a measure of early-life child health, we study Medicaid enrollment and expenditures from birth up to age 2. Impacts on either outcome could signal increased parental involvement in health. At the same time, we also expect that increased medical care at birth should reduce post-birth medical needs. In [Table 5](#), we find no statistically

significant effect on either measure of Medicaid program use. Although imprecise, the point estimate for expenditures is negative and represents a reduction of about 50 percent relative to the mean for children just above 1,500 g at birth.

Another parental response that could affect child outcomes is subsequent maternal fertility. [Table 5](#) shows that mothers of children born just under 1,500 g are about 6.4

Table 5
Impacts for Early Childhood Investment Outcomes

	Mean for BW > 1,500 grams (1)	RD Est. BW < 1,500 grams (2)	Observations (3)	Birth Cohort (4)
Maternal Care Index	2.965	0.121 (0.160)	429	2002–2014
Maternal Stress Index	3.498	−0.156 (0.338)	429	2004–2014
Maternal Employment Ages 0–2 (= 1)	0.588	−0.013 (0.044)	1,796	1991–2014
Avg. Maternal Earn. Ages 0–2 (\$)	12,355	2,053 (2,265)	1,796	1991–2014
Medicaid Enrollment Ages 0–2 (= 1)	0.455	−0.035 (0.056)	2,627	1989–2014
Medicaid Exp. Ages 0–2 (\$)	14,171	−7,528 (6,518)	1,360	1984–2016
Birth Within 3 Years (= 1)	0.132	0.064** (0.028)	2,153	1989–2014

Notes: Column 1 reports the mean of the dependent variable for children born above 1,500 g. Column 2 provides estimates of the impact of crossing the 1,500 g threshold using Eq. 3. Standard errors clustered at the gram level are presented in parentheses. Columns 3 and 4 report the total number of observations and describe the birth cohorts included in the analysis. The sample for all results includes children born with birth weight between 1,300 and 1,700 g (excluding children born within 3 g of the 1,500 g threshold). Statistical significance is denoted by *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

percentage points more likely to have a subsequent child relative to children born just above the cutoff. This estimate represents a relatively large effect size (compared to the mean fertility rate above the threshold).³⁵

Finally, Appendix Table A.16 tests whether parents of children born just under 1,500 g are more likely to send their children to higher-quality schools.³⁶ Prior research suggests that there is a strong link between school quality and child achievement (Hastings and Weinstein, 2008; Hoxby and Murarka, 2009; Abdulkadiroğlu et al., 2011; Pop-Eleches and Urquiola, 2013; Deming et al., 2014). We find statistically insignificant and small point estimates for school quality measured in elementary, middle and high school.

7. Conclusion

This paper uses comprehensive administrative data to provide new evidence on the impacts of providing additional medical and postnatal investments for children with low birth weight. We estimate causal effects using an RD approach based on the 1,500 g threshold for VLBW status. Prior studies by Almond et al. (2010) and Bharadwaj et al. (2013) document that physicians and hospitals make decisions about medical care using this threshold.

We find statistically significant, large and positive impacts of crossing the 1,500 g threshold on academic achievement, college enrollment, and reliance on social programs. Children just under the threshold for VLBW status have 0.34 standard deviations higher test scores in grades 3–8. They are also 32 percent more likely to enroll in college and receive \$66,997 less in social program expenditures by age 14. When we analyze potential mediators of

these impacts, the findings suggest there is a large role for direct effects of additional medical care. We find no statistically significant impacts of threshold crossing on measures of parental responses such as maternal stress or labor market outcomes. At the 1,500 g threshold, there are also no detectable changes in the quality of schools that children attend during childhood.

Overall, this paper suggests there are private and public benefits to investing in health for low birth weight children. However, there are two important caveats to our results. First, while we find statistically significant impacts on important outcomes, the standard errors associated with our results are often large. This lack of precision implies that the true benefits may be notably smaller than the point estimates that we report. Future research based on larger samples may be able to quantify effects more precisely. Until then, cost-benefit analysis should use our estimates cautiously. Second, it is worth noting that our results are particularly relevant to children near the VLBW threshold. More work must be done to determine whether the results that we find generalize to other groups of low birth weight children.

CRedit authorship contribution statement

Eric Chyn: Conceptualization, Methodology, Writing - review & editing, Supervision, Investigation, Formal analysis. **Samantha Gold:** Conceptualization, Methodology, Writing - review & editing, Supervision, Investigation, Formal analysis, Data curation. **Justine Hastings:** Conceptualization, Methodology, Writing - review & editing, Supervision, Investigation, Formal analysis.

Declarations of Competing Interest

None.

³⁵ Appendix Table A.15 shows the point estimate is positive but smaller when we use the optimal bandwidth selected using the method from Calonico et al. (2014).

