American Economic Review 2013, 103(5): 1862–1891
http://dx.doi.org/10.1257/aer.103.5.1862

Early Life Health Interventions and Academic Achievement[†]

By Prashant Bharadwaj, Katrine Vellesen Løken, and Christopher Neilson*

This paper studies the effect of improved early life health care on mortality and long-run academic achievement in school. We use the idea that medical treatments often follow rules of thumb for assigning care to patients, such as the classification of Very Low Birth Weight (VLBW), which assigns infants special care at a specific birth weight cutoff. Using detailed administrative data on schooling and birth records from Chile and Norway, we establish that children who receive extra medical care at birth have lower mortality rates and higher test scores and grades in school. These gains are in the order of 0.15–0.22 standard deviations. (*JEL* I11, I12, I18, I21, J13, O15)

This paper studies the effect of improved neonatal and early childhood health care on mortality and long-run academic achievement in school. Using administrative data on vital statistics and education records from Chile and Norway, we provide evidence on both the short- and long-run effectiveness of early life health interventions. The question of whether such interventions affect outcomes later in life is of immense importance for policy not only due to the significant efforts currently being made to improve early childhood health world wide, but also due to large disparities in neonatal and infant health care that remain between (and within) countries.[1] While the stated goal of many such interventions is to improve childhood health and reduce mortality, understanding spillovers and other long-run effects such as better academic achievement is key to estimating their efficacy.

Beyond the immediate policy relevance of this question, examining the role of early life health interventions in explaining academic achievement is also important

* Bharadwaj: University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92629 (e-mail: prbharadwaj@ucsd.edu); Løken: University of Bergen, Postboks 7802, 5020 Bergen Norway (e-mail:katrine.loken@econ.uib.no); Neilson: Yale University, 37 Hillhouse Avenue, New Haven, CT 06211 (e-mail:christopher.neilson@yale.edu). The authors wish to thank Danuta Rajs, Victoria Garrido, Josette Iribarne, and Daily Piedra from the Departamento de Estadisticas e Informacion de Salud del Ministerio de Salud (MINSAL) and Francisco Lagos of the Ministry of Education (MINEDUC) of the government of Chile for facilitating joint work between government agencies that produced the data from Chile used in this study. They also thank the Norwegian Research council for financial support and the Norwegian Institute of Public Health and Statistics Norway for providing access to the different data sources. In addition, they are grateful for support from the labor project at NHH funded by the Norwegian research council and to the Medical Birth Registry of Norway for providing the birth records data. Finally, the authors also wish to thank Douglas Almond, Joseph Altonji, Gordon Dahl, Joseph Doyle, Fabian Duarte, Hilary Hoynes, Amanda Kowalski, Costas Meghir, Karthik Muralidharan, Kjell Salvanes, Seth Zimmerman, and participants at the NBER Children's Program meeting and the All California Labor Conference for comments and suggestions. The authors have no relevant or financial interests related to this project to disclose.

[†] Go to http://dx.doi.org/10.1257/aer.103.5.1862 to visit the article page for additional materials and author disclosure statement(s).

[1] World Health Report (2005) documents the persistent gaps in provision of care which consequently leads to largely avoidable deaths of over 4 million babies before they reach the age of 28 days and half a million mothers at childbirth.

because it highlights the role of health and social policy more generally in the education production function. The recent literature on educational production functions tends to find that a large part of the variation in educational outcomes is explained by students' individual "initial conditions" (Almond and Currie 2011; Heckman and Masterov 2007). Successful early life health interventions would suggest that initial conditions of students are not only a function of family and individual choices, but also of public policies such as health care.[2] As we show in this paper, the fact that treatments soon after birth make a difference for schooling outcomes later on suggests that the observed heterogeneity in educational outcomes can in part be explained by heterogeneity in health care beginning at birth. By focusing on the role of health care policy, such as the introduction of standardized neonatal care in Chile or the widespread use of surfactant in Norway starting in the 1990s, we underscore the importance of early life health care as a way to improve test scores and potentially lower inequalities in achievement.

A growing literature in economics suggests that interventions during early childhood matter for later life outcomes. Papers have examined the role of child care (Havnes and Mogstad 2011), pre-school and kindergarten related interventions (Heckman et al. 2010; Chetty et al. 2011) and welfare programs (Almond, Hoynes, and Schanzenbach 2010; Currie 2006) in determining later life economic outcomes. While the literature on health and education has documented the effects of several contemporaneous health interventions and their impact on educational outcomes,[3] there are few studies in economics that causally link early childhood health interventions to academic performance later in life.[4] One example is Field, Robles, and Torero (2009), who present evidence that children born to mothers subjected to an iodine supplement program while pregnant complete more years of schooling in Tanzania. A recent working paper by Chay, Guryan, and Mazumder (2009) relates the narrowing of the black-white test score gap in the United States to improved health access for blacks during infancy after the Civil Rights Act. We contribute to this literature by providing causal evidence on the effect of improved neonatal health care on mortality and academic achievement using administrative data from two countries.

The challenge in examining the causal effect of health interventions is that they are generally not administered randomly. Hence, infants who receive special medical attention may differ along various other dimensions that affect mortality and school performance. To get around such confounding factors, we take advantage of rules and recommendations for administering medical care to children who are born with Very Low Birth Weight status (VLBW—birth weight less than 1,500 grams).

[2] For example, Hoynes, Page, and Stevens (2011) find that WIC programs led to better birth outcomes. An excellent reference on this is Currie (2006) where examples from many well known public safety net programs and their impact on child well-being is discussed.

[3] A small sampling of these studies are Miguel and Kremer (2004); Bleakley (2007); Behrman (1996); and Glewwe, Jacoby, and King (2001). In the seminal work on educational externalities of health interventions by Miguel and Kremer (2004), the intervention examined is contemporaneous with school outcomes.

[4] We differentiate ourselves from the literature examining the role of early childhood shocks in utero or otherwise, (see for example Maccini and Yang 2009) because while we might know that endowments or shocks matter for later life outcomes, this does not imply that treatments can remedy those assaults. Our paper is concerned with understanding whether treatments matter for long-run outcomes. Several papers that document the importance of early childhood health and later life outcomes are Black, Devereux, and Salvanes (2007); Currie (2011); and Conti, Heckman, and Urzúa (2010).

Following Almond et al. (2010), the underlying assumption is that an infant born with a birth weight of 1,490 grams is essentially identical to an infant born with a birth weight of 1,510 grams, except for the extra medical attention that the lower birth weight infant might receive. At these close margins, the role of confounding factors is mitigated and inference can be carried out at least locally via a regression discontinuity design.

Rules and recommendations regarding VLBW births appear to be quite salient in many countries. In guidelines published by the Ministry of Health in Chile, the medical recommendations for children born below 1,500 grams (or below 32 weeks of gestation) are explicitly stated and eligibility for several publicly funded treatments is determined by birth weight and gestational age. In Norway, a survey of 19 of the largest neonatal units revealed such cutoffs to be one of the main criteria for assigning care (Skranes, Skranes, and Skranes 2000). We focus in particular on the birth weight cutoff, which is measured at the gram interval in both Chile and Norway, and compare children just under and over 1,500 grams to examine differences in outcomes as a result of extra medical treatments.

Results from both countries strongly support the idea that children below the 1,500 gram cutoff receive extra medical attention and that this results in significantly lower mortality and better performance in school. In Chile, children born just below the cutoff have around 4.4 percentage points lower infant mortality (death within one year of birth). While slightly smaller in magnitude, we find statistically significant effects on mortality in Norway as well. Following surviving children through school from first to eighth grade in Chile, we find that those born just below the cutoff perform 0.15 standard deviations (SD) better in math than children born just above the cutoff. In Norway we find a slightly larger effect of 0.22 SD using national exams taken in tenth grade.[5] In both countries, we are able to examine a specific policy initiative of administering surfactant therapy to newborns. Using the timing of the policy together with the regression discontinuity framework described above, we find suggestive evidence that the introduction of this treatment augmented the effect of being just below the cutoff, lowering mortality and raising academic outcomes even more.

Our results are robust to standard regression discontinuity checks and additional checks relevant to the cases with potential nonrandom heaping at certain round integer values.[6] We also have a unique internal check to ensure that our results are not driven by nonrandom heaping at or around 1,500 grams. As mentioned earlier, the rules and recommendations in Chile (and to a large extent in Norway as well) explicitly mention a 32 week gestational rule: all children (regardless of birth weight) below 32 weeks of gestation are eligible for treatments. If heaping or rounding associated with socioeconomic characteristics were an important driver of the results, we would expect to find this to be true for the sample below 32 weeks in age as well as above. However, we find that birth weight cutoffs play no role in determining

[5] The sample of children observed in school is a selected sample of children who survive. In Section IVC we address the extent to which this results in bias for our results on educational achievement. Our results suggest that survival bias does not play an important role here.

[6] This is particularly a problem when birth weight is measured in grams as well as ounces (Umbach 2000; Barreca et al. 2011). However, in both Chile and Norway, birth weight is always measured in grams which helps mitigate some of the problems identified in this literature. We explore these issues in detail in Section V and in the online Appendices.

mortality or test scores for children who were born with less than 32 weeks of gestation. We do not use gestational age itself in a regression discontinuity framework as this could be a choice variable, driven by doctor or hospital characteristics/quality. Conditional on gestational age, however, birth weight should not be manipulable.

I. VLBW Births in Chile and Norway

Health care in Chile is primarily funded by the public system which consists of 29 regions, each which has at least one hospital equipped for providing specialized care to VLBW infants (and other infants who need advanced care) in a Neonatal Intensive Care Unit (NICU). In 1991 a national committee of Chilean neonatologists set uniform standards for care and equipment at all NICUs in the country. Gonzalez et al. (2006, p. e951) point out that, "A protocol has been implemented at the national level to regulate the referral of neonates who are born in hospitals without a NICU to the regional hospitals. There also are *standardized protocols for the treatment of newborns who weigh less than 1,500g* and for cases of respiratory distress syndrome" (emphasis added). Between 1992 and 2000, 99 percent of births occur under skilled care (doctor or midwife), approximately 68 percent of births occur in hospitals with a NICU, and the number of NICUs in the country did not change.[7]

Publications by the Ministry of Health in Chile list the numerous medical recommendations to be administered to children who are born with a weight of less than 1,500 grams and/or less than 32 weeks in gestational age.[8] One of the most well known programs introduced for VLBW births in Chile was the national surfactant program which began in 1998. Under this program artificial lung surfactant is used to treat respiratory distress syndrome in VLBW infants. Many public health articles on Chile's infant and neonatal mortality credit this program with reducing mortality rates among VLBW infants in Chile (e.g., Gonzalez et al. 2006 and Jiménez and Romero 2007).[9] Several public neonatal health care programs that were introduced later went even further and not only recommended treatments for births under the cutoff but made VLBW status an explicit requirement for program eligibility. For example, *PNAC prematuro* is a program introduced in 2003 which provides specialized nutritional supplements and has its eligibility determined exclusively by the cutoff birth weight and gestational age. A larger public health care expansion introduced in 2005, called *AUGE*, provided additional neonatal examinations and treatments determined again by the same cutoffs mentioned above.[10]

In Norway, prematurity is defined as births of birth weight below 2,500 grams or less than 37 weeks of gestational age. This category is again divided into subgroups which follow the WHO recommendations of very low birth weight (VLBW) of less

[7] For a review of neonatal care in Chile, its implementation during the 1990s in Chile, and evaluation in the public health literature see Gonzalez et al. (2006) and Palomino, Morgues, and Martínez (2005).

[8] A website maintained by the Committee of Neonatologists in Chile provides extensive information and recommendations for the care of premature births (www.prematuros.cl).

[9] A manual with recommendations on how to treat and monitor premature births was published in 1999 with the title including the 1,500 gram cutoff and 32 week gestational period again signaling the importance of the cutoff. This is available in PDF form from the authors.

