Spillover Effects of Early-Life Medical Interventions[*]

N. Meltem Daysal

University of Copenhagen and IZA

Marianne Simonsen

Aarhus University and IZA

Mircea Trandafir

University of Southern Denmark and IZA

Sanni Breining

Ramboll Management Consulting

Abstract: We investigate the effects of early-life medical treatments on the treated children and their families. We use a regression discontinuity design that exploits changes in medical treatments across the very low birth weight (VLBW) cutoff. Using administrative data from Denmark, we establish that VLBW children have better health and higher test scores. We find that these benefits spill over to other family members: mothers enjoy better mental health and siblings have higher test scores. Maternal mental health improvements seem to be driven by better focal child health, and sibling spillovers by improved interactions within the family and parental compensating behavior.

JEL Classifications: I11, I12, I18, I21, J13

---

1. Introduction

An extensive body of research in economics suggests that early-life conditions have long-lasting impacts on individual well-being, including health, educational attainment and labor market outcomes (Almond and Currie, 2011; Almond et al., 2018). Growing evidence also indicates that disabled children affect the socio-economic outcomes of other family members, such as parental labor supply (Gunnsteinsson and Steingrimsdottir, 2019; Deshpande, 2016; Powers, 2003), health (Burton et al., 2008), and marital stability (Kvist et al., 2013), or sibling academic achievement (Black et al. 2017). A natural question then is whether – and by how much – interventions that improve child health also affect these spillovers in the family. In this paper, we address this question by investigating the spillover effects of early-life medical treatments on the socio-economic outcomes of other family members, focusing on the specific case of treatments provided to very low birth weight (VLBW) children, i.e., children with birth weight below 1,500 grams.

Medical interventions targeting VLBW children constitute an ideal setting to study spillover effects for several reasons. To begin with, they have been found to substantially improve the health (e.g., Cutler and Meara, 1998; Almond et al., 2010; Bharadwaj et al., 2013) as well as the academic achievement (e.g., Bharadwaj et al., 2013) of treated children. Second, although VLBW children represent a small share of all births, they account for a substantial portion of newborn health care expenditures. For example, VLBW babies in the US represent around 1.5% of all births but the neonatal intensive care unit costs associated with these babies alone account for 30% of all newborn health care costs (Johnson et al., 2013). Finally, focusing on treatments provided to VLBW children allows us to overcome identification challenges arising from potentially correlated unobservables within the family, such as shared genetic factors that affect both the receipt of medical treatments by targeted children and the outcomes of other family members. Specifically, we use a regression discontinuity design that exploits changes in medical treatments across the very low birth weight threshold to address the non-random assignment of medical treatments (Almond et al., 2010; Bharadwaj et al., 2013). We restrict our analysis to the families of focal children (defined as the children with birth weight in a small window around 1,500 grams) with gestational age above 32 weeks because the medical guidelines prescribe additional medical treatments to children with gestational age below 32 weeks regardless of their birth weight.

Using population-level data from Denmark, we first replicate the finding in the previous literature that VLBW children are significantly less likely to die within the first year of life relative to newborns with slightly higher birth weight. We also confirm the finding that these children have higher math and language test scores in 9th grade. Our estimates are remarkably comparable in magnitude to the reductions in infant mortality and improvements in test scores from the previous studies using data from the United States, Chile and Norway (Almond et al., 2010; Bharadwaj et al., 2013). We add to these studies by expanding the outcome set to include a range of common childhood disabilities (intellectual disability, attention deficit hyperactivity disorder, behavioral and emotional disorders, cerebral palsy, and epilepsy) as well as by investigating effects on hospital and emergency room (ER) visits up to 15 years after birth. Our findings suggest that children slightly below the 1,500-gram threshold have the same likelihood of a childhood disability by age 10 as compared to children slightly above the threshold. While there is no impact on child disability, the medical treatments provided to VLBW children seem to result in better health during school years as proxied by reduced hospital and ER contacts.

We next turn to investigating the spillover effects of early-life medical treatments on other family members. We examine the effects on siblings' health (hospital and ER contacts) and academic achievement (9th grade math and language test scores, enrollment beyond compulsory education). For parents, we focus on labor market outcomes (employment status, number of days worked), annual gross income, and mental health as proxied by antidepressant use. Where possible, we present the short-term effects (1-5 years after the birth of the focal child) separately from the long-term effects (6-15 years after the birth of the focal child). We try to address the potential multiple inference issues resulting from the large number of outcomes we study in two ways. First, we create indices summarizing the relevant variables in each outcome domain and time horizon, standardized at the level of the birth cohort of the focal child. Second, we adjust the p-values to take into account the multiple inference problem using a procedure proposed by Anderson (2008).

Our results suggest that early-life medical interventions have no impact on parental behavior that affects total household resources: we do not observe discontinuous changes across the VLBW cutoff in parental labor force participation or income, either in the short-run or in the long-run. This is perhaps not surprising given that Denmark has one of the most generous social safety nets in the world, including 52 weeks of parental leave, generous cash payments to families with children, heavily subsidized childcare, and free and universal health care. Our results instead

2

suggest that early-life medical interventions may improve intra-household interactions. In particular, we find that the mothers of VLBW focal children are significantly less likely to use antidepressants soon after the focal child is born and that these mental health gains diminish as the focal child ages. We provide evidence suggesting that both focal child survival and improved focal child health are important drivers of these mental health gains.

Our results also suggest that early-life medical treatments have substantial positive spillovers on sibling test scores. We find that the siblings of VLBW newborns have 9th grade test scores that are on average 0.375 standard deviations higher. Our results further highlight that test scores gains are observed across skill domains: language and math test scores are higher by 0.386 and 0.255 standard deviations, respectively. These results are economically large, corresponding to a reduction of 33-69% in the test score gap between the children in the top and bottom 10% of the income distribution. We show that these sibling spillovers are unlikely to be driven by correlated health shocks within the family that would result in siblings' exposure to early-life medical treatments themselves, or by differential focal child survival at the cutoff. Instead, our evidence points to improved intra-family interactions as a potential channel behind the sibling spillovers.

Finally, we provide indirect evidence on the role of parental reallocation of resources within the family. If the production of human capital exhibits dynamic complementarities, then parental investments have higher returns for children with high initial endowment than for children with low initial endowment. If parents engage in compensating behavior, then our results suggest that the siblings of VLBW focal children should receive more parental resources than the siblings of focal children with birth weight slightly above 1,500 grams. We find that the siblings of VLBW focal children experience higher test score gains if they have high initial endowments themselves (as proxied by their birth weight), a pattern consistent with compensating behavior by their parents, especially in the presence of dynamic complementarities in human capital accumulation.

Our paper makes three contributions. First, we add to the studies that document spillover effects of child health. The majority of this research examines the effects of having a disabled child on parental outcomes, such as labor supply (Gunnsteinsson and Steingrimsdottir, 2019; Deshpande, 2016; Powers, 2003) and health (Burton et al., 2008). One notable exception is Black et al. (2017), who investigate spillovers to siblings and find that the second child in a family has worse test

scores when the third child is disabled.[1] Our paper documents that improvements in child health that do not operate through child disability status may still result in significant sibling spillovers.

Second, we contribute to the economic literature on the returns to early-life medical interventions. These studies almost exclusively investigate the effects on treated children (e.g., Cutler and Meara, 1998; Chay et al., 2009; Field et al., 2009; Almond et al., 2010; Bharadwaj et al., 2013; Daysal et al., 2015; Hjort et al., 2017; Bütikofer et al., 2019; Daysal et al., 2019).[2] The main insight that emerges from our work is that medical treatments may have far-reaching effects on family well-being through spillovers even in developed countries with generous welfare systems.

Third, our results speak to the economic literature that relies on sibling fixed effects models to account for unobserved heterogeneity across households in estimating the effects of various exposures. To the extent that siblings have spillovers on each other, sibling fixed effects models would not estimate the true treatment effects.

Our results are also pertinent to the ongoing discussions about the cost effectiveness of early-life medical treatments. During the past few decades, medical spending for the very young increased substantially faster than spending for the average individual. For example, US annual spending on individuals aged 1 to 64 increased by 4.7 percent between 1960-1990, while per capita spending on infants under 1 year old increased by 9.8 percent per year (Cutler and Meara, 1998). Technological innovations are widely considered the main driver of this medical cost growth, both in general and in the specific case of early-life treatments (Newhouse, 1992; Cutler and Meara, 1998). As medical expenditures keep increasing, understanding the benefits of early-life medical interventions becomes even more important. Our finding that medical treatments for VLBW children have positive externalities on other family members indicates that conventional calculations understate the net benefits of these treatments.

---

[1] There is also evidence on sibling spillovers more generally. For example, Dahl et al. (2014) show that take-up of family friendly policies affects siblings' subsequent use of these policies, Joensen and Nielsen (2018) and Nicoletti and Rabe (2019) find that siblings' education choices and test scores causally affect younger siblings' academic outcomes, and Altonji et al. (2017) and Heissel (forthcoming) document that older siblings' risky behavior impact younger siblings' outcomes.