³⁶ Appendix Table A.17 provides results from robustness tests to test whether the findings from Table A.16 are sensitive to different choices for the RD analysis.

Acknowledgements

We thank the team at Research for Improving People's Lives (RIPL) and the State of Rhode Island for making this project possible. We are also thankful for helpful comments that we received from Prashant Bharadwaj and Amanda Kowalski. Finally, we are grateful for financial support from the Smith Richardson Foundation and the Laura and John Arnold Foundation.

Appendix C. Supplementary data

Supplementary material related to this article can be found, in the online version, at <https://doi.org/10.1016/j.jhealeco.2020.102400>.

References

- Abdulkadiroğlu, Atila, Angrist, Joshua D., Dynarski, Susan M., Kane, Thomas J., Pathak, Parag A., 2011. *Accountability and flexibility in public schools: evidence from Boston's Charter and Pilots*. *Q. J. Econ.* 126 (2), 699–748.
- Almond, Douglas, Currie, Janet, 2011. *Killing me softly: the fetal origins hypothesis*. *J. Econ. Perspect.* 25 (3), 153–172.
- Almond, Douglas, Doyle, Joseph J., Kowalski, Amanda E., Williams, Heidi, 2010. *Estimating marginal returns to medical care: evidence from at-risk newborn*. *Q. J. Econ.* 125 (2), 591–634.
- American Academic of Pediatrics, Committee on Fetus and Newborn, 2012. *Levels of neonatal care*. *Pediatrics* 130 (3), 587–597.
- Anderson, Michael L., 2008. *Multiple inference and gender differences in the effects of early intervention: a reevaluation of the abecedarian, perry preschool, and early training projects*. *J. Am. Stat. Assoc.* 103 (484), 1481–1495.
- Angert, Robert, Adam, Henry, 2009. *Care of the very low-birthweight infant*. *Pediatr. Rev.* 30 (1), 32–35 (January 2009).
- Bald, Anthony, Chyn, Eric, Hastings, Justine, Machelett, Margarita, 2019. *The causal impact of removing children from abusive and neglectful homes*. In: NBER Working Paper #25419. National Bureau of Economic Research.
- Barreca, Alan I., Guldi, Melanie, Lindo, Jason M., Waddell, Glen R., 2011. *Saving babies? Revisiting the effect of very low birth weight classification*. *Q. J. Econ.* 126 (4), 2117–2123.
- Benjamini, Yoav, Krieger, Abba M., Yekutieli, Daniel, 2006. *Adaptive linear step-up procedures that control the false discovery rate*. *Biometrika* 93 (3), 491–507.
- Bharadwaj, Prashant, Loken, Katrine Veelsen, Neilson, Christopher, 2013. *Early life health interventions and academic achievement*. *Am. Econ. Rev.* 103 (5), 1862–1891.
- Black, Sandra E., Devereux, Paul J., Salvanes, Kjell G., 2007. *From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes*. *Q. J. Econ.* 122 (1), 409–439.
- California Perinatal Quality Care Collaborative, 2018. *High Risk Infant Follow-up*, Accessed August 24, 2018 <https://www.cpqcc.org/perinatal-programs/high-risk-infant-follow-up>.
- Calonico, Sebastian, Cattaneo, Matias D., Titiunik, Rocio, et al., 2014. *Robust data-driven inference in the regression-discontinuity design*. *Stata J.* 14 (4), 909–946.
- Chetty, Raj, Friedman, John N., Hilger, Nathaniel, Saez, Emmanuel, Schanzenbach, Diane Whitmore, Yagan, Danny, 2011. *How does your kindergarten classroom affect your earnings? Evidence from Project STAR*. *Q. J. Econ.* 126 (4), 1593–1660.
- Chetty, Raj, Friedman, John N., Rockoff, Jonah E., 2014. *Measuring the impacts of teachers I: evaluating Bias in teacher value-added estimates*. *Am. Econ. Rev.* 104 (9), 2593–2632.
- Chetty, Raj, Hendren, Nathaniel, Katz, Lawrence, 2016. *The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity project*. *Am. Econ. Rev.* 106 (4), 855–902.
- Chyn, Eric., 2018. *Moved to opportunity: the long-run effects of public housing demolition on children*. *Am. Econ. Rev.* 108 (10), 3028–3056.
- Currie, Janet, 2011. *Inequality at birth: some causes and consequences*. *Am. Econ. Rev.* 101 (3), 1–22.
- Currie, Janet, Hyson, Rosemary, 1999. *Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight*. *Am. Econ. Rev.* 89 (2), 245–250.
- Currie, Janet, Thomas, Duncan, 1995. *Does Head Start Make a Difference?* *Am. Econ. Rev.* 85 (3), 341–364.