[10] These include (i) screening for Retinopathy of Prematurity (ROP), which helps avoid blindness; (ii) screening and follow-up treatment for Sensorineural Hearing Loss (SHL); and (iii) treatment for Bronchopulmonary Dysplasia (BPD), which is a chronic lung disease common in VLBW births.

than 1,500 grams or less than 32 weeks of gestational age and extremely low birth weight (ELBW) of less than 1,000 grams or less than 28 weeks of gestational age (Markestad and Halvorsen 2007). Bratlid and Nordermoen (2010) provide a 40 year overview of the treatment for VLBW children in Norway and give evidence that the VLBW cutoff was important from the 1980s and onward.

The specific recommendations regarding VLBW births begin to appear in documents in the 1980s, several of which specifically state the cutoffs mentioned above (Meberg 1988; Finne et al. 1988).[11] Several recent studies provide direct evidence on the practices in Norwegian neonatal wards. Bratlid and Nordermoen (2010) report that only 14 percent of children born below 32 weeks of gestational age in 1970 received respiratory treatment and only half of them survived; however, by the 1980s these treatments had become more commonplace. At the end of the 1980s 75 percent of children born below 32 weeks of gestational age or below 1,500 grams received respiratory treatment and beginning in 1989, surfactant became common practice in the care of VLBW children in all hospitals in Norway (Saugstad 2010). Skranes, Skranes, and Skranes (2000) surveyed all the main neonatal wards in Norway and *all* hospitals that responded to the survey listed less than 1,500 grams as their main indicator for having children in extra treatment and follow-up programs. While other factors also determine care, VLBW is the only one common across all neonatal wards. Similar to Chile and the United States, there are numerous medical publications that recommend treatments for children less than 1,500 grams and/or less than 32 weeks (Metodebok i nyfødtmedisin 1998).

These policies and recommendations show a general trend in which the medical community in Chile and Norway give special importance to the births below the weight of 1,500 grams. In sum, it appears that the "rules of thumb," as mentioned in Almond et al (2010), are very much present in the Chilean and Norwegian context. In Section V, using hospital level data from both countries, we directly provide evidence for discontinuity in treatments around 1,500 grams for children greater than 32 weeks of gestational age.

II. Economic and Empirical Framework

We model birth weight BW_i of an individual i as a noisy signal of initial health at birth H_i, which is unobserved to the econometrician. D_i represent the collection of hospital inputs that newborns receive at hospitals. These treatments are assumed to depend on a decreasing function of health at birth, $g(H_i)$, and a random component v_i. However, due to the behavior of midwives, doctors, and clinics regarding the needs of very low weight births, there is a discontinuous break in treatments provided at a point in the birth weight distribution c. Given the evidence presented in the previous section, we can think of the amount of treatment as shifted upwards by some discrete amount κ below the cutoff c.

[11] For example, Haugen and Markestad (1997, p. 305) specifically state, "At the neonatal intensive care unit, Haukeland Hospital, University of Bergen, all infants born in the period 1/1/89–12/31/93 with birth weight less than 1,500g or gestational age less than 32 weeks were examined for ROP if they still remained in the hospital 4–5 weeks after birth."

(1) $$BW_i = H_i + e_i \qquad \text{Birth weight and initial health}$$

(2) $$D_i = g(H_i) + \kappa \cdot 1[BW_i < c] + v_i \qquad \text{Additional initial medical care}$$

In this framework, treatments D_i will be correlated with the unobserved health component not captured by birth weight through $g(H_i)$, thus confounding direct inference that conditions on birth weight. A regression discontinuity framework helps identify the role of medical treatments at the cutoff c. We adopt this approach following Lee and Lemieux (2010) and estimate variants of the following equation for different outcome variables y_i:

(3) $$y_i = f(BW_i - c) + \alpha \cdot 1[BW_i < c] + X_i\beta + \varepsilon_i,$$

where $f(\cdot)$ is a polynomial in the distance from the cutoff (we allow for different slopes on either side of the cutoff), X_i is a vector of covariates, and ε_i is an error term. Threshold crossing will induce a discrete jump in treatment $\Delta D_i = \kappa$ which will be uncorrelated with other determinants of outcome y_i.

While a regression discontinuity framework generates randomization of preconditions across the treatment threshold c, behavior of post-hospital investments can potentially be influenced by treatment. Thus the interpretation of the estimated coefficients should consider the possible role of parental or other nonhospital inputs that may react to treatment, and which can amplify or reduce the effect of medical interventions on measured long-run outcomes. For example, academic achievement has a long horizon, allowing for post-hospital investments to respond to initial treatment D over time. To make this idea more precise, let $I_t^{post}(H, D, \zeta)$ represent all accumulated investments up to period t, and be a function of initial health, treatment at birth, and a vector of all subsequent shocks to health or educational ability ζ. Let academic achievement be determined by initial conditions and the accumulated effects of all subsequent inputs as in Todd and Wolpin (2007):

(4) $$A_{it} = \phi_t H_i + \psi_t D_i + \varphi_t I_t^{post}(H_i, D_i, \zeta_i) + X_{it}\beta_t + \epsilon_{it} \qquad \text{Academic Achievement at } t,$$

where A_{it} is the academic outcome for child i at time t. A regression discontinuity approach will help solve the problem of nonrandom assignment of D_i at least locally. However, this framework also makes explicit that post-hospital investments may react to treatments and that the estimated coefficient α from the regression discontinuity in equation (3) will reflect the combination of the effect of initial treatment and the reinforcing or countervailing effect of later investments. Specifically, we can write the following expression for the coefficient of interest from the regression discontinuity estimation from equation (3):

$$\hat{\alpha} = \psi_t \cdot \kappa + \varphi_t \cdot \Delta I_t^{post}(c),$$

where $\psi_t \cdot \kappa$ is the structural effect of additional treatments at birth on academic achievement in t and $\Delta I_t^{post}(c)$ is the difference in average post-hospital investments children will receive as a consequence of obtaining additional treatment at the cutoff. The estimated coefficient $\hat{\alpha}$ should thus be interpreted as the total policy relevant effect of the increased medical care at this margin, which may include any possible reaction by post-investments. In our empirical analysis we attempt to gauge how important post-investments may be. We observe different sources of parental investments: time use surveys, quality of child care and school, timing of the mother's return to the labor force, etc., and study how these vary across the cutoff to search for evidence of differential post-investments.

An additional point to be made is that if treatment is effective in lowering mortality, the composition of children who survive to school age will also be affected. We deal with the composition bias in two ways. First, we assign counterfactual scores to children who died above the cutoff and examine the percentiles at which these children would have to score to nullify our results. The idea is to test how well children who died above the cutoff would have had to perform to smooth out our discontinuity in test scores. Second, we compute Lee (2008) bounds that specifically account for this type of attrition. These results are presented in Section V.

We estimate equation (3) using triangular weighted OLS within a window around the cutoff, and report the coefficients with robust standard errors clustered at the gram level (Lee and Card 2008). Since the cutoffs are only valid for births greater than or equal to 32 weeks in gestational age, we estimate separately for births below and above the gestational age cutoff. For births below 32 weeks in gestational age, we expect to see no discontinuity in outcomes.[12] We examine mortality using a similar specification.

We primarily use a window of 1,400–1,600 grams in Chile and a window of 1,300–1,700 grams in Norway for this study. In Section V, we explore the sensitivity of our results to a wide range of windows and polynomials on either side of 1,500 grams. To keep the set of covariates consistent across countries, we control for maternal characteristics (education, age, and marital status), type of birth service (doctor or midwife), birth region (in the case of Norway we use county), sex, and year of birth. We control for heaping at the 1,500 gram point as suggested by Barreca et al. (2011) in both regressions and graphical analysis. While these controls form the basis of our preferred specification, in Section V we explore a variety of issues, some common to RD designs and some specific to our context of examining birth weight as a running variable.

III. Data

A. *Chile*

The data we use from Chile comes from matching the population of births between 1992 and 2007 to death certificate data for the same years, and test score

[12] A general concern with the approach of dividing the sample into less than and greater than 32 weeks of gestational age is that the problems faced by VLBW children of greater gestational age (for example, these children might be small for gestational age) could be different from that faced by children of lesser gestational age. In order to directly examine children closer together in gestational age, online Appendix B, Table 11 reproduces some of the main results using gestational age of 30, 31, 33, and 34 weeks. The results are very similar using this restricted sample.

and transcript records between 2002 and 2010. As most children in the later years of the data are too young to be observed in school, we use births between 1992 and 2002 for our main sample and concentrate on academic achievement between first and eighth grade. The data on births and deaths come from administrative records provided by the Health Ministry of the Government of Chile (MINSAL). The data with valid identification accounts for 99 percent of all births and deaths in published aggregate figures (online Appendix B, Table 1). This dataset provides data on the sex, birth weight, birth length, weeks of gestation, and several demographic characteristics of the parents such as the age, education, and occupational status. In addition, the dataset provides a variable describing the type of birth, be it a single birth, double (twins), triple (triplets), etc. Focusing on births of weight within the relevant window of 1,400 grams to 1,600 grams, we see that mothers in this part of the birth weight distribution are surprisingly similar to the average mother. They have similar education levels, age, and are only slightly less likely to be married at the time of birth. However, 17 percent of births in this range are multiple, which is much higher than the population average of 1.8 percent. Births in this low birth weight window are also more likely to be attended by a doctor (54.9 percent) instead of a midwife (44.3 percent), (online Appendix B, Table 2).

We observe 4.02 million births between 1992 and 2007, out of which 0.9 percent (approximately 35,000 births) are observed to be below 1,500 grams in birth weight and are considered VLBW. Within the bandwidths we examine in this paper (between 1,400 and 1,600 grams) we observe 12,247 births. Among these, 6,782 births are for infants who are equal to or above 32 weeks of gestation. Dropping observations that are missing important covariates such as mother's education and marital status, and restricting the sample of births to those with mothers in the age range of 15–43 years leaves us with a sample of 6,109 births.[13] Our estimating equations use triangular weights which give the end points of 1,400 grams and 1,600 grams a weight of 0, so that our final estimating sample contains 5,129 observations for the mortality sample.

The data on academic achievement comes from two sources. The first dataset on school achievement comes from administrative transcript data for the population of students in school between 2002 and 2010. This data was made available by the Ministry of Education of Chile (MINEDUC) and covers all students in the country. The detailed transcripts include grades by subject for each student in a given year. We construct language and math averages and standardize grades for each student at the school-classroom level and average across first and eighth grade.[14] Ninety-five percent of all births between 1992 and 2002 are matched to this measure of their academic success. Using similar restrictions as above (and not counting the end points of 1,400 and 1,600 grams), we are left with a sample of 2,877 births above the gestational age of 32 weeks for regressions involving academic performance. Online Appendix B, Table 3 presents the outcome of the merge between vital stats and different educational records taking into account the births that have not survived until schooling age. This measure of academic achievement is useful both because it gives the maximum

[13] Our results are unchanged if we include some of these missing observations by adding a dummy variable to denote missing status (for example, mother's marital status).

[14] Alternative measures of academic achievement we study are average GPA, different ways of standardizing grades, and averaging over different grade levels.

possible number of observations, and because it also provides a measure of performance that is calculated over the entire school year and across several grades.

The second source of data is a national test administered to all fourth grade students in Chile called the SIMCE. We observe test scores for fourth graders in 2002 and yearly from 2005 to 2010 and standardize the scores by cohort. In cohorts that would have been in fourth grade (based on age), the match rate between vital statistics and fourth grade SIMCE is approximately 90 percent for the full distribution but 80 percent for births in the window of birth weight studied. Tables in online Appendix B show the details of this merge rate. While providing rich data on student characteristics, the amount of observations with SIMCE scores in the VLBW range is limited both because it was administered in years that cover about half the births between 1992 and 2002 and because of overall lower match rates due to missing or corrupted IDs in the SIMCE data. An important consideration here is that the match rates for both the administrative data on grades and SIMCE test data show no significant discontinuity at the cutoff of 1,500 grams.