[2] One exception is Adhvaryu and Nyshadham (2016), who examine the effects of a large-scale iodine supplementation program in Tanzania on parents' investments in children and find that the siblings of treated children were more likely to be immunized.

2. Institutional Background

The majority of Danish health care services, including birth related procedures, are free of charge and all residents have equal access (Danish Ministry of Health and Prevention, 2008). The first European neonatal intensive care unit was established in 1965 at Rigshospitalet in Denmark and the use of early-life medical technologies has since followed the international development (Mathiasen et al., 2008). Danish neonatal medicine textbooks pay particular attention to VLBW children (i.e., children weighing less than 1,500 grams, regardless of gestational age) and very premature newborns (i.e., those with a gestational age less than 32 weeks, regardless of birth weight). These birth weight and gestational age classifications are frequently found in medical research papers based on Danish data where the focus is often on their higher mortality rates (e.g., Thomsen et al., 1991; Hertz et al., 1994). Medical handbooks suggest courses of treatment based on either birth weight or gestational age (Schiøtz and Skovby, 2001). Specific recommendations in terms of nutrition and vitamin supplements exist for VLBW children (Peitersen and Arrøe, 1991). In addition, papers indicate that children below 1,500 grams or born before 32 weeks of gestation are more likely to receive additional treatments such as cranial ultrasound (Greisen et al., 1986), antibiotics (Topp et al., 2001), prophylactic treatment with nasal continuous positive airway pressure, prophylactic surfactant treatment and high priority of breast feeding, and use of the kangaroo method (Jacobsen et al., 1993; Verder et al., 1994; Verder, 2007; Mathiasen et al., 2008).

Anecdotal evidence from hospital and regional specific notes also outline special services that are provided to families with children below 1,500 grams or below 32 weeks of gestational age. These services include referrals to a physiotherapist who guides and instructs parents on how to stimulate the development of the child and on various baby exercises. It is also mentioned that all children below 1,500 grams or below 32 weeks of gestational age are routinely checked 1-2 months after discharge and again when they are five months, one year and two years old.[3]

3. Empirical Strategy

Identification of the (spillover) effects of early-life health interventions is complicated by the non-random assignment of medical treatments. In particular, there may be unobserved determinants of the outcomes of other family members that are correlated with the receipt of medical treatments

---

[3] Unfortunately, our data does not include any information on specific early-life treatments.

by targeted children, such as shared genetic factors. In order to address this endogeneity, we follow Almond et al. (2010) and Bharadwaj et al. (2013) and use a regression discontinuity design that exploits changes in medical treatments across the VLBW threshold. Specifically, we estimate:

$$y_{ijt} = f(bw_j - 1500) + \beta VLBW_j + \delta X_{ijt} + \epsilon_{ijt} \qquad (1)$$

where y_{ijt} is an outcome of family member i of focal child j at time t after the birth of the focal child, bw_j is the birth weight of focal child j, $f(\cdot)$ is a first-degree polynomial in distance to the VLBW cutoff that is allowed to differ on both sides of the cutoff, $VLBW_j$ is an indicator for focal child j having very low birth weight (i.e., $bw_j < 1500$), and X_{ijt} is a vector of covariates.[4]

We start our analysis by replicating and extending the findings in the previous literature on the impact of medical technologies on focal children themselves, i.e., we set $i = j$ in Equation (1). We then turn to effects on other family members. The parameter of interest, β, is an intention-to-treat estimate of the effects that additional medical treatments received by VLBW newborns may have on themselves and on their families.

Our baseline regressions use a triangular kernel that assigns decreasing weights to observations farther away from the cutoff. We choose our bandwidth based on a rule-of-thumb procedure suggested by Calonico et al. (2014), which yields optimal bandwidths between 118 grams and 251 grams with an average of 189 grams (see Appendix Table A2). We choose 200 grams as our preferred bandwidth to ensure that newborns on either side of the VLBW cutoff are nearly identical. This bandwidth is the same as the one used by Bharadwaj et al. (2013) for Norwegian data and reflects the relatively small number of observations available in Denmark and Norway. The vector of covariates, X_{ijt}, includes indicators for heaping at multiples of 50 grams in all specifications unless mentioned otherwise (Barreca et al., 2011).[5] Some of our robustness checks additionally control for child and family characteristics (see Section 4 below).

We are interested in exploring a variety of outcomes across multiple domains for several family members. This gives rise to a multiple inference problem: we may estimate statistically significant

---

[4] An alternative strategy would rely on the 32-week cutoff for gestational age. This strategy is infeasible with our data because gestational age is recorded in full weeks.

[5] Given that birth weight is measured in grams, heaping is generally symmetric around our cutoff point and hence our strategy is less likely to be affected by the criticism raised by Barreca et al. (2011). Indeed, we show in Section 5.3 that our results are robust to the exclusion of these controls.

effects of the VLBW status on some outcomes simply by chance. We address this issue in two ways following the recommendations of Anderson (2008). First, we create indices for each family member and each domain (see Section 4 below and Appendix Table A1 for details). While this procedure helps with reducing the dimensionality, we are still left with a relatively large number of indices. Hence, as a second step, we adjust the *p*-values to take into account the multiple inference problem based on a procedure proposed by Anderson (2008).[6]

Finally, we construct robust confidence intervals following Calonico et al. (2014, 2019). These confidence intervals are centered on bias-corrected estimates instead of the usual (conventional) estimates and use the standard errors from a specification with a higher-order polynomial in the running variable, which in our case is a second-degree polynomial. Therefore, in addition to the coefficient estimates and their robust standard errors, we also report the bias-corrected estimates that are used to construct these robust confidence intervals.

4. Data

Our key data set is the Birth Register, which includes information about the universe of births in Denmark starting from 1970. For each child, the data includes information on the exact date of birth, gender, and plurality. Birth weight is recorded in 250-gram intervals between 1973-1978, in 10-gram intervals in the period 1979-1990, and at the gram level since 1991. Gestational age is added beginning in 1982. Using parental identifiers, we are able to link children to their parents and siblings and determine parity. We also link this data to other register data that provide information on both parents and children regarding demographic characteristics, labor market outcomes, health outcomes, and academic achievement.

We first use data on focal children to investigate whether early-life medical interventions impact focal child health and academic achievement. Our mortality index includes two previously studied short-term outcomes, 28-day and 1-year mortality. In addition, we construct two health indices. Our short-term health index uses indicators for being hospitalized during each year between the ages of 1-5, while our long-term health index uses separate indicators for being hospitalized and

---

[6] The false discovery rate (FDR) is the average fraction of true null hypotheses among the rejected hypotheses, and the *q*-value is the level of the FDR desired by the researcher (Benjamini and Hochberg, 1995; Benjamini et al., 2006). We report the lowest *q*-value, i.e., the lowest sharpened FDR, at which an estimated effect is still significant (see Anderson, 2008, for details). This is conceptually similar to a *p*-value in that it represents the probability of a type-I error.

for having an ER visit during each year between the ages of 6-15. Given the previous medical literature linking very low birth weight to child developmental disabilities (Schieve et al., 2016), we also construct a disability index based on separate indicators for being diagnosed by age 10 with one of the following conditions: intellectual disability, attention deficit hyperactivity disorder, behavioral and emotional disorders, cerebral palsy, and epilepsy. Our first measure of human capital accumulation is a test-score index based on course-specific test scores from 9th grade qualifying exams in reading and math, available between 2001 and 2010.[7] Finally, we create an index of enrollment beyond compulsory education (9 years during our sample period) using indicators for enrollment in high school or vocational school at age 18, enrollment in an academic track at age 18, enrollment in higher education at age 24, and enrollment in a university at age 24.

We then turn to spillover effects on the family. The outcomes for siblings mirror the outcomes for focal children with the exception of mortality and disability diagnosis. In particular, we create a short-term health index using indicators for being hospitalized during each year when the focal child is 1-5 years old, and a long-term health index using indicators for being hospitalized or having an ER visit during each year when the focal child is 6-15 years old. We construct a test-score index based on 9th grade math and language test scores and an index of enrollment beyond compulsory education based on sibling's enrollment in (higher) education at ages 18 and 24.

For parents, we focus on mental health and labor market outcomes, separately for mothers and fathers. Our two mental health indices are based on indicators for having filled at least one antidepressant prescription during each year when the focal child is 2-5 and 6-15 years old.[8] For labor market outcomes, we study effects on employment and income. We create two employment indices for each parent based on the number of days worked as well as indicators for being employed in each calendar year when the focal child is 1-5 and 6-15 years old. We similarly create income indices based on the log real annual gross income over the same range of focal child age.

We construct each index in two steps. We first standardize each variable by the birth cohort of the focal child such that it has a mean of zero and a standard deviation of one. Next, we take the average of the standardized variables that make up the index and, because some of these variables are correlated, we re-standardize the index at the level of the birth cohort of the focal child.

---

[7] All exams are graded by the teacher and by an external examiner, who can overrule the teacher.
[8] The prescription drug register begins recording data from 1995 so we can only construct measures of antidepressant use starting from focal child age two.

Appendix Table A1 lists each variable included in the construction of each index. We provide results using selected outcomes or alternative aggregation strategies in Appendix Tables A6-A8.