- Datar, Ashlesha, Kilburn, Rebecca M., Lohran, David S., 2010. *Endowments and parental investments in infancy and early childhood*. *Demography* 47 (1), 145–162.
- Daysal, N., Meltem, Mircea, 2019. *Trandafir, Marianne Simonsen, and Sanni Breining*. In: *Spillover Effects of Early-Life Medical Interventions*, Manuscript.
- Deming, David J., Hastings, Justine S., Kane, Thomas J., Staiger, Douglas O., 2014. *School choice, school quality, and postsecondary attainment*. *Am. Econ. Rev.* 104 (3), 991–1013.
- Duncan, Greg J., Magnuson, Katherine, 2013. *Investing in preschool programs*. *J. Econ. Perspect.* 27 (2), 109–132.
- Guldi, Melanie, Hawkins, Amelia, Hemmeter, Jeffrey, Schmidt, Lucie, 2018. *Supplemental security income and child outcomes: evidence from birth weight eligibility cutoffs*. In: NBER Working Paper #24913. National Bureau of Economic Research.
- Harrison, Wade, Goodman, David, 2015. *Epidemiologic trends in neonatal intensive care*. *JAMA Pediatr.* 169 (9), 855–862.
- Hastings, Justine S., 2019. *Fact-based policy: how do state and local governments accomplish it?* In: *Research Improving People's Lives*. Brookings Institute.
- Hastings, Justine S., Weinstein, Jeffrey M., 2008. *Information, school choice, and academic achievement: evidence from two experiments*. *Q. J. Econ.* 123 (4), 1373–1414.
- Hastings, Justine S., Howison, Mark, Lawless, Ted, Ucles, John, White, Preston, 2019. *Unlocking data to improve public policy*. In: *Communications of the ACM*, Forthcoming.
- Havnes, Tarjei, Mogstad, Magne, 2011. *No child left behind: subsidized child care and children's long-run outcomes*. *Am. Econ. J. Econ. Policy* 3 (2), 97–129.
- Heckman, James, Pinto, Rodrigo, Savelyev, Peter, 2013. *Understanding the mechanisms through which an influential early childhood program boosted adult outcomes*. *Am. Econ. Rev.* 103 (6), 2052–2086.
- Hoxby, Caroline M., Murarka, Sonali, 2009. *Charter schools in New York City: who enrolls and how they affect their students' achievement*. In: NBER Working Paper #14852. National Bureau of Economic Research.
- Hoynes, Hilary, Schanzenbach, Diane Whitmore, Almond, Douglas, 2016. *Long-Run Impacts of Childhood Access to the Safety Net*. *Am. Econ. Rev.* 106 (4), 903–934.
- Kane, Thomas J., Staiger, Douglas O., 2008. *Estimating teacher impacts on student achievement: an experimental evaluation*. In: NBER Working Paper #14607. National Bureau of Economic Research.
- Kramer, Michael S., 1987. *Intrauterine Growth and Gestational Duration Determinants*. *Pediatrics* 80 (4), 502–511.
- Lee, David S., Card, David, 2008. *Regression discontinuity inference with specification error*. *J. Econom.* 142 (2), 655–674.
- McCrary, Justin., 2008. *Manipulation of the running variable in the regression discontinuity design: a density test*. *J. Econ.* 142 (2), 698–714.
- Oreopoulos, Philip, Stabile, Mark, Wald, Randy, Roos, Leslie L., 2008. *Short-, medium-, and long-term consequences of poor infant health: an analysis using siblings and twins*. *J. Hum. Resour.* 43 (1), 88–138.
- Pop-Eleches, Cristian, Urquiola, Miguel, 2013. *Going to a better school: effects and behavioral responses*. *Am. Econ. Rev.* 103 (4), 1289–1324.
- Rhode Island Department of Education, 2013. *Annual Per Pupil Expenditure Report*.
- Royer, Heather., 2009. *Separated at birth: US twin estimates of the effects of birth weight*. *Am. Econ. J. Appl. Econ.* 1 (1), 49–85.
- SSA, 2018a. *100.00 Low Birth Weight and Failure to Thrive - Childhood*, Retrieved March 2019 from <https://www.ssa.gov/disability/professionals/bluebook/100.00-GrowthImpairment-Childhood.htm#100.04B>.
- SSA, 2018b. *Understanding Supplemental Security Income SSI for Children*, 2018 edition, Retrieved March 2009 from <https://www.ssa.gov/ssi/text-child-ussi.htm>.
- U.S. Department of Education, 2015. *Local Education Agency Universe Survey*.
- Vohr, Betty, Stephens, Bonnie, Tucker, Richard, 2010. *35 Years of Neonatal Follow-up in Rhode Island*. *Med. Health R. I.* 93 (5), 151–253.