B. *Norway*

For Norway, the primary data source is the birth records for all Norwegian births over the period 1967–1993. We obtained this data from the Medical Birth Registry of Norway. The birth records contain information on year and month of birth, birth weight, gestational length, age of mother, and a range of variables describing infant health at birth including APGAR scores,[15] malformations at birth, transfer to a neonatal intensive care unit, and infant mortality. We are also able to identify twin births. Using unique personal identifiers, we match these birth files to the Norwegian Registry Data, a linked administrative dataset that covers the entire population of Norwegians aged 16–74 in the 1986–2008 period, and is a collection of different administrative records such as the education register, the family register, and the tax and earnings register. These data are maintained by Statistics Norway and provide information about educational attainment, labor market status, earnings, and a set of demographic variables (age, gender), as well as information on families.

We can link data on grades from tenth grade to children in the birth files using unique identifiers. These records are provided directly from the schools to Statistics Norway. Written and oral exams are administered in the final year of junior high school at the national level and are externally graded. The written exam could be in either math, Norwegian, or English, with exam subjects determined at the school level. The students are informed of which exams they will take three days before the exam date. The oral exam is administered in a quasi-randomly selected subject and is also graded externally. As tenth grade is the last of the compulsory years of schooling, the grade obtained on this national test is important when applying for admission to selective high schools. The grades on this test range from 1 to 6, in discrete integers. We standardize the tests at the yearly national level. This data is available for cohorts born between 1986 and 1993.

[15] APGAR scores are a composite index of a child's health at birth and take into account Activity (and muscle tone), Pulse (heart rate), Grimace (reflex irritability), Appearance (skin coloration), and Respiration (breathing rate and effort). Each component is worth up to 2 points for a maximum of 10.

Mothers who give birth in this part of the birth weight distribution are quite similar to the average mother in the overall population of births, although they are slightly less likely to go to college and be married. Births in this range are much more likely to be multiple. Between 1,300 and 1,700 grams, 25 percent of births were twins or triplets, which is much higher than the population average of 2.4 percent. See online Appendix B, Table 2 for more characteristics of VLBW births in this sample.

We observe 460,507 births between 1986 and 1993, out of which 0.8 percent (3,741 births) are observed to be below 1,500 grams in birth weight and are considered VLBW. Within the birth weight window we examine for Norway (between 1,300 and 1,700 grams) we observe 2,477 births. We use a different window in Norway to increase sample size and to get more stable estimates for the academic achievement results, although our results are statistically significant even for smaller windows. We explain this window choice more in the results section. Among these 2,477 births about 1,498 births are for infants who are above 32 weeks of gestation (inclusive). More than 72 percent of all births born between 1986 and 1993 are matched to their educational records. We lose some observations due to deaths (neonatal, infant, and later deaths) and some to missing information on grades (this could be due to illness, strikes during the exam period, or other reasons for not taking the exams). Descriptive statistics for the sample is available in online Appendix B. As in Chile, we find no discontinuity in match rates around 1,500 grams.

IV. Results

A. *Treatments*

The rules and recommendations for medical treatment of premature births in Chile and Norway highlight the importance of providing special care for births below 1,500 grams or less than 32 weeks of gestation. Confirming the discontinuity of treatments quantitatively is difficult given the lack of micro data on hospital inputs. However, data on NICU usage from Norway and hospitalization records from public hospitals for a subset of approximately 30 percent of births from 2001 to 2007 in Chile, provide evidence that is consistent with differential health treatments across the relevant threshold. We also see that as expected from the description of the discontinuity in Section II, the evidence suggests a break in treatment at 1,500 grams for births of 32 weeks of gestation and above, but none for births of 31 weeks and below. This also provides evidence that this shift is not due to issues associated with nonrandom heaping.

Since all graphs in the paper largely follow the same format, it is useful to understand how these graphs are constructed. We drop observations at 1,500 gram point from the data while constructing graphs to be consistent with our regression results that control for the heap at this point (as we show later, our regressions are robust to dropping this point from the analysis). The dark lines are a linear fit using triangular weights on either side using the micro data (these triangular weights result in a weight of 0 to the 1,400 and 1,600 gram point, which is consistent with our regressions) using no coviariates. The dots represent averages of 30 gram bins (approximately 1 ounce) centered at 10 gram intervals. Graphs with different window and bin widths are presented in online Appendix B.

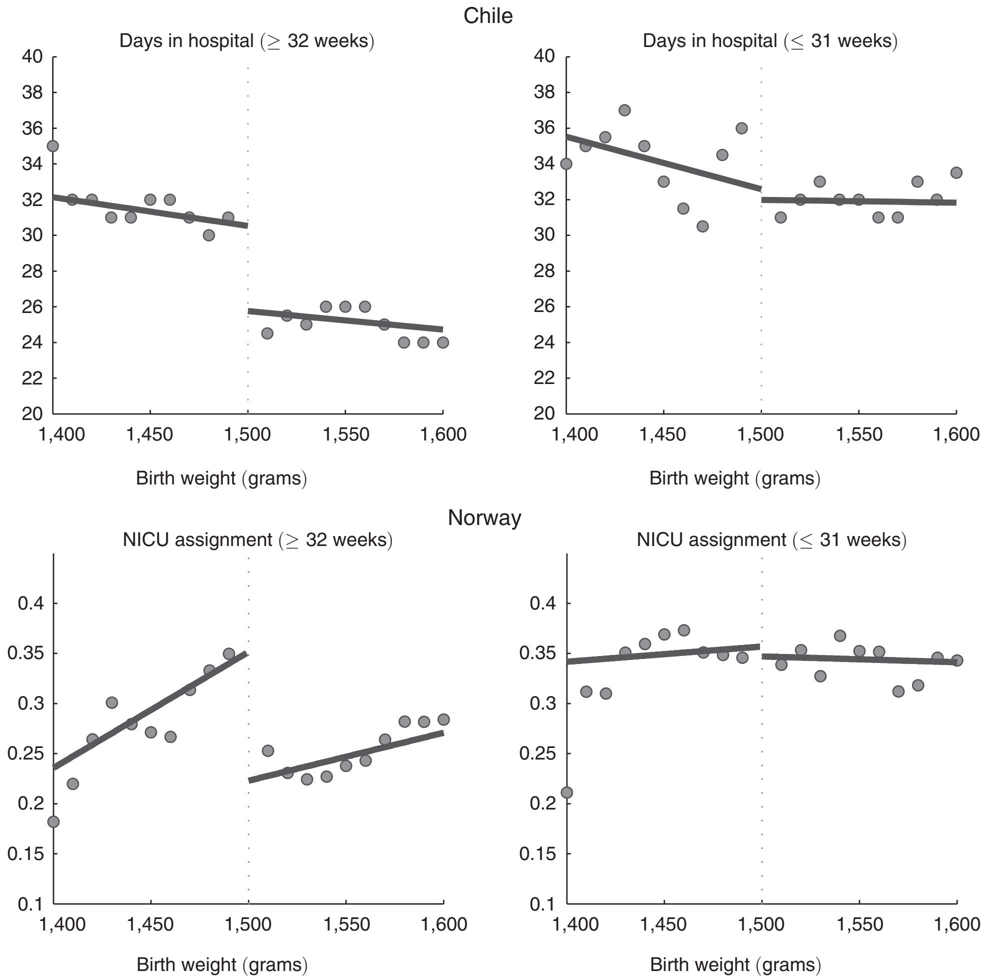

Figure 1. Treatments around 1,500 Grams

Notes: Top panel of this figure shows the relationship between birth weight and median days spent in public hospitals counting all hospitalizations that begin during the first month of life in Chile. The bottom panel shows the relationship between birth weight and whether or not the child was admitted to a Neonatal Intensive Care Unit in Norway. Data from Chile covers cohorts born 2001–2006, Norway data covers cohorts 1980–1993.

General graphing notes: We drop observations at 1,500 gram bin from the data while constructing graphs to be consistent with our regression results that control for the heap at this point. The dark lines are a linear fit using triangular weights on either side using the micro data. The dots represent averages of 30 gram bins (approximately 1 ounce) centered at 10 gram intervals. Graphs with different window and bin widths are presented in online Appendix B.

Figure 1 shows the relationship between birth weight and days of hospitalization that begin within the first month of life in Chile and NICU usage in Norway.

Table 1 provides the regression analog of these graphs. Children just below 1,500 grams in Norway are about 14 percentage points (43 percent) more likely to be admitted to a NICU, and children in Chile just below the cutoff spend around four days more in the hospital (over a mean of 28 days).[16] The NICU data from Norway,

[16] Hospital days is analyzed using a quantile regression at the median. The reason for this is that the number of observations are small and there are significant outliers which influence the results when using means. We experimented with other specifications which are presented in online Appendix B, Table 4 and find the results are largely

TABLE 1—TREATMENTS AROUND 1,500 GRAMS

	All gestational ages	Gestational age $\geq$ 32 weeks	Gestational age < 32 weeks
Panel A. Chile: number of days spent in hospital within a month of birth			
Birth weight < 1,500	1.576	3.976**	0.91
	(1.465)	(1.6)	(3.374)
Mean of dependent variable	32.95	28.89	37.38
Observations	862	449	413
Panel B. Norway: whether child was transferred to a NICU			
Birth weight < 1,500	0.087**	0.143**	0.004
	(0.035)	(0.052)	(0.034)
Mean of dependent variable	0.31	0.28	0.35
Observations	2,111	1,224	887

Notes: Window of 100 grams on either side of 1,500 grams used. Regression controls for mother's age, education and marital status, year of birth, and region/municipality of birth fixed effects, type of birth service, and 100 gram heap fixed effect. Linear slopes on either side of 1,500 grams are included and regression is weighted using triangular weights (only in Norway in this case). Standard errors are clustered at the gram level. Due to some outliers driving the results in small sample sizes in Chile, reported regressions are quantile regressions evaluated at the median.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

and hospitalization data from Chile, are consistent with the idea that a broad set of medical inputs have been applied differentially across the birth weight cutoff of 1,500 grams for births that are at least 32 weeks of gestational length.

B. *Mortality*

Mortality is a relatively short-run outcome which additional medical treatment would affect differentially across the relevant threshold. Figure 2 shows infant mortality, defined as death before the first year of life, in both Chile and Norway. Comparing the different panels in Figure 2 it is clear that most of the impact of being just below the cutoff of 1,500 grams is for children who were above 32 weeks of gestational age.

Table 2 estimates equation (3) and shows the results for infant mortality by gestational age. We find, as expected, that the 1,500 gram cutoff does not seem relevant for children less than 32 weeks in gestational age in either country. Column 2 indicates that children below 1,500 grams are 4.4 percentage points less likely to die within a year compared to children just above 1,500 grams in Chile (average infant mortality for this birth weight range is 10.9 percent). Given the low average mortality within this birth weight range, this is a fairly large effect.[17]

In Norway, children below 1,500 grams are 3.1 percentage points less likely to die within a year if they are born at or greater than 32 weeks in gestational age. This is a substantial effect given the already low average infant mortality rate for this group

consistent. Public hospitals were identified using the name of the institution which was available for 77 percent of births in the sample after 2001.

[17] In online Appendix B, Table 18, we show that the mortality effect in Chilean hospitals is most prominent in public hospitals where one expects stricter adherence to such rules of thumb. In addition, the effect around the cutoff is greater in hospitals that have a NICU.