Some of our checks use focal child characteristics (gestational age, indicators for gender, parity, plurality, birth year, and birth region), maternal characteristics at the birth of the focal child (age, years of education, indicators for marital status and immigrant status), and sibling characteristics (gestational age, birth weight, and indicators for gender, parity, plurality, and birth year).[9]

We define the analysis sample in several steps (see Appendix Table A3). First, we select focal children born between 1982 and 1993.[10] We then exclude observations for which either birth weight or gestational age are missing and restrict the sample to those with birth weight within 1,300-1,700 grams. Given that we are particularly interested in sibling spillovers, we further restrict the sample to the 3,677 focal children with siblings.[11] As discussed in Section 2, newborns with a gestational age of less than 32 weeks are always covered by the medical guidelines for receiving additional medical interventions, irrespective of their VLBW classification. Since there is no discontinuity in eligibility for medical treatments (Bharadwaj et al., 2013), we do not expect to observe a discontinuity in focal child outcomes or in the outcomes of their family members. Therefore, we use the 1,521 focal children with gestational age below 32 weeks and their families only in a falsification check, and from here on we focus exclusively on the 2,156 focal children with gestational age of at least 32 weeks (hereafter the *FC sample*) and their families.

Parents are identified from the birth register. Our data includes parental identifiers for all the mothers. If the mother is married to a man at the time of birth, authorities automatically register the husband as the biological father. When the mother is unmarried, the biological father needs to claim paternity of the child. Parental identifiers for the fathers are missing for only 40 of the focal children in the *FC sample*. Thus, the *parent sample* virtually overlaps with the FC sample.

---

[9] Maternal education is missing for 315 observations corresponding to 154 mothers. We replace these with the median number of years of education by birth cohort and include an indicator for imputed maternal education. Our results (available upon request) are robust to excluding these observations or the indicator for missing mother's education.

[10] Our sample includes focal children born after 1982, when both birth weight and gestational age are recorded. We include cohorts born before 1994 for two reasons. First, this allows us to study human capital accumulation information for all cohorts, which makes it possible to compare the effects of early-life health interventions on the focal children in our context to those in previous studies. Second, evidence suggests that medical guidelines around the VLBW cutoff are less likely to be binding in recent years (see, for example, footnote 20 in Bharadwaj et al., 2013).

[11] The results for the sample including the 922 focal children who have no siblings born within our sample period are quantitatively and qualitatively similar (available upon request).

Siblings are defined as children born to the same mother from different pregnancies. We include both older and younger siblings because the receipt of additional medical treatments around the VLBW cutoff does not seem to impact future fertility decisions.[12] We focus on siblings who are old enough for us to observe their academic outcomes. Tests are administered when children are around 15-16 years old, so data on test scores are available for cohorts of siblings born between 1986-1997. Enrollment outcomes are measured at ages 18 and 24 and include siblings born between 1970-1993. The resulting sample includes 3,311 siblings of focal children with gestational age of at least 32 weeks (the *sibling sample*).[13]

5. Results

5.1. Tests of the Validity of the Regression Discontinuity Design

The validity of an RD design rests on the assumption that individuals do not have precise control over the assignment variable. Since women cannot precisely predict the birth weight of their children, the variation in birth weight near the VLBW cutoff is plausibly as good as random (Almond et al., 2010; Bharadwaj et al., 2013). The key identification assumption of the RD design could be violated if physicians systematically misreport birth weight, especially in the presence of financial incentives for manipulation (Shigeoka and Fushimi, 2014; Jürges and Köberlein, 2015).

In order to test this assumption, we examine the frequency of births by birth weight within our bandwidth around the cutoff. Appendix Figure A1 plots the distribution of observations in the FC/parent sample and in the sibling sample by birth weight of the focal child. We use 10-gram

---

[12] A focal child may have more than one sibling. We treat each sibling-focal child pair as an independent observation. This is not a concern for our identification because parity and total family size are relatively smooth across the cutoff in the FC sample. In addition, we find no evidence of a discontinuity at the cutoff when we examine the probability of having a younger sibling, the number of younger siblings, and the birth spacing between focal children and younger siblings (see Table 1). Finally, our results are qualitatively similar when we cluster the standard errors at the mother level in order to correct for the bias in standard errors caused by the potential correlation in the error terms between pairs of siblings from the same household (available upon request).

[13] Test scores are missing for approximately 20% of the eligible cohorts in the sibling sample. This is because children can be exempted from taking the test if, for example, they have a documented disability. This could be a concern if medical treatments provided to focal children impact the test-taking of siblings. We find no discontinuity at the cutoff in the probability that a sibling takes the language test (estimate 0.029, bias-corrected estimate b.c.e. 0.051, s.e. 0.070, mean 0.808) or the math test (estimate 0.048, b.c.e. 0.069, s.e. 0.068, mean 0.804), or in the age when they take the test (estimate -0.139, b.c.e. -0.106, s.e. 0.120, mean 16.035). Enrollment information is available for all eligible cohorts.

bins because birth weight is reported in 10-gram intervals for most of our sample period.[14] Similar to previous studies (Almond et al., 2010; Bharadwaj et al., 2013), we observe reporting heaps at multiples of 50 and 100 grams but there is no evidence of irregular heaping around the VLBW cutoff. We check this more formally by estimating a local-linear regression similar to our baseline model, using the number of births in each birth weight bin as the dependent variable (McCrary, 2008). We do not find any evidence of a discontinuity in the frequency of births at the VLBW cutoff.[15] These results suggest that birth weight is unlikely to be manipulated in our context.

In the remainder of this section, we check if there are differences in observable characteristics across the VLBW cutoff by estimating our baseline model with the covariates as dependent variables. If the RD design is valid, then there should be no discontinuities at the VLBW cutoff.[16] Table 1 provides the results. Panels A, C and D use the FC/parent sample and check whether focal child and parental characteristics are balanced, while Panels B uses the sibling sample to check for discontinuities in the covariates of siblings. Column 1 provides the conventional point estimate from the local-linear regression. Columns 2 and 3 present the bias-corrected estimate and the robust standard error (Calonico et al., 2019). We report sharpened q-values, i.e., p-values based on the numbers in Columns 2 and 3 further corrected for multiple inference, in Column 4 (see section 3 and Anderson, 2008, for details). Finally, Column 5 reports the mean of the covariate in the sample of (family members of) focal children with birth weight of at least 1,500g. The results show that observations just below the VLBW cutoff are generally similar to those just above the VLBW cutoff. There are few characteristics with marginally significant discontinuities at the threshold and none of these survive the adjustment for multiple inference: the lowest q-value is 0.5.

Overall, the analyses in this section indicate that there is no evidence of manipulation of the running variable around the VLBW cutoff or of discontinuities in the observable characteristics of focal children, their parents and their siblings.

---

[14] Some degree of rounded running variables is common across studies relying on regression discontinuity designs; a prominent example is to use age in quarters (e.g., Card et al., 2008) or years (e.g., Oreopoulos, 2006). While rounding can, of course, cause discretization bias (Dong, 2015), we believe this is a minor issue in our case with access to relatively fine-grained data.

[15] The estimates corresponding to Appendix Figures A1(a)-(b) are 0.092 (b.c.e. -7.507, standard error s.e. 6.955) and 0.196 (b.c.e. -12.614, s.e. 17.429). The results are qualitatively similar when using the logarithm of the number of births as the dependent variable instead. In this case, the estimated coefficients are 0.013 (b.c.e. -0.223, s.e. 0.188) and 0.027 (b.c.e. -0.238, s.e. 0.324).

[16] Visual evidence on selected covariates is provided in Appendix Figures A2-A4.

5.2. Baseline Results

Figure 1 provides visual evidence on the relationship between birth weight and the outcomes of focal children. Figures 1(a)-1(d) focus on child health, while Figures 1(e)-(f) focus on academic achievement. The Figure shows that focal children with birth weight slightly lower than 1,500 grams have lower mortality than children who weigh slightly more than 1,500 grams. Conditional on survival, however, the short-term health of children seems to be similar across the VLBW cutoff. We also do not observe any discontinuity in the disability index in Figure 1(d) but there is some indication of improved long-term health during primary school-age years from the index based on hospital admissions and ER visits in Figure 1(c). Turning to academic achievement, Figure 1(e) shows that focal children with birth weight slightly lower than 1,500 grams have visibly higher test scores in 9th grade. The improved test scores, however, do not seem to result in higher enrollment beyond compulsory education (Figure 1(f)).

Column 1 in Tables 2 and 3 presents the corresponding regression results from our baseline model. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the summary index. Bias-corrected estimates are listed in square brackets, robust standard errors in brackets, and sharpened *q*-values in curly braces. We also report the mean of the outcome for observations above the cutoff. Consistent with the graphical evidence, we find that VLBW newborns have lower mortality in the short-run and lower hospital and ER visits in the long run: mortality is 0.508 standard deviations lower (Table 2) and hospitalizations and ER visits during school years are 0.324 standard deviations lower (Table 3).[17] We also find that focal children just below the VLBW cutoff have on average 0.314 standard deviations higher test scores in the 9th grade but they are not significantly more likely to be enrolled beyond compulsory education.