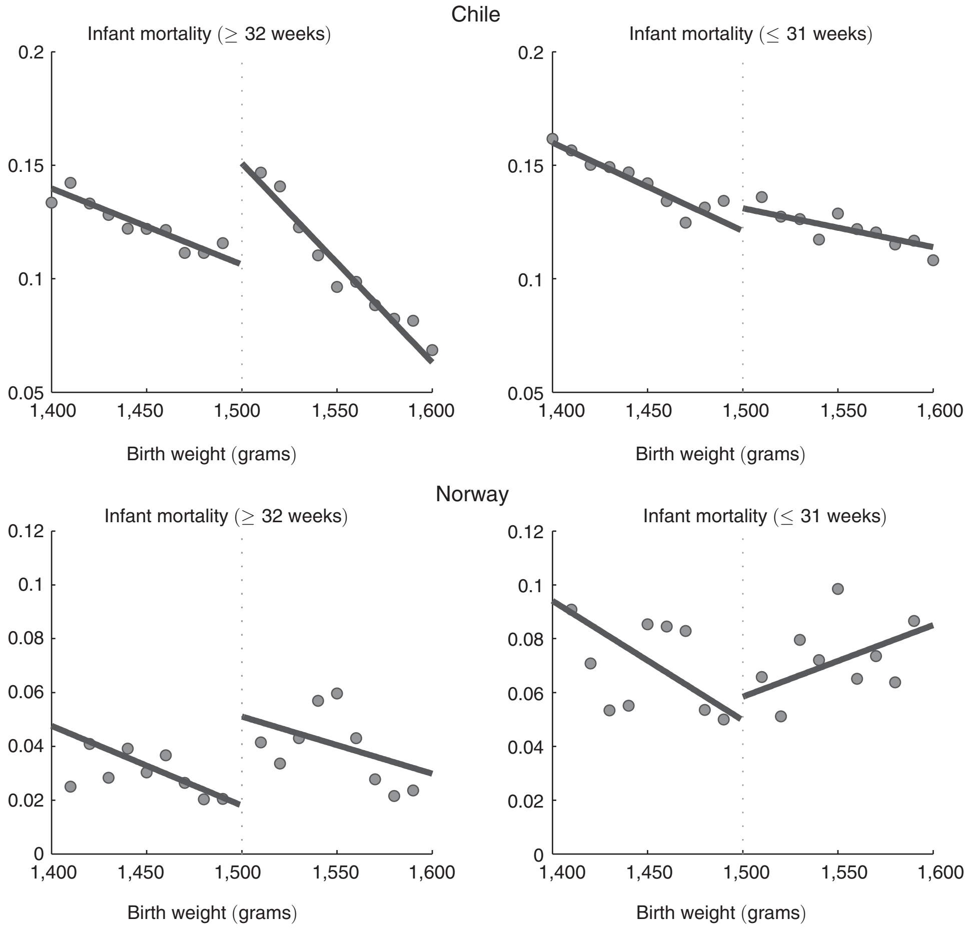

Figure 2. Infant Mortality

Notes: This figure shows the relationship between birth weight and infant mortality in Chile and Norway. Cohorts born between 1992 and 2007 in Chile, and 1980 and 1993 in Norway are used for this graph. General notes from Figure 1 apply.

of around 4 percent. We consider these results in line with children receiving extra treatments below the cutoff.

C. *Academic Achievement*

Whether medical interventions have a lasting impact on human capital can be analyzed by examining the relationship between academic achievement later in life and birth weight around the cutoff. Figure 3 presents a visual representation where it is clear that there is an effect, and that most of the impact of being below the cutoff is for children born with greater than 32 weeks of gestation in Chile and Norway.[18]

[18] Note that while there might appear to be a cutoff for Chile for less than 32 weeks, note that it goes in the *opposite* direction and also it is statistically insignificant.

TABLE 2—MORTALITY AROUND 1,500 GRAMS BY GESTATIONAL AGE

Chile: Birth cohorts 1992–2007	All gestational ages	Gestational age $\geq$ 32 weeks	Gestational age < 32 weeks
Infant mortality (death within one year of birth)			
Birth weight < 1,500	−0.0261*	−0.0449**	−0.00228
	(0.0134)	(0.0181)	(0.0196)
Mean of dependent variable	0.116	0.109	0.125
Observations	9,348	5,129	4,219
Norway: Birth cohorts 1980–1993			
Infant mortality (death within one year of birth)			
Birth weight < 1,500	−0.03*	−0.031**	−0.028
	(0.015)	(0.013)	(0.027)
Mean of dependent variable	0.053	0.036	0.08
Observations	4,035	2,437	1,598

Notes: Window of 100 grams on either side of 1,500 grams used for Chile and window of 200 grams on either side of 1,500 grams used for Norway. Regression controls for mother's age, education and marital status, year of birth and region/municipality of birth fixed effects, type of birth service, and 100 gram heap fixed effect. Linear slopes on either side of 1,500 grams are included and regression is weighted using triangular weights. Standard errors are clustered at the gram level.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 3 estimates equation (3) using school performance as the dependent variable. As before, these estimates are simply regression analogs of Figure 3. In Chile, we find consistent results when looking at different measures of academic achievement such as transcript level grades in math and language, overall GPA, and scores from a national exam (SIMCE) administered in fourth grade.[19] Table 3 shows that children below 1,500 grams perform around 0.15 SD better in math classes compared to students just above 1,500 grams. Online Appendix A, Table 1 shows an analogous result for language. The impact of being below the cutoff on average GPA from first to eighth grade is also positive and significant. As discussed in Section IV, the SIMCE is a national test administered to all fourth graders in Chile on a subsample of years. Table 3 shows that children born just below the cutoff on average obtain scores that are 0.13 SD higher, although this is only significant at the 15 percent level. Note that estimates using the SIMCE have fewer observations since we observe this test only for children in fourth grade and for less cohorts since the test was administered every year starting in 2005 (we have data until 2010). The general pattern of the results from the SIMCE, even though they are not statistically significant, appear consistent with our overall results.

[19] Online Appendix A, Table 1 provides other measures of school performance that restrict transcript data to grades 1–4 or standardize at the national level instead of the classroom level. The results are found to be very similar. Another point to note is that given the nature of the transcript data in Chile, some observations have more math grades available than other observations, depending on how long we observe them in school. We have tried specifications where we put more weight on the students with more observations and this does not change the basic import of the results. In particular, while the effect size decreases to around 0.101 SD, the effect remains statistically significant at the 5 percent level.

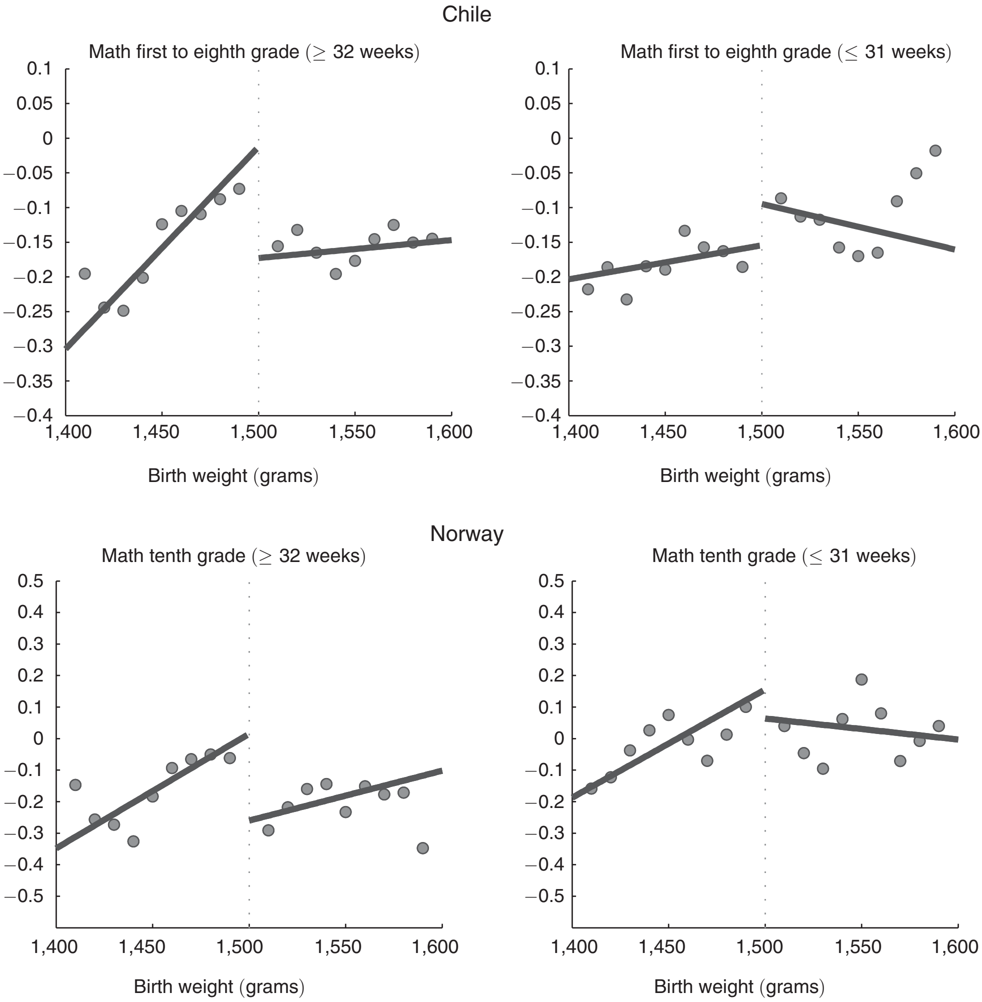

Figure 3. School Performance

Notes: This figure shows the relationship between birth weight and standardized math grades, averaged over grades one to eight in Chile and nationally standardized test scores (tests are either in math or language) administered in tenth grade in Norway. Chilean data consists of cohorts born 1992–2002 and Norwegian data consists of cohorts born 1986–1993. General notes from Figure 1.

The Norwegian results use the tenth grade national exam administered yearly starting in 1986. We use the standardized average of the written and oral portion of the national exam. Cohort sizes being much smaller in Norway (compared to Chile), Table 3 shows that the results are sensitive to choice of window length around 1,500 grams. Using the same 100 gram window as in Chile results in significant but rather large estimates of the impact of being below 1,500 grams. From online Appendix A, Table 7 it is clear that the size of this coefficient falls by half when we use a window of 120 grams on either side of 1,500 grams. The size of the coefficient however remains stable after that. We thus prefer using a 200 gram window on either side of 1,500 grams in Norway. Not only does the magnitude appear more in line with what we find in Chile, but we obtain more precision since a larger window provides more observations. Our preferred estimates from Norway suggest a 0.22 SD

Table 3—School Performance around 1,500 Grams by Gestational Age

	Chile school outcomes		
Birth cohorts 1992–2002	All gestational ages	Gestational age ≥ 32 weeks	Gestational age < 32 weeks
Classroom standardized math scores			
Birth weight < 1,500	0.0676	0.152**	–0.0363
	(0.0484)	(0.0583)	(0.0750)
Mean of dependent variable	–0.155	–0.153	–0.157
Observations	5,022	2,877	2,145
School GPA			
Birth weight < 1,500	0.0247	0.0905**	–0.0594
	(0.0222)	(0.0364)	(0.0344)
Mean of dependent variable	5.771	5.786	5.752
Observations	5,114	2,935	2,179
SIMCE Scores in math (administered only in 2002 and yearly from 2005–2010)			
Birth weight < 1,500	–0.0176	0.135	–0.232*
	(0.0845)	(0.0906)	(0.135)
Mean of dependent variable	–0.156	–0.157	–0.154
Observations	2,469	1,463	1,006
Birth cohorts 1986–1993	Norway tenth grade national exam		
100 gram window on either side of 1,500 grams			
Birth weight < 1,500	0.275*	0.476***	0.025
	(0.150)	(0.097)	(0.334)
Mean of dependent variable	–0.081	–0.145	0.011
Observations	940	556	384
200 gram window on either side of 1,500 grams			
Birth weight < 1,500	0.179*	0.228**	0.101
	(0.089)	(0.087)	(0.171)
Mean of dependent variable	–0.114	–0.166	–0.03
Observations	1,880	1,163	717

Notes: Window of 100 grams on either side of 1,500 grams used for Chile and window of 100 and 200 grams on either side of 1,500 grams used for Norway. Regression controls for mother's age, education and marital status, year of birth and region/county of birth fixed effects, type of birth service, and 100 gram heap fixed effect. Linear slopes on either side of 1,500 grams are included and regression is weighted using triangular weights. Standard errors are clustered at the gram level.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

increase in scores for children born below the cutoff. Some of the results for less than 32 weeks appear sizable, although they are usually insignificant (except in the case of SIMCE) and in the opposite direction.

Bias Due to Selection into Survival.—The results on differential mortality around the 1,500 gram cutoff suggest that there is selection into being observed in school which will introduce bias as survivors are likely to get different scores than those who do not survive. In general we think that the bias would lead to an underestimate

Table 4—Counterfactuals using Nonsurvivors of Infancy

	Only survivors	Percentile of test score assigned to nonsurvivors above 1,500 grams					
		Median	55th	60th	65th	75th	80th
Chile							
Birth weight < 1,500	0.152**	0.145**	0.145**	0.140**	0.118**	0.0582	−0.00244
	(0.0583)	(0.0581)	(0.0581)	(0.0567)	(0.0583)	(0.0596)	(0.0612)
Observations	2,877	3,166	3,166	3,166	3,166	3,166	3,166
Norway							
Birth weight < 1,500	0.228**	0.231**	0.232**	0.226**	0.224**	0.216**	0.205**
	(0.087)	(0.088)	(0.088)	(0.087)	(0.087)	(0.086)	(0.086)
Observations	1,163	1,184	1,184	1,184	1,184	1,184	1,184

Notes: This table assigns counterfactual scores to children with birth weight above 1,500 grams who are not observed in the data due to death within the first year of their lives. These children are assigned scores at the percentile (indicated at each column) within their 10 gram birthweight bin. Window of 100 grams on either side of 1,500 grams used for Chile and window of 200 grams on either side of 1,500 grams used for Norway. Regression controls for mother's age, education and marital status, year of birth and region/municipality of birth fixed effects, type of birth service, and 100 gram heap fixed effect. Linear slopes on either side of 1,500 grams are included and regression is weighted using triangular weights. Standard errors are clustered at the gram level.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

of the true effect. This is because the weakest children survive below the cutoff, and these children might get the worst grades among their birth weight cohort. On the other hand, the weakest children above the cutoff end up dying, hence, raising the average test scores for those birth weight groups.

In Table 4 we offer some counterfactual scenarios where we examine the extent of this bias. We consider pessimistic scenarios and start by assigning nonsurviving children above 1,500 grams the median score of their birth weight group. In both countries we use birth weight grouped at the 10 gram level to assign these counterfactual scores. We subsequently assign the nonsurvivors better and better scores, ranging from the 55th percentile to the 80th percentile within their nearest 10 gram birth weight bin. Under the counterfactual scenario of the nonsurvivors scoring at the 75th percentile (or higher) of their birth weight bin class in Chile, we no longer find evidence for a discontinuity. In the Norwegian case, there does not appear to be a point beyond which we would not find a discontinuity in scores. Hence, the selection into mortality above the 1,500 gram mark has to consist of some of the smartest children in their birth weight bin class for our results to disappear.

More formally, we can adapt the procedure outlined in Lee (2008) to create bounds for our treatment effects. The main idea behind the bounding procedure is to identify the extra people who are treated into survival below 1,500 grams and trimming the upper and lower tails of the test score distribution by this number. For conciseness, we only discuss the Chilean case here. The results for both countries are presented in online Appendix A, Table 2. Since this is not a standard treatment-control design experiment, the average test scores for children in the 1,400–1,500 gram range is not higher than the average score for children in the 1,500–1,600 gram range. To execute the Lee (2008) procedure, we choose a smaller window where the average scores for treated children are higher than the average scores for untreated children.

This occurs in a 50 gram window around 1,500. Indeed Figure 3 would indicate that this is the case. As listed in online Appendix A, Table 2, within this 50 gram window the mean treatment effect, without adding any covariates, is 0.09 SD. Accounting for differential mortality within this window (treated children are 1.4 percent more likely to be alive), the upper and lower bounds for the test score effect are 0.069 and 0.11 SD. The upper and lower bounds in Norway are 0.15 and 0.21 respectively.

Introduction of Surfactant.—One specific treatment we explore further is that of surfactant use. Surfactant is a soap like material produced in the lungs and is essential for proper lung function. Infants who have not produced enough surfactant on their own cannot fully utilize their lungs for breathing. Hence, surfactant therapy is considered a breakthrough in preventing deaths due to Respiratory Distress Syndrome (RDS) and Bronchopulmonary Dysplasia (Schwartz et al. 1994). Moreover, the medical literature cites Bronchopulmonary Dysplasia and early childhood lung diseases to be significantly correlated with cognitive outcomes (Singer et al. 1997; D'Angio et al. 2002; Marlow et al. 2005). One of the pathways by which preterm birth might affect cognitive outcomes appear to be related to the development of the lung and the delivery of oxygen to the brain. Hypoxia (reduction in oxygen supply to tissues) or ischemia (a severe low oxygen state) in the perinatal period is one of the leading causes of brain injury in preterm infants (Luciana 2003).

As mentioned earlier, Norway in 1989 and Chile in 1998 introduced universal surfactant therapy to be administered to VLBW infants (Gonzalez et al. 2006; Saugstad et al. 2006). We explore the timing of the introduction of surfactant to provide suggestive evidence of the long-term impacts of surfactant treatment on school outcomes. We estimate equation (3) in the time periods before and after the introduction of surfactant to show that most of the effect on test scores we see come from the period when surfactant was used.

Table 5 shows that after the introduction of surfactant, the cutoff of 1,500 begins to play an even more important role in determining school outcomes. In Chile, the impact of being below the cutoff after 1998 is 0.19 SD (just shy of significance at the 10 percent level), and in Norway it is 0.34 SD. In the case of Chile, we find substantial reductions in infant and neonatal mortality around the cutoff after the introduction of the surfactant program.[20] We view these results as suggestive evidence that the introduction of surfactant played an important role in improving the different outcomes we measure.[21]

Parental Responses.—As emphasized earlier, interpreting long-run impacts of early life events is complicated by the fact that parents might respond to these

[20] We restrict the post-period for surfactant to 2003 in Chile, since after 2003 other programs like PNAC and AUGE started which also affected births at precisely this cutoff. In Norway (results not presented) we do not find large impacts on mortality. This is likely due to the fact that 1980–1989 were the most dramatic in terms of the decline in infant and neonatal mortality in the country. Infant mortality before 1989 was around 10 percent, but after 1989 is around 2 percent. We are unable to detect a differential impact around 1,500 grams in the post-surfactant period perhaps due to the low levels of infant mortality.

[21] To the best of our knowledge, no major policies were implemented around these time periods. In Chile specialized nutritional programs were introduced only in 2003 (PNAC). However, in 1999 the Ministry of Health published and distributed a handbook for training programs on the following and caring of VLBW births. This might have also emphasized cutoffs and generated an alternative reason for mortality to improve more under the cutoff after 1998. We are not aware of any competing medical programs for VLBW infants in Norway around 1989.

Table 5—Role of Surfactant

	Chile: Surfactant introduced 1998		Norway: Surfactant introduced 1989	
	Pre-surfactant (1992–1997)	Post-surfactant (1998–2002)	Pre-surfactant (1986–1988)	Post-surfactant (1989–1993)
Test scores				
Birth weight < 1,500	0.103**	0.197	−0.044	0.349***
	(0.0509)	(0.134)	(0.260)	(0.130)
Mean of dependent variable	−0.132	−0.198	−0.107	−0.302
Observations	1,990	887	354	809
	Chile: Infant mortality		Chile: Neonatal mortality	
	Pre-surfactant (1992–1997)	Post-surfactant (1998–2002)	Pre-surfactant (1992–1997)	Post-surfactant (1998–2002)
Mortality				
Birth weight < 1,500	−0.0152	−0.0693**	−0.0155	−0.0548**
	(0.0309)	(0.0296)	(0.0327)	(0.0252)
Mean of dependent variable	0.13	0.1	0.021	0.025
Observations	2,021	1,801	2,021	1,801

Notes: Window of 100 grams on either side of 1,500 grams used for Chile and window of 200 grams on either side of 1,500 grams used for Norway. Regression controls for mother's age, education and marital status, year of birth and region/municipality of birth fixed effects, type of birth service, and 100 gram heap fixed effect. Linear slopes on either side of 1,500 grams are included and regression is weighted using triangular weights. Standard errors are clustered at the gram level.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

shocks. We explore if there is any evidence of differential parental investment decisions around the cutoff as a way to gauge how important this mechanism may be in determining the results we have found.

The first dimension of parental responses that we examine in Chile and Norway is that of school choice. Within the framework described so far, we examine whether school quality varies across the cutoff. We use the average of the national standardized math scores by school to examine whether students below the cutoff attend schools of different quality on average. In Table 6 we find that this is not the case. Hence, it appears that at least on school choice, parents do not invest differentially around the cutoff.

We also explore different avenues of parental investments by examining data on parental time use. In Chile, when the SIMCE is administered, a detailed survey is handed out to parents and students. The content of these surveys vary from year to year, but in several years the surveys contained a set of detailed time investment questions to the parents. Hence, for a sample of fourth graders, we have detailed information on time spent by parents in activities such as reading to their children.[22]

In Norway, while we lack direct measures of parental time investments in their children, we can proxy for parental time by examining when the mother returns to the labor force after giving birth. In addition we can examine whether the child was

[22] Generally the questions from year to year do not have much overlap, except for the questions regarding reading investments. Hence, we choose to show results for this type of investment so as to maximize the number of observations over two cohorts. These data are not from time use diaries. Responses to the questions on investments typically range from 1 to 4 where 1 is no "Never" and 4 is "Very often."

TABLE 6—PARENTAL INVESTMENT RESPONSES AROUND CUTOFF

Chile	Average raw SIMCE score in school	Private school	Grade size	School in top 25 percent of SIMCE score distribution	Parents read "often" to child during the week
Birth weight < 1,500	2.433	0.0144	−6.147	−0.0465	0.0149
	(2.327)	(0.0272)	(4.361)	(0.0364)	(0.106)
Mean of dependent variable	251.0	0.0672	62.34	0.301	0.365
Observations	2,094	2,174	2,174	2,094	641
Norway	**Enrolled in child care at age five**	**Average exam score in school**	**Return to work after paid maternity leave**	**log parental income at the time of test**	**Mother employed at the time of test**
Birth weight < 1,500	0.009	−0.004	−0.034	0.042	0.029
	(0.041)	(0.033)	(0.036)	(0.070)	(0.027)
Mean of dependent variable	0.83	0	0.66	13.2	0.79
Observations	1,249	683	1,594	1,507	1,594

Notes: Standard errors clustered at the gram level. See Table 3. School level measures in Chile are measured as of grade four. Grade size refers to the number of students in the entire grade. Average number of classrooms per grade is two. Parental reading measures come from self reported surveys administered along with the SIMCE in 2002 and 2007. Answers range from very often = 1, to never = 4. We create a binary variable which is 1 if parents read very often or often (answers 1 and 2), and 0 otherwise. Child care in Norway is coded as 1 if the care if formal, and 0 if care was informal i.e., nannies at home, grandparents, etc. Return to work variable in Norway is coded as 1 if mother returns to work after the end of maternity leave.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

enrolled in formal child care by age five. Formal care includes public and private child care centers, while informal care typically involves nannies hired by parents, grandparents, and the like (for more on Norway's child care system see Black et al. 2012). If parents of VLBW children stay home more or provide different types of child care then we might expect to see discontinuities along these parental investment measures around 1,500 grams. As Table 6 shows, we find no evidence of differential parental responses around 1,500 grams along any of these measures.