We next turn to spillover effects on the siblings. Figure 2 provides visual evidence while the corresponding regression results are presented in Column 2 of Tables 2 and 3. Figure 2 shows that the siblings of focal children with birth weight slightly lower than 1,500 grams have visibly higher test scores in 9th grade. On the other hand, there is no evidence of important spillovers on health or enrollment outcomes. The regression results confirm that the early-life medical treatments

---

[17] As an example, the 95% robust confidence interval for the mortality effect (-0.508) is constructed using the bias-corrected estimate and the robust standard error as: -1.011 ± 0.408 * 1.96 = [-1.811, -0.211].

offered to VLBW children have significant positive spillovers on the test scores of the siblings without gains in higher education or health outcomes. In particular, we find that siblings of VLBW newborns have 9th grade test scores that are on average 0.375 standard deviations higher.[18]

Finally, in Figures 3-4 and in the remaining columns of Tables 2-3 we examine potential spillovers to parental outcomes. Figure 3 suggests that the mothers of VLBW newborns have potentially better mental health, as proxied by reduced antidepressant use, than the mothers of heavier babies. On the other hand, there are no apparent discontinuities in maternal labor market outcomes or income, and surely not immediately after the birth of the focal child. Figure 4 similarly shows that the distributions of paternal labor market outcomes and income are generally smooth across the VLBW cutoff. However, the fathers of VLBW newborns do not seem to experience the mental health improvements enjoyed by the mothers. The corresponding regression results reported in Columns 3-4 of Tables 2-3 confirm the visual evidence. We generally do not find significant discontinuities at the VLBW cutoff in the measures of family resources (parental labor market outcomes and income). In the few cases when we find marginally significant gains, the results do not survive the adjustment for multiple inference: the lowest q-value is 0.213. However, we do find evidence of improved maternal mental health soon after the birth of the focal child that dissipates as the child ages. In particular, our results indicate that antidepressant use by the mothers of VLBW newborns is on average 0.347 standard deviations lower. Consistent with the visual evidence in Figure 4, we find no evidence of a similar effect for the fathers.

5.3. Robustness Checks

In this section we present robustness checks for the indices that were statistically significant in the baseline regressions and that survived the adjustment for multiple inference: focal child mortality, focal child long-term health, focal child test scores, sibling test scores, and maternal short-term health (Appendix Figures A6-A8 and A10-12, and Appendix Table A5 provide the corresponding checks for the remaining indices). Appendix Figure A5 and Column 1 of Appendix Table A4 investigate the robustness of our estimates to the choice of bandwidth. Appendix Figure A5 presents the results for all bandwidths between 100-300 grams in 10-gram steps. Our baseline effect is indicated with a square and the vertical bars plot the corresponding 95% robust confidence

---

[18] Among the test-takers in the sibling sample, the maximum age difference between older siblings and focal children is 7.6 years, meaning that none of the older siblings take the test before the focal children are born.

interval following Calonico et al. (2014, 2019). The Figure shows that the magnitudes of the estimates are remarkably consistent across different bandwidths. In Column 1, we allow the bandwidths to differ across outcomes using the optimal bandwidths suggested by the Calonico et al. (2014) strategy. Given the stability of the estimates to alternative bandwidths, it is not surprising that the results are again very robust.

We next check the sensitivity of our results to the choice of degree of polynomial in birth weight. The results in Column 2 show that our findings are robust to using a second degree of polynomial. Column 3 investigates the sensitivity of the results to the inclusion of the control variables described in Section 4. If the key assumption in our RD design is satisfied (i.e., birth weight is as good as random around the cutoff), then including additional relevant covariates should not impact the estimates much but increase precision instead. The results show that this is generally the case.

Columns 4-5 turn to the role of heaping. Heaping can lead to biased estimates if it does not occur in a symmetric way around the cutoff. Following Barreca et al. (2011, 2016), our main specification controls for heaping at 50-gram intervals. We conduct two checks to probe this further. First, we estimate models excluding the heaping dummies (Column 4). Second, in Column 5 we estimate "donut" regressions that exclude the (family members of) focal children who weighed 1,500 grams (Barreca et al., 2016). The results are again similar to the main estimates, suggesting that our baseline results are not driven by heaping.

Our baseline model uses a triangular kernel. In Column 6, we show that our findings are robust to using a rectangular kernel that places equal weights to each observation. Column 7 checks the sensitivity of our inference by clustering standard errors at the birth weight level and confirms that the results remain statistically significant at conventional levels.

Finally, we conduct two falsification tests. First, we estimate our baseline model in the sample of (family members of) focal children with a gestational age of less than 32 weeks. Since these children are eligible to receive additional medical treatments regardless of their birth weight, any discontinuity in their outcomes or in the outcomes of their family members would suggest a violation of the key identification assumptions underlying the RD design. The results in Column 8 indicate that the indices studied are relatively smooth across the VLBW threshold in this sample.

Second, we check whether we observe similar discontinuities in the indices at other points in the distribution of birth weight of the focal child. If the observed gains are indeed driven by the medical

treatments received by focal children, then we should not observe systematic discontinuities in the outcomes at other potential cutoffs. We examine cutoffs from 1,300 grams to 3,100 grams, keeping the bandwidth fixed at 200 grams. The results presented in Appendix Figure A9 indicate that the discontinuities observed at 1,500 grams are indeed distinct. Although the effects at 1,300 grams are more noisily estimated, it is clear in most figures that the largest and only significant discontinuity is found at 1,500 grams.[19] Overall, these findings strongly suggest that the observed (spillover) effects are due to the impact of medical treatments provided to the VLBW focal children.

5.4 Discussion

In the first part of our paper, we confirm the findings in the previous literature that early-life medical treatments have significant effects on focal child survival and academic achievement. In order to compare our findings with the previous literature, we present the results using selected components of the mortality and the test-score indices in Appendix Table A6. We show that the probability of death within the first 28 days (1 year) of life is 4.1 (5.4) percentage points lower among VLBW newborns. These are large gains when compared to the average mortality rates of those above the cutoff (6.2 and 7.7 percent, respectively) but they are comparable in magnitude to the reductions in infant mortality from previous studies: 1 percentage point (mean: 5.5 percent) in the US (Almond et al., 2010); 4.5 percentage points (mean: 11 percent) in Chile and 3.1 percentage points (mean: 3.6 percent) in Norway (Bharadwaj et al., 2013). We find that VLBW newborns have language and math test scores higher on average by 0.229 and 0.315 standard deviations, respectively.[20] The estimated effect on math test scores is comparable to those found by Bharadwaj et al. (2013), who estimate effects of 0.152 standard deviations in Chile and 0.476 standard deviations in Norway. We add to this literature by investigating effects on focal child human capital accumulation beyond compulsory schooling as well as disability and hospital/ER contacts. We do not find effects on enrollment beyond compulsory schooling. Further analyses based on individual components of the index shows that there is also no impact on the "intensive" margin,

---

[19] In the case of focal child and sibling test scores, we also find marginally significant discontinuities at 2,500 grams, which is another birth weight cutoff for specialized medical treatments. These effects are three to four times smaller than the estimated effects at 1,500 grams.

[20] These results are not driven by delayed school entry as proxied by the age at which focal children take the 9th grade test (Landersø et al., 2017). Indeed, we find that the distribution of the age when focal children take the test is smooth across the VLBW threshold (estimate -0.033, b.c.e. -0.115, s.e. 0.127, mean 16.137).

as the share of focal children enrolled in an academic track at age 18 is not significantly higher among VLBW children (see Appendix Table A6). In order to reconcile the test score gains with the lack of effects on enrollment, we estimate a quantile RD specification based on the method proposed by Frandsen et al. (2012). The results presented in Appendix Table A9 indicate that early-life medical treatments do not improve the test scores of the focal children at the bottom of the test score distribution. This suggests that the children affected by the medical treatments are not on the margin of dropping out or of making the choice between an academic or a vocational track. While we also do not find any effects on focal child disability status, our results consistently point to health improvements during school years, highlighting a potential channel behind the academic achievement results.[21]

The main novelty in our paper is the investigation of spillover effects to other family members. Our results suggest that early-life medical interventions have little impact on parents' decisions that affect total household resources, as we see no discontinuity in the labor market or income indices of either mothers or father. Since Denmark is a developed country with a particularly generous social safety net, this is maybe not surprising. We do, however, find evidence that early-life treatments provided to VLBW children improve maternal mental health. The mental health gains are short-lived and tend to dissipate as focal children age. One natural explanation for this pattern could be improved focal child survival. We check whether this is the case in two ways. First, we estimate our baseline model in the sample of family members of focal children who survive past the first year of life. The estimated effect shown in Column 1 of Appendix Table A10 is smaller than our baseline estimate, suggesting that child survival is indeed a channel through which early-life medical treatments may affect maternal mental health. Second, we consider several groups of focal children whom we expect to be affected differently by early-life medical interventions. For example, twins are on average lighter than singletons, meaning that a VLBW singleton may be in poorer health than a VLBW twin. Therefore, we would expect the VLBW singletons in our sample to benefit more from medical treatments than twins. In addition, there is a large medical literature on the "fragile male" hypothesis, which states that the male fetus is simply more at risk than the female fetus (Naeye et al., 1971). Hence, we would expect that the

---

[21] In order to shed some light on this, we estimate our baseline model with the focal child test score index as the outcome while additionally controlling for focal child long-term health index. The estimated coefficient of the *VLBW* variable in this case is 0.260 (b.c.e. 0.496, s.e. 0.233).