D. *Discussion*

The results presented above indicate that health investments in early childhood matter for infant survival and educational achievement later in life. Evidence on both the short- and long-run effectiveness of early life health interventions is crucial for estimating their efficacy and orienting public policy. These results support recent evidence that early childhood is a critical period for determining adult outcomes and highlight the role of health policy in promoting better educational outcomes later in life.[23] From a policy perspective, our findings would suggest that an important

[23] Recent work by Veramendi and Urzúa (2011) and Noboa-Hidalgo and Urzúa (2012) have shown in the context of Chile, the importance of publicly provided child care centers in improving not only cognitive abilities among children, but also noncognitive abilities. These are more short-term outcomes but are consistent with the longer run results presented here. Similar results have been found while analyzing the impacts of cash transfer programs on young children in many Latin American countries (see Schady 2006 for an excellent review). However, most of

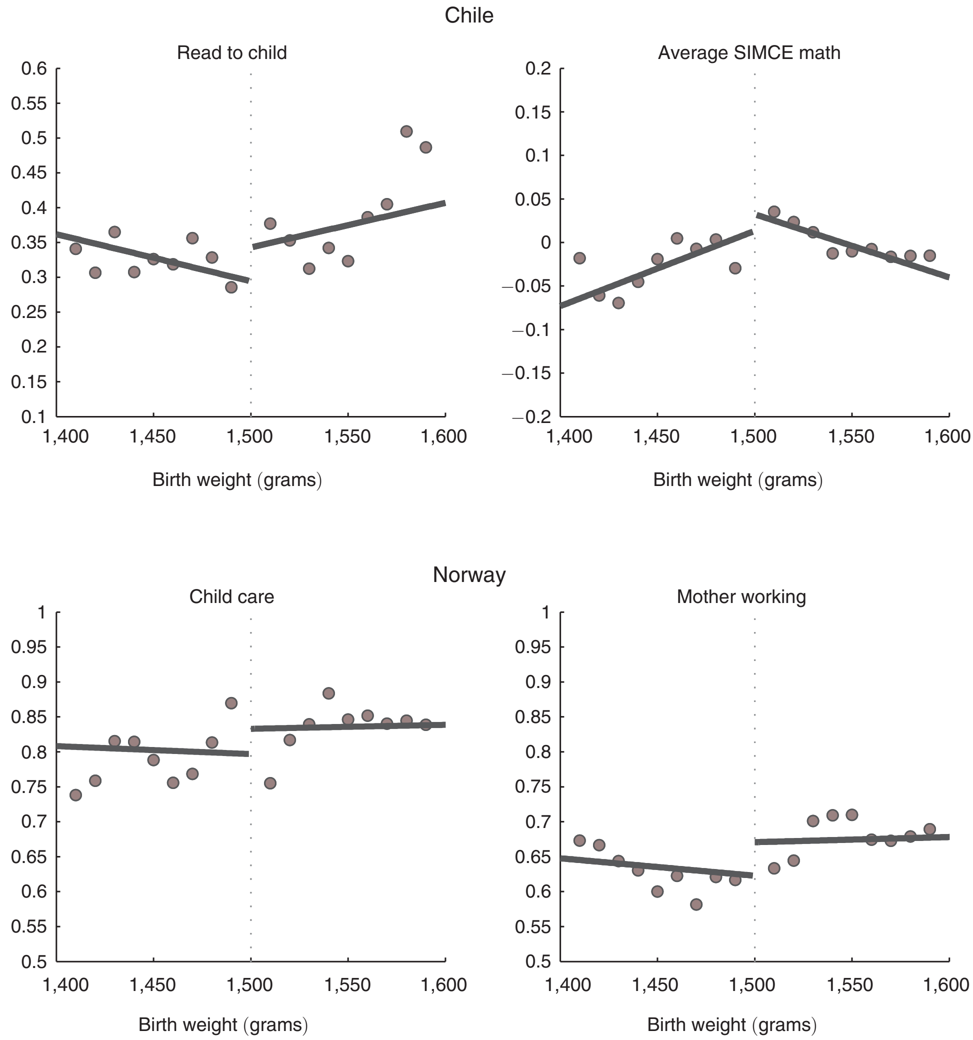

Figure 4. Parental Responses

Notes: Data on average math school scores in Chile come from averaging fourth grade SIMCE scores from tests administered in 2002, 2005–2010. In Norway, the average school score is constructed by averaging school scores for test years 2002–2009. Reading investment variable is obtained from parental survey component of Chile's SIMCE questionnaire, and responses range from 1–4, with 1 being very often. Reading related questions were asked in 2002 and 2007. Data on child care in Norway is available for 1980–1993. General notes from Figure 1 apply and the sample consists of only birth of 32 weeks of gestation or more.

source of inequalities in later life cognition or labor market success might arise from differences in access to health interventions, in this case, access to specialized neonatal treatment at birth. It is also important to note that the results have been found in countries at very different stages of development suggesting the importance of neonatal health care applies more generally to a broad set of countries.

---

these studies analyze rather short-term impacts of such investments and few studies are able to analyze long-term impacts. Studies evaluating the Perry Preschool Program, Head Start, or Project STAR do find long-term effects (see Garces, Thomas, and Currie 2002 or Chetty et al. 2011); although these are not health-based interventions.

The results found for mortality are large. To place our mortality results in context, we compare our findings to that of Almond et al. (2010), who find large mortality effects around 1,500 grams in the United States. A comparison of magnitudes suggests that the effects seen in Chile and Norway are larger than those found in the United States. Almond et al. (2010) find that children just below 1,500 grams have a 1 percentage point lower infant mortality rate compared to children just above 1,500 grams. Relative to the mean of around 5 percent mortality in their sample, this is a large effect.[24] As noted earlier, our results from Chile suggest a reduction in mortality of 4.4 percentage points relative to a mean infant mortality rate of nearly 11 percent for this sample. The magnitudes from Norway are even larger.

Apart from potentially different institutional settings, one reason we find larger effects is our focus on children above 32 weeks of gestational age. When we include all gestational ages, the coefficient of interest on infant mortality declines. Hence, it is likely that some treatments are administered to all children regardless of birth weight. One of the downsides of our study relative to the one by Almond et al. (2010) is that we are unable to provide details on the treatments. Future research will hopefully shed more light on the mechanisms that lead to long-term effects. What is certainly evident is that early childhood health interventions play an important role in determining mortality across three countries that differ in their socioeconomic characteristics.

While the academic achievement results from Chile and Norway appear similar, they differ in a few important ways. The results from Norway are for cohorts that were born 1986–1993, and in Chile the results are for cohorts born 1992–2002. Hence, it is possible that later cohorts in Chile received more advanced treatments as the treatment of at-risk newborns has changed over time. Moreover, the Norwegian test results are from grade ten, whereas the Chilean scores are a combination of grades achieved between grades one to eight. Most of the effects seen in Chile appear to come from earlier grades rather than later grades. Hence, the Norwegian sample provides evidence of rather long-term effects that we are not able to detect in the Chilean sample. It is possible that differences in care for newborns or better practices in Norway lead to more long-term effects, but without a systematic and detailed comparison of the medical technologies from the two countries and the relevant time periods, this is hard to assess.

These differences aside, our results on educational achievement in the context of education specific interventions are quite sizable. While the obvious thrust of the medical interventions we examine is to save lives, we can attempt to think about the monetary benefits by examining the results in Chetty et al. (2011). One of their results suggest that a one standard deviation increase in kindergarten entry test scores is correlated with an 18 percent increase in earnings. While keeping in mind that this is a correlation and that this correlation is based on data from the United States, we translate this correlation in the Chilean context to result in an increase of 2.7 percent increase in incomes (using an effect size of 0.15 SD increase in test scores). The increase in the Norwegian context would be around 1.8 percent.

[24] The effect size is similar in their paper as long as the focus is on low-quality hospitals, due to the comment raised by Barreca et al. (2011).

V. Robustness Checks

We explore the robustness of several aspects of our empirical strategy in this section. Some of these are threats to identification which are general and apply to any application of an RD strategy while others are specific to using birth weight as a running variable.

A first general check is that the running variable is being manipulated in its assignment across the cutoff. Our regression discontinuity design will not identify the effects of extra medical treatment if doctors or parents were systematically manipulating the birth weight variable. If they were, then we might expect to find many births around 1,490 and fewer births around 1,510. One visual way of checking for manipulation of the running variable is to simply plot a detailed histogram of the data and to check whether abnormal heaps occur to the left- or right-hand side of the cutoff.

As can be seen in Figure 5 this does not appear to be the case in either country. We test this (as do Almond et al. 2010) by collapsing the data at the gram level at which the data was naturally collected and testing in a framework similar to equation (3), whether more (or less) births are reported just below the cutoff compared to just above the cutoff.

In the greater than 32 week gestation sample, the coefficient (standard error) on the cutoff dummy is -16.78 (30.33). In Norway the analogous coefficient and standard error is 5.3 (10.7). These tests suggest that there is no manipulation of the running variable in this case.[25]

Another standard check in applications with an RD design is to verify that no other predetermined variables should display discontinuities around the cutoff apart from the treatment. In online Appendix A, Table 3 we show for both countries that a number of demographic characteristics like mother's education, mother's age, mother's employment status, twin or singleton status, and whether the mother was married at the time of birth appear smooth around the cutoff of 1,500 grams. In addition in Norway, we can examine APGAR scores and family income at the time of birth, both of which appear to be smooth at the cutoff. A graphical equivalent of this is Figure 6 and Figure 7. Were these to show discontinuous jumps, we would be concerned that socioeconomic characteristics determine which side of the cutoff an infant is observed on, invalidating the random assignment assumption.[26]

We also examine the role of covariates by adding them sequentially in the framework of equation (3). Online Appendix A, Table 4 shows how the coefficient on the cutoff dummy changes as we add more and more covariates (analogous table for mortality in online Appendix B, Table 10). Overall, the results show a rather limited role for covariates in determining the size of the coefficient on the cutoff dummy.[27]

[25] Manipulation in the context of birth weight and medical care is a potential concern as shown to be the case in Japan in a recent working paper by Shigeoka and Fushimi (2011).

[26] Online Appendix A, Table 3 only shows the smoothness of covariates for the schooling sample. Since we have a different sample for analyzing mortality results, we show in online Appendix B, Table 12 that covariates are smooth for various subsamples analyzed in the paper. Figures in online Appendix B also show covariates obtained from the fourth grade SIMCE surveys.

[27] Another way to understand the extent to which mother level unobservables might be driving the estimates is to examine children of the same mother. We can do this using twins and siblings that are identified in the data using the unique identifier for the mother. Certainly the demands of the data are rather high—the sample used for identifying

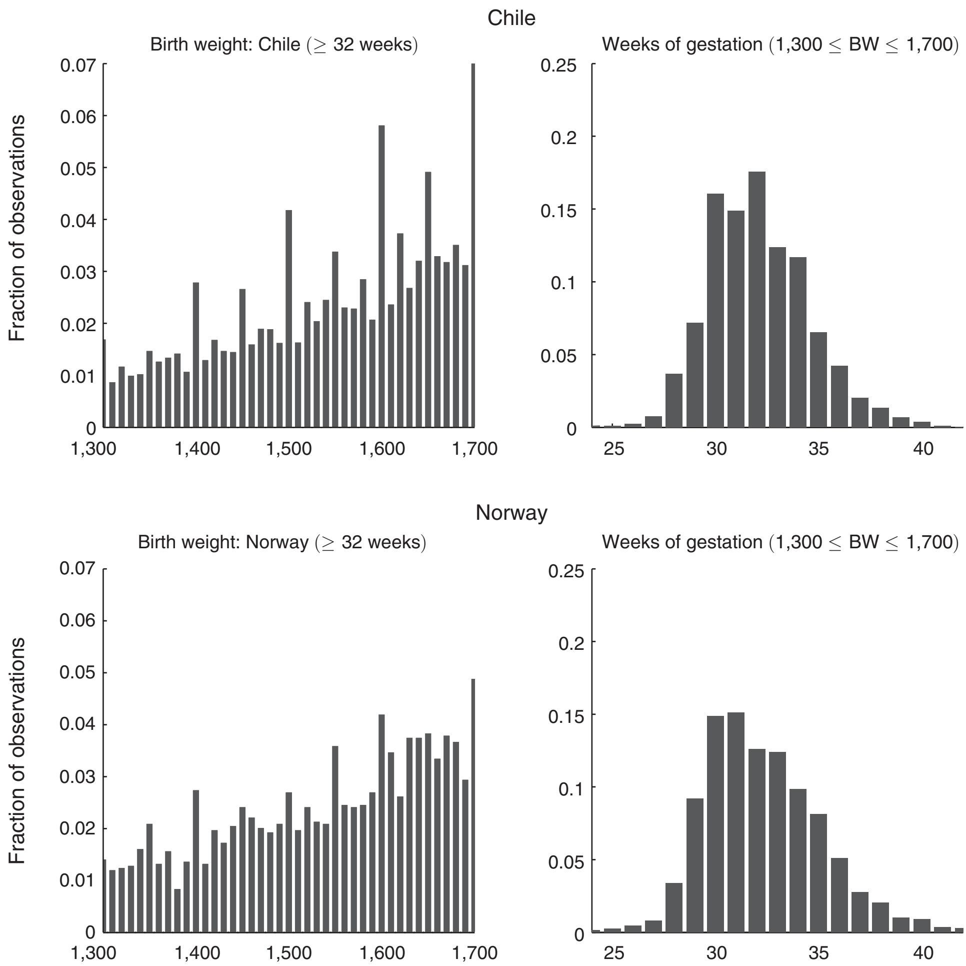

FIGURE 5. HISTOGRAMS OF BIRTH WEIGHT AND GESTATIONAL AGE

Notes: The histogram of birth weight is presented for birth with 32 weeks of gestation or more. The histogram of weeks of gestation are presented for births with weight between 1,300 grams and 1,700 grams. Chile graphs use data from 1992–2007, Norway graphs use data from 1980–1993.

We also verify that we do not observe similar results as those presented above at other intervals of 100 grams. If we observed that children below 1,700 grams, for example, had higher test scores than children slightly above 1,700 grams, then we would be concerned that something inherent about getting heaped at 100 gram intervals is driving the results rather than exposure to treatments specific to being less than 1,500 grams. In general, this is less of a concern in our context since if

the RD within a twin or sibling fixed effect requires one twin (or sibling) on either side of the cutoff, both twins (or siblings) above 32 weeks of gestation and a birth weight difference of no more than 200 or 400 grams (*both* have to fall between the range of 1,400 and 1,600 in Chile and between 1,300 and 1,700 in Norway). With caveats for small samples in place, we estimate mortality regressions (sample is too small for schooling outcomes) around the cutoff using twins and siblings. The point here is not to compare these estimates to the overall estimates we showed earlier, but rather to understand how much difference the fixed effect makes. In online Appendix B, Table 9 we show that OLS and FE estimates for both twins and siblings are very similar. This suggests that unobserved mother characteristics or propensities to manipulate birth weights say, are not playing an important role in this setting.

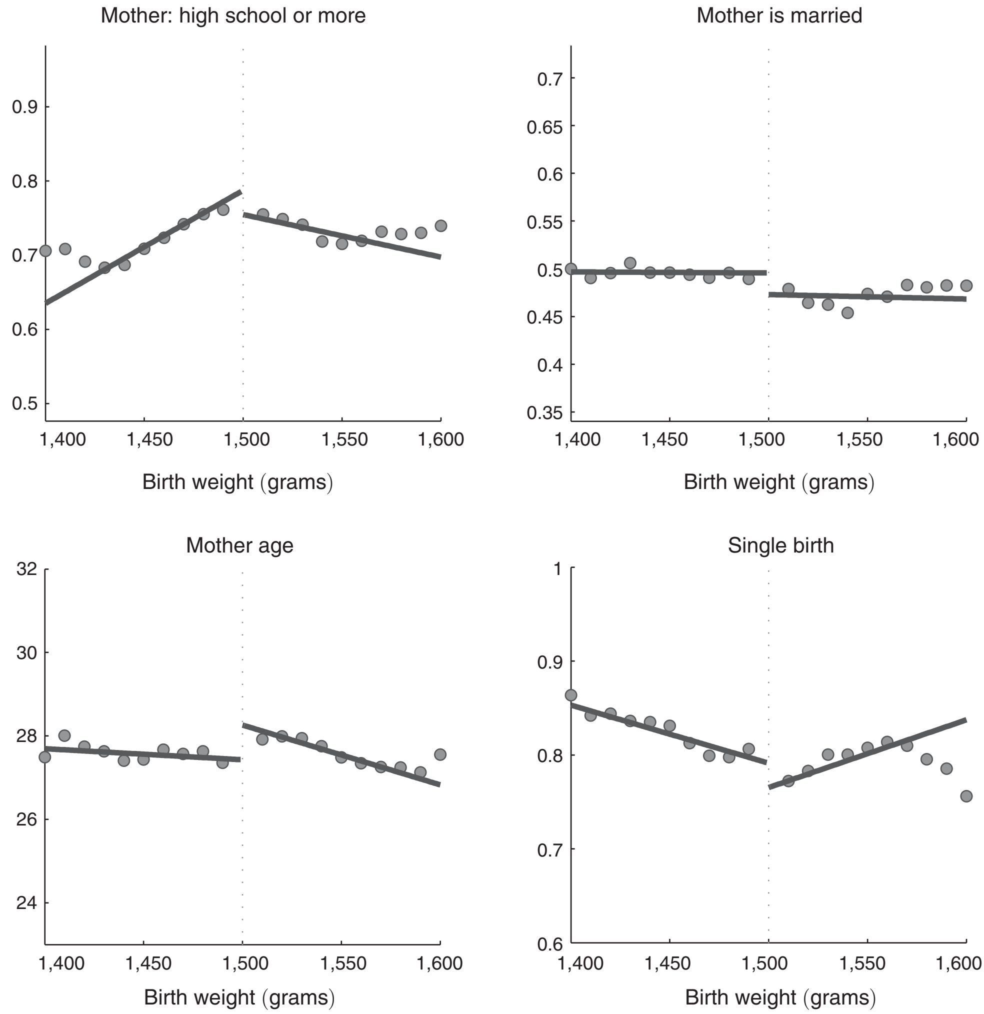

Figure 6. Baseline Covariates around 1,500 Grams: Chile

Notes: This figure shows the relationship between birth weight and other covariates in Chile. Cohorts born between 1992–2007 used for this graph. For other covariates, please see online Appendix B. General notes from Figure 1 apply and the sample consists of only birth of 32 weeks of gestation or more.

this were true, we should find that 1,500 gram matters even for gestational age less than 32 weeks.[28] Nevertheless, in online Appendix A, Table 5 we examine every 100 gram cutoff in a similar estimation strategy as in equation (3) (similar table for mortality, hospitalizations, and NICU admissions is in online Appendix B, Table 14). We find that in both countries, test scores are significantly affected only around the 1,500 gram cutoff.

[28] Moreover, given the long data series in Norway, we can show that the 1,500 gram point as a discontinuity only occurs starting in the 1980s. WHO recommendations and documents in Norway show that this was the period in which focus on VLBW birth was most apparent. Prior to 1980 it is unclear whether such rules existed. These results are available upon request.

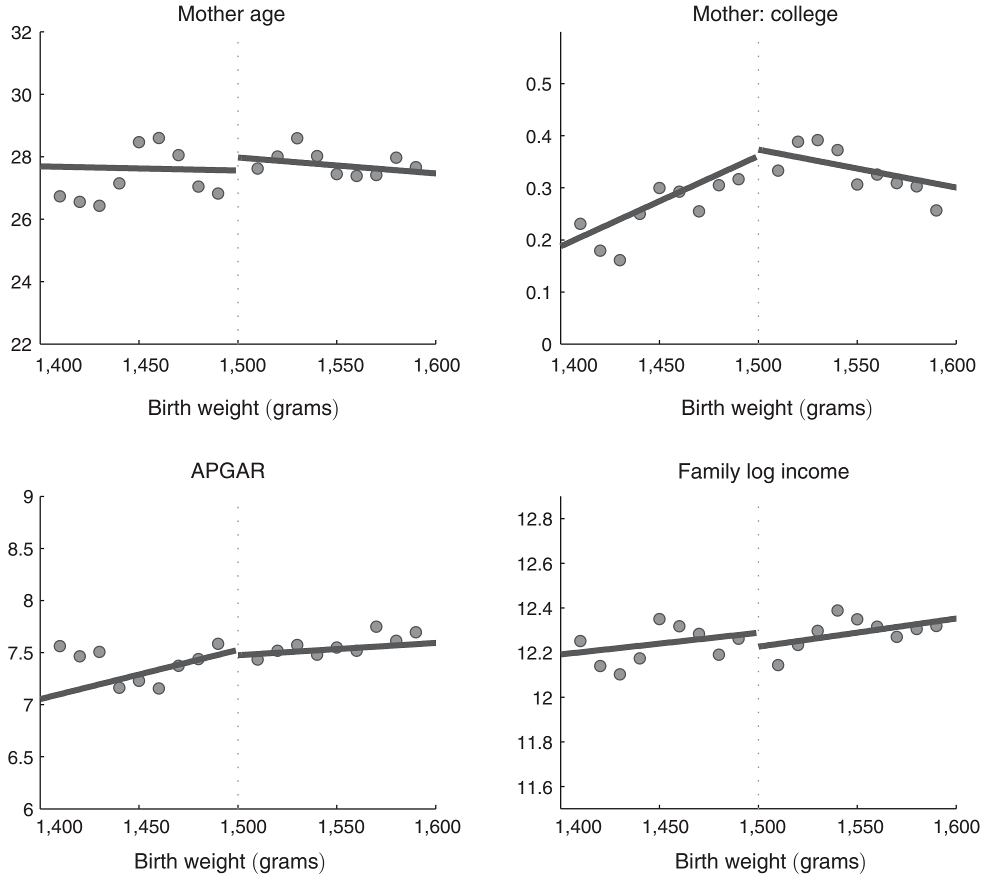

Figure 7. Baseline Covariates around 1,500 Grams: Norway

Notes: This figure shows the relationship between birth weight and other covariates in Norway. Cohorts born between 1980–1993 used for this graph. For other covariates, please see online Appendix B. General notes from Figure 1 apply and the sample consists of only birth of 32 weeks of gestation or more.

We explore the robustness of the estimates of equation (3) using a wide variety of birth weight windows and polynomials on either side of the 1,500 gram cutoff (online Appendix A, Table 6). While the results are largely consistent across different bandwidths for a given polynomial selection, the results across different polynomials for a given bandwidth do tend to differ, especially at smaller bandwidths. We attribute the sensitivity of our results to higher order polynomials to over fitting the data with few data points. To the extent that the results are largely similar for polynomials of up to order 3 and for bandwidths reaching up to 150 grams on either side of 1,500, we consider our results to be quite robust to bandwidth and polynomial selection. Moreover, visual inspection of the data and the check suggested by Lee and Lemieux (2010) (inclusion of 10 gram bin dummies and jointly testing that the coefficients on these dummies are zero) indicate that linear trends on either side is a good fit of the data. Results for mortality with different window sizes and polynomials are presented in online Appendix B, Tables 8 and 15.

One concern with using birth weight as a running variable is that of heaping (Barreca et al. 2011). In Chile, birth weight tends to be recorded at 10 gram intervals and more than 93 percent of births have birth weight ending in a zero (see Figure 5). Recall that in Norway all birth weight data is only recorded in 10 gram intervals. In addition, in both countries, there appear to be heaps at 50 and 100 gram intervals. Since birth weight is observed at heaps it is natural to worry about whether irregular rounding up (or down) of the data could affect our results. In our data, rounding at 50 and 100 gram intervals is significantly correlated with demographic characteristics as shown in online Appendix B. In the birth weight window used for our analysis, the main heap of interest is at the cutoff of 1,500.

Barreca et al. (2011) suggest two ways of dealing with rounding in this context: a fixed effects approach and a "donut" RD. Following these ideas, all graphs in the paper omit data from the 1,500 gram bin and all baseline regressions control for it. In addition, tables in online Appendix B show the stability of the results when we use fixed effects for heaping at 10, 50, and 100 gram intervals. We also show the results for heaping dummies interacted with linear slopes so the effect of the heap can be different on either side of 1,500. This makes no difference to the overall results. The results are also quite stable when we simply remove points at 10, 50, and 100 gram bins, even though this decreases sample size by a significant amount. We also adopt a donut RD approach and find that our results are valid even when we exclude points that are 7 grams to either side of 1,500 grams. These results are presented in online Appendix B, Tables 5–7. Indeed, this should not be surprising since in Figure 3, it can be clearly seen that even points at 1,490 are quite different from points at 1,510. Hence, the heaped point of 1,500 grams itself is not driving our results.[29]

For a subsample of our data we can observe the exact hospital name, and note that using hospital fixed effects mitigates the correlation between rounding and demographic characteristics. This suggests that while hospitals round, the rounding is not manipulated *within* hospitals. In Norway, we can directly add hospital fixed effects to the estimation and we find that the results do not change (online Appendix A, Table 4). For Chile, while we can add hospital fixed effects for regressions that examine mortality, we are unable to do so for regressions that examine school scores. This is because we only have hospital information starting in 2002 and cohorts born after this are too young to be observed in school.