VLBW boys in our sample benefit more than girls from the additional treatments provided to VLBW children. Indeed, Columns 2-5 in Appendix Table A10 indicate that our mortality results are driven by singletons and by boys. However, there are long-term improvements in health across all four groups, supporting our earlier conclusion that early-life medical treatments have health benefits beyond just survival. More importantly, the Table shows large improvements in the mental health of the mothers of focal children with no survival benefits (twins and girls). This suggests that general improvements in the health of focal children, and not just their survival, is a channel behind the spillover effects on maternal mental health.

Turning to siblings, we find that early-life medical interventions have economically significant long-run gains in sibling academic achievement. The results based on course-specific test scores presented in Appendix Table A8 show that the test score gains are driven by both math and language test scores. Siblings of VLBW newborns have on average 0.386 (0.255) standard deviations higher language (math) test scores relative to the siblings of newborn who weigh slightly more than 1,500 grams. One way to gauge the magnitudes of these effects is to compare it to other policy-relevant test score gaps. For example, among all children born during the period covered by our sibling sample, the difference in language (math) scores between the children of non-immigrants and immigrants is 0.264 (0.404) standard deviations. Our results imply that medical interventions are equivalent to eliminating the language disadvantage for children of immigrants and reducing the gap in math scores by more than half. We also calculate that the difference in language (math) test scores among those born in households above the 90th income percentile and those born in households below the 10th income percentile is 0.557 (0.769) standard deviations. Our coefficients imply that medical interventions can reduce the income-based test score gap at age 16 by 33-69%. These effects are in line with those found by Duncan and Sojourner (2003) for income-based test score gaps at ages 3 through 8 for children exposed to an early-education program targeting low-birth-weight children in the US.

Similar to the focal child results, we find that siblings' test score gains do not translate into a higher likelihood of pursuing education beyond the compulsory level. In contrast to the effects on focal children, however, the results from the quantile RD approach indicate that early-life medical treatments improve the test scores of siblings across the entire test score distribution (see Appendix Table A9). This seems to still not affect the children on the margin of dropping out, likely due to the fact that only 22% of students do not continue beyond compulsory schooling. However, as

more students from the lower half of the test score distribution see improvements in their academic achievement, we now find effects at the "intensive" margin: siblings of VLBW children are more likely to follow an academic track instead of a vocational track (see Appendix Table A8).

These positive spillovers on academic achievement are unlikely to be driven by correlated health shocks within the family that make siblings themselves more likely to receive medical interventions early in life. The fact that we do not observe discontinuities in the hospital/ER visits of siblings at the cutoff offers the first evidence that this is unlikely to be the case. We conduct two more checks to shed more light on this issue. First, we exclude VLBW siblings and confirm that our main results are not driven by them.[22] Second, if the families of VLBW focal children are more prone to having health shocks than the families of slightly heavier children, then our human capital achievement results may capture the effects of these unobserved family traits instead of the spillovers from early-life medical interventions. In that case, we may expect to see differences across the VLBW cutoff in the short-term survival rates of older siblings before being exposed to the VLBW focal children. Using the 28-day and 1-year mortality rate of older siblings as outcomes suggests that this is not a concern in our context.[23] The sibling spillovers are also not driven by differential focal child survival at the cutoff. In contrast to mother's mental health, we find similar improvements in sibling test scores to our baseline results when we estimate the baseline model in the sample of families where focal children survive the first year (Column 1 Appendix Table A10).

The fact that we observe significant test score gains among siblings of VLBW children without effects on total household resources or sibling health suggests that early-life medical treatments may be changing intra-family interactions or the intra-household allocation of resources. Data limitations do not allow us to investigate these hypotheses directly, but we provide indirect evidence on both. It is well understood in economics that the family, especially parent-child interactions, plays a central role in the human capital accumulation of children (Cunha and Heckman, 2007; Cunha et al., 2010; Almond and Currie, 2011). In order to further shed light on this, we estimate our baseline model while controlling for the maternal short-term mental health

---

[22] The estimated coefficient of the *VLBW* variable is 0.401 (b.c.e. 0.466, s.e. 0.205, N=1,456) for the language test score and 0.263 (b.c.e. 0.416, s.e. 0.183, N=1,465) for the math test score. After excluding VLBW siblings, only 10 siblings with a gestational age below 32 weeks remain in the sample. Further dropping these from the sample does not change the results.

[23] The estimated coefficient of the *VLBW* variable is 0.016 (b.c.e. 0.027, s.e. 0.020, N = 3,594) for 28-day mortality and 0.021 (b.c.e. 0.031, s.e. 0.022, N = 3,594) for 1-year mortality.

index and find that this reduces the estimated effect on sibling text scores by about 50%.[24] This suggests that improved parent-child relations may be important for sibling academic achievement.

The existing research also indicates that children's early-life health endowments impact the academic outcomes of other children in the family by changing parental investments. Evidence of such spillover effects are found both in developing and in developed countries and the magnitudes of the effects are economically large (Yi et al., 2015; Black et al., 2017). We can provide some indirect evidence on this if we make the assumption that there are dynamic complementarities in the production of human capital, as suggested by Cunha and Heckman (2007). In this case, children with high initial endowments would benefit most from parental investments because "skills beget skills." To illustrate, consider two children with low initial endowment, A and B, who are identical in every respect except that A has a sibling with birth weight slightly below the VLBW cutoff while B has a sibling with birth weight slightly above the cutoff. If both sets of parents engage in compensating behavior, then child B has more resources taken away from her and allocated to her sibling than child A does (because the VLBW sibling of child A benefits from the additional medical treatments). Therefore, in the long-term child B ends up with a lower level of skills than child A. Now consider a similar pair of identical children, C (who has a VLBW sibling) and D (who does not), but with high initial endowment. Just as before, child D has more resources taken away from her and so she ends up with a lower level of skills in the long term than child C. However, because of dynamic complementarities, child D is harmed even more by the fewer resources she receives because the return to those resources would be higher for her than for child B. Therefore, the difference in skills between children C and D (high initial endowment) is larger than the difference in skills between children A and B (low ability).[25]

To check whether we observe this pattern in our data, we rely on birth weight as an indicator of initial endowments because the previous literature finds that it is highly correlated with later-life academic, health, and labor market outcomes (e.g., Black et al., 2007; Figlio et al., 2014). We define "high endowment" siblings as those whose birth weight is higher than the birth weight of the median child born during our sample period. The results, shown in Appendix Table A11,

---

[24] The estimated coefficient of the *VLBW* variable is 0.182 (b.c.e. 0.437, s.e. 0.250, N = 546).

[25] Alternatively, the difference in skills between children C and D can be larger than the difference in skills between children A and B if parents reallocate more resources to the focal child in order to compensate for the larger difference in endowments within the family. This explanation is also consistent with compensating behavior by the parents.

suggest that high-endowment siblings benefit more than low-endowment siblings from the additional medical treatments received by VLBW focal children. This suggests that parental compensating behavior (possibly combined with dynamic complementarities in the production of human capital) may also be one of the factors behind the observed spillover effects.

6. Conclusions

This article investigates the spillover effects of early-life medical treatments provided to VLBW children on other family members. Using register data from Denmark, we confirm the findings in the previous literature that VLBW children eligible to receive early-life treatments are less likely to die in the first year of life and have higher academic achievement in 9th grade. We add to this literature by showing that focal children's likelihood of having a childhood disability is not impacted by early-life medical treatments, but they are still more likely to enjoy better health during school years, as proxied by reduced hospital/ER contacts.

The main innovation in our study is that we document the presence of spillover effects to other family members. While total household resources do not differ between the families of VLBW children and the families of slightly heavier newborns, the mothers of VLBW children are significantly less likely to consume antidepressants soon after the birth of the children. These health gains diminish as the focal child ages. Our results suggest that both increased focal child survival and improvements in focal child health contribute to the improvements in maternal mental health. Turning to siblings, we find that the siblings of focal children who were slightly below the VLBW cutoff have better 9th grade language and math test scores. These gains are not driven by correlated health shocks within the family or by differential focal child survival at the cutoff. Instead, we present evidence suggesting that improved interactions in the family and parental compensating behavior may be important reasons behind the spillovers to siblings.

Our results underscore the importance of health interventions targeted to other family members as an important factor in the accumulation of human capital. Our findings also have important implications for understanding the efficacy of early-life medical interventions. In particular, they underline the need to consider potential externalities when assessing the net benefits of medical treatments. Finally, our results have implications for studies on the effects of early-life health endowments using sibling fixed-effects estimators. The fact that we find substantial positive spillovers on the siblings of treated children suggests that within-sibling comparisons of

achievement gains may underestimate the true impact of initial health endowments on later-life outcomes.

References

Adhvaryu, Achyuta, and Anant Nyshadham. 2016. "Endowments at Birth and Parents' Investments in Children." *Economic Journal* 126 (593): 781-820.

Almond, Douglas, and Janet Currie. 2011. "Human Capital Development before Age Five." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card, Vol. 4B. Amsterdam; New York; New York, N.Y., U.S.A.: North Holland.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature* 56 (4): 1360–1446.

Almond, Douglas, Joseph J. Doyle, 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.

Altonji, Joseph G., Sarah Cattan, and Iain Ware. 2016. "Identifying Sibling Influence on Teenage Substance Use." *Journal of Human Resources* 52 (1): 1-47.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495.

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): 2117–2123.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell. 2016. "Heaping-Induced Bias in Regression-Discontinuity Designs." *Economic Inquiry* 54 (1): 268–293.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. 2006. "Adaptive Linear Step-up Procedures That Control the False Discovery Rate." *Biometrika* 93 (3): 491–507.

Bharadwaj, Prashant, Katrine Vellesen Løken, and Christopher Neilson. 2013. "Early Life Health Interventions and Academic Achievement." *American Economic Review* 103 (5): 1862–1891.

Black, Sandra E, Sanni Breining, David N. Figlio, Jonathan Guryan, Krzysztof Karbownik, Helena Skyt Nielsen, Jeffrey Roth, and Marianne Simonsen. 2017. "Sibling Spillovers." NBER Working Paper No. 23062.

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.

Burton, Peter, Lynn Lethbridge, and Shelley Phipps. 2008. "Children with Disabilities and Chronic Conditions and Longer-Term Parental Health." *Journal of Socio-Economics* 37 (3): 1168–86.

Bütikofer, Aline, Katrine V. Løken, and Kjell Salvanes. 2019. "Infant Health Care and Long Tern Outcomes," *Review of Economics and Statistics* 101 (2): 341–354.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Roció Titiunik. 2019. "Regression Discontinuity Designs Using Covariates." *Review of Economics and Statistics* 101 (3): 442–451.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Card, David, Carlos Dobkin, and Nicole Maestas. 2008. "The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare". *American Economic Review* 98 (5): 2242–2258.

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." NBER Working Paper No. 15078.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47.

Cunha, Flavio, James Heckman, and Susanne Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78 (3): 883–931.

Cutler, David M., and Ellen Meara. 1998. "The Medical Costs of the Young and Old: A Forty-Year Perspective." In *Frontiers in the Economics of Aging*, edited by David A. Wise, 215–46. University of Chicago Press.

Dahl, Gordon, Katrine V. Løken, and Magne Mogstad. 2014. "Peer Effects in Program Participation." *American Economic Review* 104 (7): 2049–2074.

Danish Ministry of Health and Prevention [Ministeriet for Sundhed og Forbyggelse]. 2008. "Health Care in Denmark."

Daysal, N. Meltem, Mircea Trandafir, and Reyn van Ewijk. 2015. "Saving lives at birth: The impact of home births on infant outcomes." *American Economic Journal: Applied Economics* 7 (3): 1-24.

Daysal, N. Meltem, Mircea Trandafir, and Reyn van Ewijk. 2019. "Low-risk isn't no-risk: Perinatal treatments and the health of low-income newborns." *Journal of Health Economics* 64: 55–67.

Deshpande, Manasi. 2016. "Does Welfare Inhibit Success? The Long-Term Effects of Removing Low-Income Youth from the Disability Rolls." *American Economic Review* 106 (11): 3300–3330.

Dong, Yingying. 2015. "Regression Discontinuity Applications with Rounding Errors in the Running Variable". *Journal of Applied Econometrics* 30 (3): 422–446.

Duncan, Greg J. and Aaron J. Sojourner. 2013. "Can Intensive Early Childhood Intervention Programs Eliminate Income-Based Cognitive and Achievement Gaps?" *Journal of Human Resources* 48 (4): 945–68.

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

Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth. 2014. "The Effects of Poor Neonatal Health on Children's Cognitive Development." *American Economic Review* 104 (12): 3921–3955.

Frandsen, Brigham R., Markus Frölich, and Blaise Melly. 2012. "Quantile Treatment Effects in the Regression Discontinuity Design." *Journal of Econometrics* 168 (2): 382–95.

Greisen, G, M. B. Petersen, S. A. Pedersen, and P. Bekgaard. 1986. "Status at Two Years in 121 Very Low Birth Weight Survivors Related to Neonatal lntraventricular Haemorrhage and Mode of delivery". *Acta Paediatrica Scandinavica* 75: 24-30.

Gunnsteinsson, Snaebjorn and Herdis Steingrimsdottir. 2019. "The Long-Term Impact of Children's Disabilities on Families." Mimeo.

Heissel, Jennifer A. Forthcoming. "Teen Fertility and Siblings' Outcomes: Evidence of Family Spillovers Using Matched Samples." *Journal of Human Resources*.

Hertz, Birgitte, Eva-Bettina Holm, and Jørgen Haahr. 1994. "Prognosen for børn med meget lav fødselsvægt i Viborg Amt." *Ugeskrift for læger* 156 (46): 6865−6868.

Hjort, Jonas, Mikkel Sølvsten, and Miriam Wüst. 2017. "Universal Investment in Infants and Long-run Health: Evidence From Denmark's 1937 Home Visiting Program." *American Economic Journal: Applied Economics* 9 (4): 78-104.

Jacobsen, Thorkild, John Grønvall, Sten Petersen, and Gunnar Eg Andersen. 1993. "'Minitouch' Treatment of Very Low-birth-weight Infants." *Acta Paediatrica Scandinavica* 82: 934-938.

Joensen, Juanna and Helena Skyt Nielsen. 2018. "Spillovers in Educational Choice." *Journal of Public Economics* 157: 158–183.

Johnson, T. J., Patel, A. L., Jegier, B. J., Engstrom, J. L., and Meier, P. P. 2012. "Cost of Morbidities in Very Low Birth Weight Infants." *The Journal of Pediatrics* 162 (2): 243-249e1.

Jürges, Hendrik, and Juliane Köberlein. 2015. "What explains DRG upcoding in neonatology? The roles of financial incentives and infant health." *Journal of Health Economics* 43: 13-26.

Kvist, Anette Primdal, Helena Skyt Nielsen, and Marianne Simonsen. 2013. "The Importance of Children's ADHD for Parents' Relationship Stability and Labor Supply." *Social Science & Medicine* 88 (July): 30–38.

Landersø, Rasmus, Helena Skyt Nielsen, and Marianne Simonsen. 2017. "School Starting Age and the Crime-Age Profile." *Economic Journal* 127: 1096-1111.

Mathiasen, René, Bo M. Hansen, Anne Løkke, and Gorm Greisen. 2008. " Behandling af tidligt fødte børn på Rigshospitalet i perioden 1955–2007" [Treatment of Preterm Children at Rigshospitalet During the Period 1955-2007]. *Bibliotek for Læger* 200 (4): 528–546.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714.

Naeye, Richard L., Leslie S. Burt, David L. Wright, William A. Blanc, and Dorothy Tatter. 1971. "Neonatal Mortality, the Male Disadvantage." *Pediatrics* 48 (6): 902–906.

Newhouse, Joseph P. 1992. "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives* 6 (3): 3–21.

Nicoletti, Cheti and Birgitta Rabe. 2019. "Sibling spillover effects in school achievement." *Journal of Applied Econometrics* 34 (4): 482–501.

Oreopoulos, Philip. 2006. "Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter." *American Economic Review* 96 (1): 152–175.

Peitersen, Birgit and Mette Arrøe. 1991. *Neonatologi – Det raske og det syge nyfødte barn*. Nyt Nordisk Forlag Arnold Busck.

Powers, Elizabeth T. 2003. "Children's Health and Maternal Work Activity." *Journal of Human Resources* 38 (3): 522–556.

Shigeoka, Hitoshi, and Kiyohide Fushimi. 2014. "Supplier-Induced Demand for Newborn Treatment: Evidence from Japan." *Journal of Health Economics* 35 (May): 162–178.

Schieve, Laura A., Lin H. Tian, Kristin Rankin, Michael D. Kogan, Marshalyn Yeargin-Allsopp, Susanna Visser, and Deborah Rosenberg. 2016. "Population Impact of Preterm Birth and Low Birth Weight on Developmental Disabilities in US Children." *Annals of Epidemiology* 26 (4): 267–274.

Schiøtz, Peter Oluf and Skovby, Flemming. (2001). *Praktisk pædiatri*, Munksgaard Danmark.

Thomsen, Ketty Dahl, Helle Hansen, Finn Ebbesen, and Vibeke Jacobsen. 1991. "Neonatal mortalitet hos børn med meget lav fødselsvægt i Nordjyllands Amt: en retrospektiv opgørelse." *Ugeskrift for læger* 153 (47): 3310−3313.