Finally, the results presented in Section V show that there is no significant discontinuity at the 1,500 gram cutoff for birth with 31 weeks of gestation once we apply the above mentioned controls. We view this as robust evidence that nonrandom heaping is not driving our results. If this were the case, it would be expected to affect births of all gestational length, not just at or above 32 weeks of gestation. We thus conclude that after applying the appropriate modifications recommended in the literature, we find that nonrandom heaping is not a significant driver of the results found.

VI. Conclusion

In this paper we provide evidence that children who receive extra medical care at birth have lower mortality rates and higher academic achievement in school. Using detailed administrative data from two countries we show that children who by virtue

[29] A concern in the Chilean context could be that the results are all driven by points that are *not* at 10 gram intervals. This is not the case.

of having been born with a birth weight of just less than 1,500 grams, are less likely to die and go on to have higher grades and test scores later in life. These results add to the growing body of research indicating the importance of early childhood care for health outcomes such as mortality. More importantly, it also provides new evidence on long-run externalities which should be considered when evaluating such policies. The results also provide suggestive evidence that the introduction of surfactant played an important role in reducing mortality and raising academic outcomes. More generally, the fact that additional medical treatment has long-run effects indicates that the observed inequalities in academic achievement and other outcomes later in life can arise at least in part due to inequalities in health care starting at birth. Efforts to improve educational outcomes should therefore focus not only on policies affecting contemporaneous educational inputs in school like better teachers, books, and school infrastructure, but also on broader public policies such as improved neonatal care.

While this paper's main contribution lies in linking early childhood medical interventions and later life educational achievement, we hope future research can highlight the pathways by which this link emerges. In this instance, children receive a "bundle" of medical interventions and although we show that surfactant likely plays a major role, understanding which intervention or what combination of interventions lead to the greatest impacts would be useful from a policy perspective. Another important avenue for future research is to better understand the way post-hospital inputs such as parental investments react to health interventions and affect long-run outcomes. Our results suggest a limited role for differential investments in this application, but we hope future research in this area can shed more light on this important behavioral response.

Finally, it is important to note that the results in this paper have been found in countries at very different stages of development (Chile and Norway) and are consistent with evidence on mortality for the United States from Almond et al. (2010). This suggests that the evidence presented on the importance of early childhood health care applies more generally to a broad set of countries at different stages of economic and social development.

REFERENCES

Almond, Douglas, and Janet Currie. 2011. "Human Capital Development before Age Five." In *Handbook of Labor Economics*, Vol. 4, Part B, edited by Orley Ashenfelter and David Card, 1315–1486. Amsterdam: Elsevier.

Almond, Douglas, Joseph J. Doyle Jr., Amanda E. Kowalski, and Heidi Williams. 2010. "Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns." *Quarterly Journal of Economics* 125 (2): 591–634.

Almond, Douglas, Hilary Hoynes, and Diane Whitmore Schanzenbach. 2010. "Childhood Exposure to the Food Stamp Program: Long-run Health and Economic Outcomes." Unpublished.

Barreca, Alan I., Melanie Guldi, Jason M. Lindo, and Glen R. Waddell. 2011. "Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification." *Quarterly Journal of Economics* 126 (4): 2117–23.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell. 2011. "Heaping-Induced Bias in Regression-Discontinuity Designs." National Bureau of Economic Research Working Paper 17408.

Behrman, Jere R. 1996. "The Impact of Health and Nutrition on Education." *World Bank Research Observer* 11 (1): 23–37.

Bharadwaj, Prashant, Katrine Vellesen Løken, and Christopher Neilson. 2013. "Early Life Health Interventions and Academic Achievement: Dataset." *American Economic Review*. http://dx.doi.org/10.1257/aer.103.5.1862.

Black, Sandra E., Paul J. Devereux, Katrine V. Løken, and Kjell G. Salvanes. 2012. "Care or Cash? The Effect of Child Care Subsidies on Academic Performance." University of Bergen Working Papers in Economics 09/12.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2007. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *Quarterly Journal of Economics* 122 (1): 409–39.

Bleakley, Hoyt. 2007. "Disease and Development: Evidence from Hookworm Eradication in the American South." *Quarterly Journal of Economics* 122 (1): 73–117.

Bratlid, Dag, and Anja Nordermoen. 2010. "Behandlingen av svært for tidlig fødte barn de siste 40 år." *Tidsskrift for Den norske legeforening* 130 (11): 1127–29.

Chay, Kenneth Y., Jonathan Guryan, and Bhashkar Mazumder. 2009. "Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon After Birth." National Bureau of Economic Research Working Paper 15078.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4): 1593–660.

Conti, Gabriella, James Heckman, and Sergio Urzua. 2010. "The Education-Health Gradient." *American Economic Review* 100 (2): 234–38.

Currie, Janet M. 2006. *The Invisible Safety Net: Protecting the Nation's Poor Children and Families*. Princeton: Princeton University Press.

Currie, Janet. 2011. "Inequality at Birth: Some Causes and Consequences." *American Economic Review* 101 (3): 1–22.

D'Angio, Carl T., Robert A. Sinkin, Timothy P. Stevens, Nancy K. Landfish, Joan L. Merzbach, Rita M. Ryan, Dale L. Phelps, et al. 2002. "Longitudinal, 15-Year Follow-up of Children Born at Less than 29 Weeks' Gestation after Introduction of Surfactant Therapy into a Region: Neurologic, Cognitive, and Educational Outcomes." *Pediatrics* 110 (6): 1094–102.

Field, Erica, Omar Robles, and Maximo Torero. 2009. "Iodine Deficiency and Schooling Attainment in Tanzania." *American Economic Journal: Applied Economics* 1 (4): 140–69.

Finne, P., A. Meberg, H. Orbeck, H. Vogt, J. Ek, I. Fuglevik, P. Scheel, E. Bertnes, and D. Bratlid. 1988. "Oppfølging av barn med fødselsvekt < 1500 gram." *Tidsskrift for legeforeningen* 108: 2254–57.

Garces, Eliana, Duncan Thomas, and Janet Currie. 2002. "Longer-Term Effects of Head Start." *American Economic Review* 92 (4): 999–1012.

Glewwe, Paul, Hanan G. Jacoby, and Elizabeth M. King. 2001. "Early Childhood Nutrition and Academic Achievement: A Longitudinal Analysis." *Journal of Public Economics* 81 (3): 345–68.

Gonzalez, Rogelio, Mario Merialdi, Ornella Lincetto, Jeremy Lauer, Carlos Becerra, René Castro, Pedro García, Ola D. Saugstad, and José Villar. 2006. "Reduction in Neonatal Mortality in Chile between 1990 and 2000." *Pediatrics* 117 (5): e949–54.

Haugen, Olav H., and Trond Markestad. 1997. "Incidence of Retinopathy of Prematurity (ROP) in the Western Part of Norway: A Population-Based Retrospective Study." *Acta Ophthalmologica Scandinavica* 75 (3): 305–07.

Havnes, Tarjei, and Magne Mogstad. 2011. "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes." *American Economic Journal: Economic Policy* 3 (2): 97–129.

Heckman, James J., and Dimitriy V. Masterov. 2007. "The Productivity Argument for Investing in Young Children." *Review of Agricultural Economics* 29 (3): 446–93.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94 (1-2): 114–28.

Hoynes, Hilary, Marianne Page, and Ann Huff Stevens. 2011. "Can Targeted Transfers Improve Birth Outcomes?" *Journal of Public Economics* 95 (7-8): 813–27.

Jiménez, Jorge, and María Inés Romero. 2007. "Reducing Infant Mortality in Chile: Success in Two Phases." *Health Affairs* 26 (2): 458.

Lee, David S. 2008. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–1102.

Lee, David S., and David Card. 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics* 142 (2): 655–74.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.

Luciana, Monica. 2003. "Cognitive Development in Children Born Preterm: Implications for Theories of Brain Plasticity Following Early Injury." *Development and Psychopathology* 15 (4): 1017–47.

Maccini, Sharon, and Dean Yang. 2009. "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall." *American Economic Review* 99 (3): 1006–26.

Markestad, Trond, and Bjørg Halvorsen. 2007. *Faglige retningslinjer for oppfølging av for tidlig fødte barn*. Sosial -og helsedirektoratet, Oslo, Norway.

Marlow, Neil, Dieter Wolke, Melanie A. Bracewell, and Muthanna Samara. 2005. "Neurologic and Developmental Disability at Six Years of Age after Extremely Preterm Birth." *New England Journal of Medicine* 352 (1): 9–19.

Meberg, A. 1988. "Kostandsanalyse av nenatalomsorg." *Tidsskrift for legeforeningen*, 108: 3261–64.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217.

Noboa-Hidalgo, Grace E., and Sergio S. Urzua. 2012. "The Effects of Participation in Public Child Care Centers: Evidence from Chile." *Journal of Human Capital* 6 (1): 1–34.

Nyfødt Intensiv, Barneavdelingen. 1998. *Metodebok i nyfdtmedisin*. Universitetssykehuset Nord-Norge.

Palomino, Maria Angélica, Mónica Morgues, and Fernando Martínez. 2005. "Management of Infants with Chronic Lung Disease of Prematurity in Chile." *Early Human Development* 81 (2): 143–49

Saugstad, Ola Didrik. 2010. "Bedre prognose for de minste for tidlig fødte." *Tidsskrift for Den norske legeforening* 130 (1): 52–54.

Saugstad, Ola Didrik, Tore Curstedt, Henry L. Halliday, Bengt Robertson, and Christian P. Speer. 2006. "Surfactant Replacement Therapy from 1986 to 2006: A 20-Year Success Story." *Biology of the Neonate* 89: 282–83.

Schady, Norbert. 2006. "Early Childhood Development in Latin America and the Caribbean." *Economia: Journal of the Latin American and Caribbean Economic Association* 6 (2): 185–213.

Schwartz, Rachel M., Anastasia M. Luby, John W. Scanlon, and Russell J. Kellogg. 1994. "Effect of Surfactant on Morbidity, Mortality, and Resource Use in Newborn Infants Weighing 500 to 1500 g." *New England Journal of Medicine* 330 (21): 1476–80.

Shigeoka, Hitoshi, and Kiyohide Fushimi. 2011. "Supply Induced Demand in Newborn Treatment: Evidence from Japan." Unpublished.

Singer, Lynn, Toyoko Yamashita, Lawrence Lilien, Marc Collin, and Jill Baley. 1997. "A Longitudinal Study of Developmental Outcome of Infants with Bronchopulmonary Dysplasia and Very Low Birth Weight." *Pediatrics* 100 (6): 987–93.

Skranes, Lars Petter, Janne Skranes, and Jon S. Skranes. 2000. "Nevrologisk oppfølging av for tidlig fødte barn og andre neonatale risikopasienter i Norge." *Tidsskrift for Den norske legeforening* 120 (11): 1332–36.

Todd, Petra E., and Kenneth I. Wolpin. 2007. "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps." *Journal of Human Capital* 1 (1): 91–136.

Umbach, David M. 2000. "Unit Conversion as a Source of Misclassification in US Birthweight Data." *American Journal of Public Health* 90 (1): 127–29.

Veramendi, Gregory, and Sergio Urzúa. 2011. "The Impact of Out-of-Home Childcare Centers on Early Childhood Development." Inter-American Development Bank, Research Department Publications 4723.

World Health Organization. 2005. *The World Health Report—Make Every Mother and Child Count*. Geneva: World Health Organization.

This article has been cited by:

1. Hilary Hoynes, Diane Whitmore Schanzenbach, Douglas Almond. 2016. Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review* **106**:4, 903-934. [Abstract] [View PDF article] [PDF with links]
2. Dora L. Costa. 2015. Health and the Economy in the United States from 1750 to the Present. *Journal of Economic Literature* **53**:3, 503-570. [Abstract] [View PDF article] [PDF with links]