Topp, Monica, Peter Uldall, and Gorm Greisen. 2001. "Cerebral Palsy Births in Eastern Denmark, 1987–90: Implications for Neonatal Care." *Paediatric and Perinatal Epidemiology* 15: 271–277.

Verder, Henrik. 2007. "Nasal CPAP has become an indispensable part of the primary treatment of newborns with respiratory distress syndrome." *Acta Pædiatrica* 96: 482–484.

Verder, Henrik, Bengt Robertson, Gorm Greisen, Finn Ebbesen, Per Albertsen, Kaare Lundstrøm, and Thorkild Jacobsen. 1994. "Surfactant Therapy and Nasal Continuous Positive Airway Pressure for Newborns with Respiratory Distress Syndrome." *New England Journal of Medicine* 331 (16): 1051-1055.

Yi, Junjian, James Heckman, Junsen Zhang, and Gabriella Conti, 2015. "Early Health Shocks, Intra-Household Resource Allocation and Child Outcomes." *Economic Journal* 125: F347-F371.
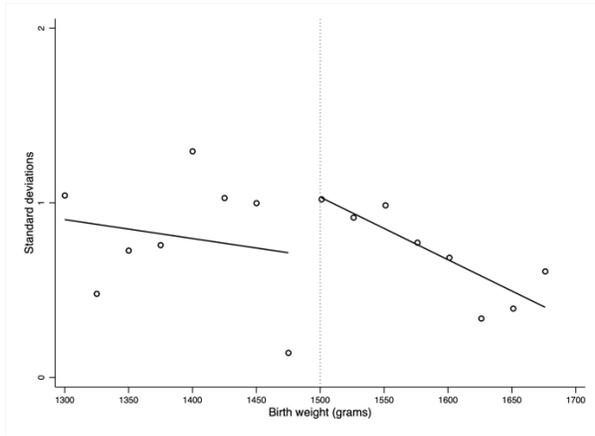
(a) Mortality
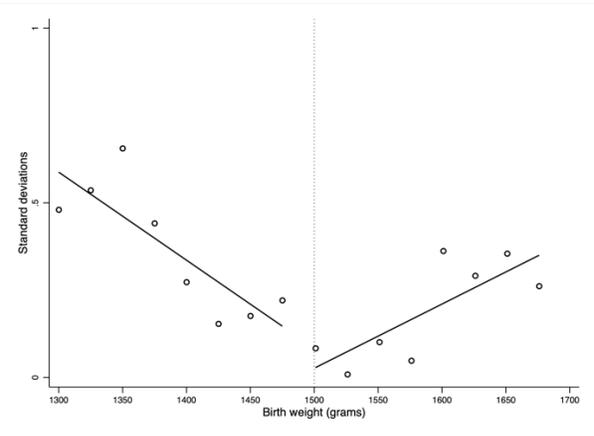
(b) Short-term health
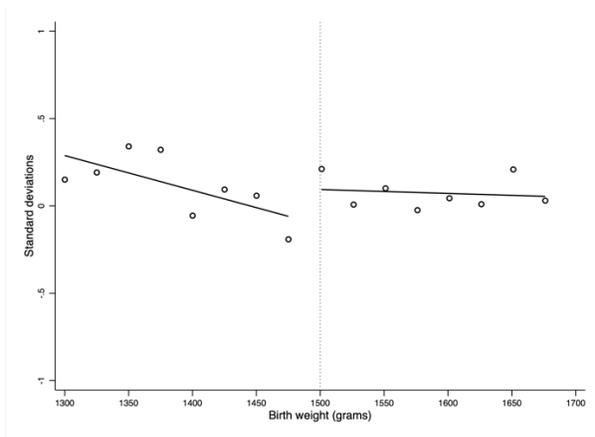
(c) Longer-term health
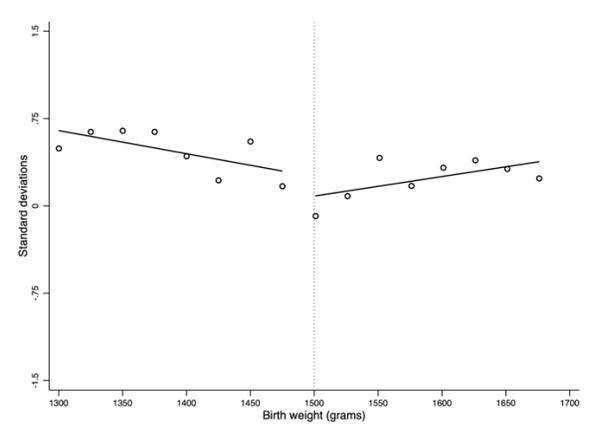
(d) Disability
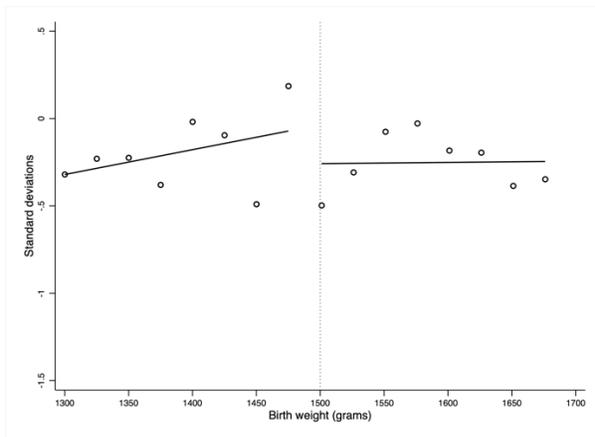
(e) Test scores
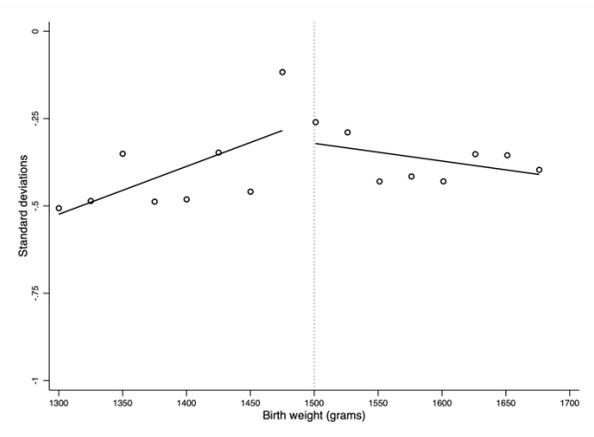
(f) Higher education

Notes: Sample of focal children with gestational age of at least 32 weeks. Each dot represents the average of the summary index indicated in the panel for a 40g bin. Focal children with birth weight of 1,500g are excluded. The lines plot a first-degree polynomial estimated separately on either side of the VLBW cutoff.

Figure 1: Evolution of summary indices of focal children around the VLBW cutoff
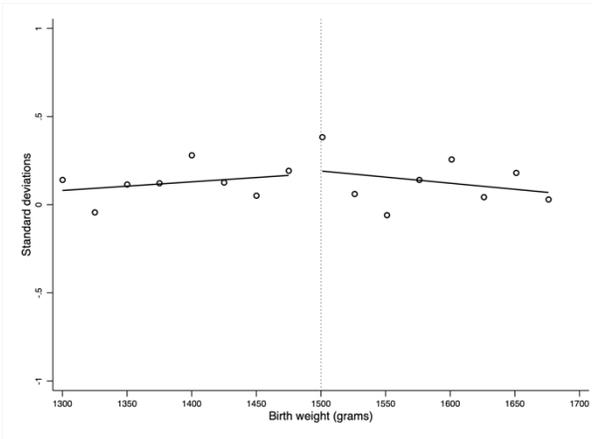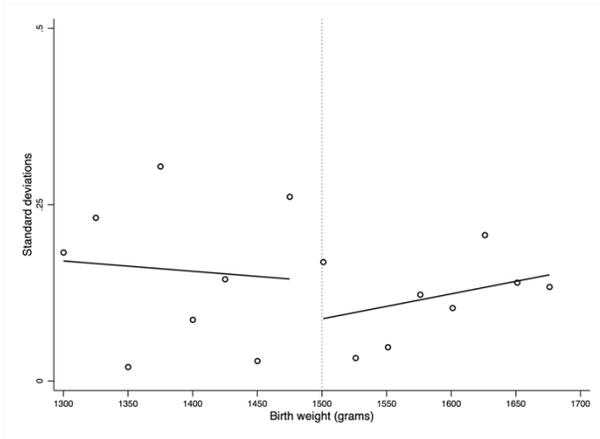
(a) Short-term health   (b) Long-term health
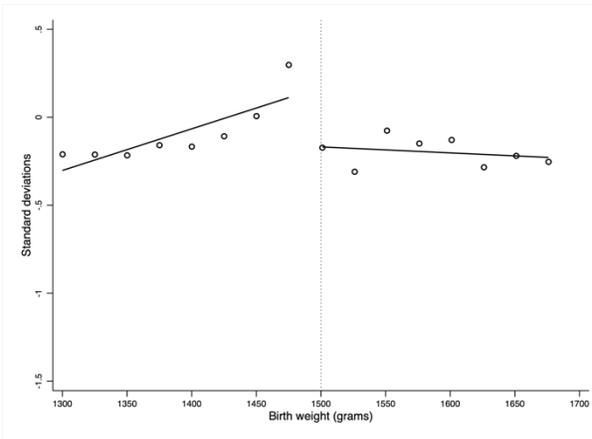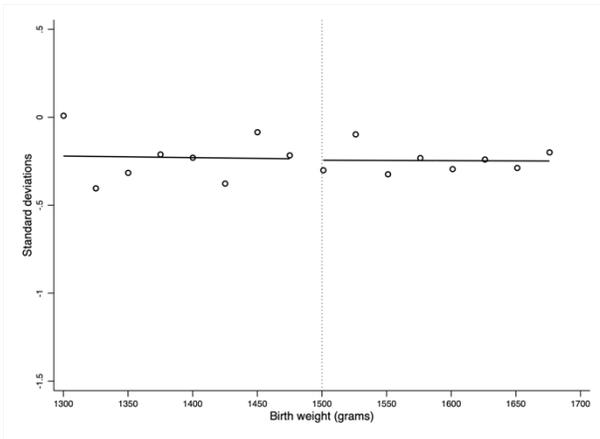
(c) Test scores   (d) Higher education

Notes: Sample of siblings of focal children with gestational age of at least 32 weeks. Each dot represents the average of the summary index indicated in the panel for a 40g bin. Siblings of focal children with birth weight of 1,500g are excluded. The lines plot a first-degree polynomial estimated separately on either side of the VLBW cutoff.

Figure 2: Evolution of summary indices of siblings around the VLBW cutoff, GA32+ sample
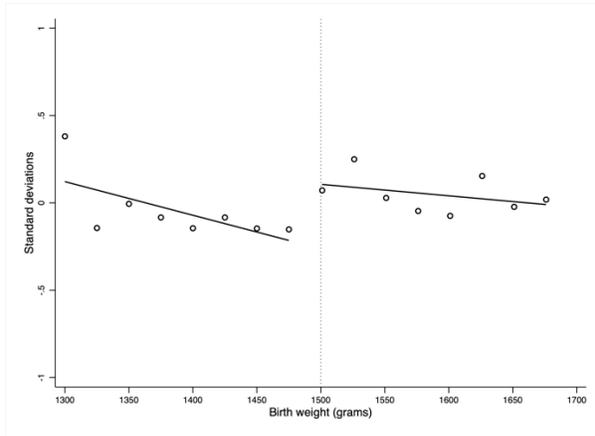
(a) Short-term health
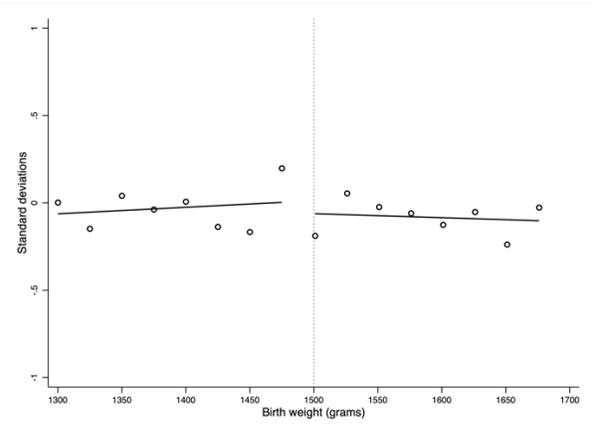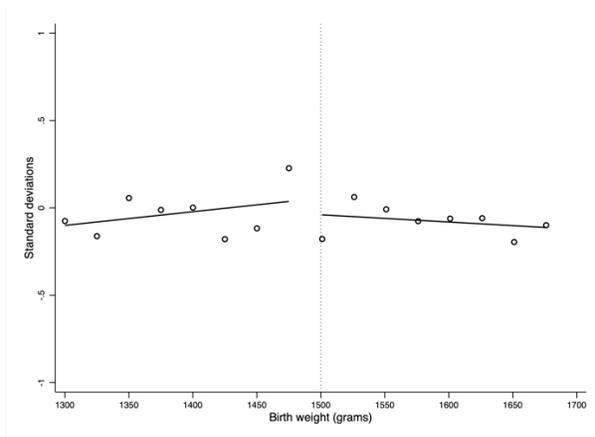
(b) Short-term labor market

(c) Short-term income
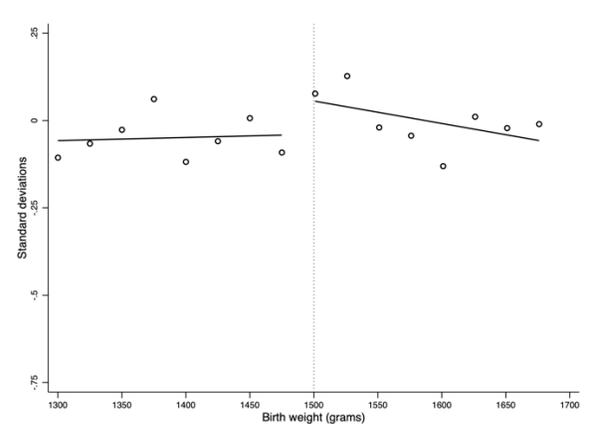
(d) Longer-term health
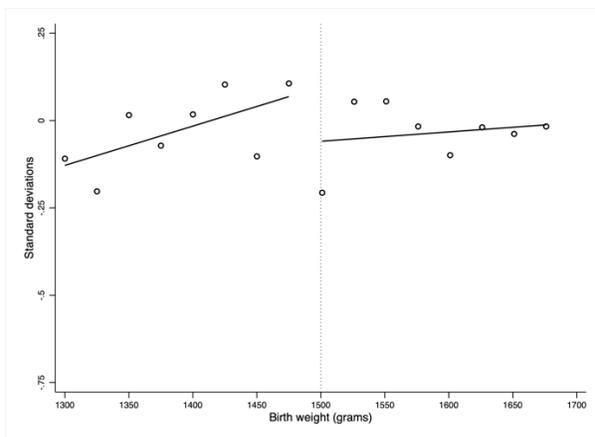
(e) Longer-term labor market
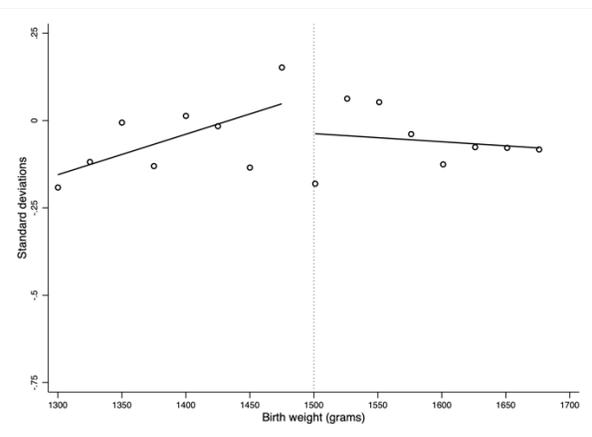
(f) Longer-term income

Notes: Sample of mothers of focal children with gestational age of at least 32 weeks. Each dot represents the average of the summary index indicated in the panel for a 40g bin. Mothers of focal children with birth weight of 1,500g are excluded. The lines plot a first-degree polynomial estimated separately on either side of the VLBW cutoff.

Figure 3: Evolution of summary indices of mothers of focal children around the VLBW cutoff
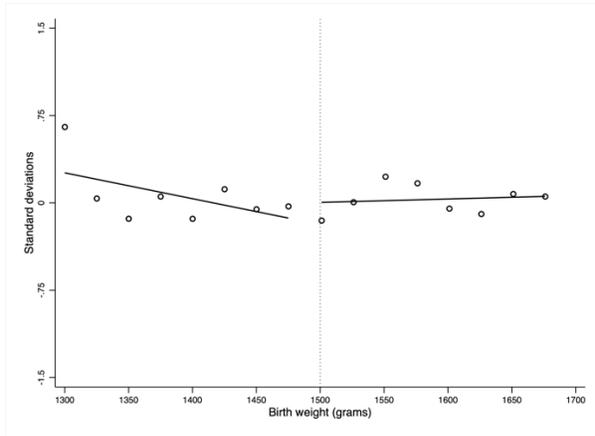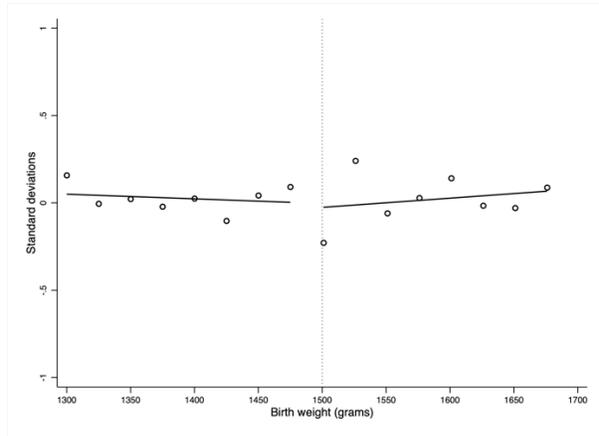
(a) Short-term health

(b) Short-term labor market
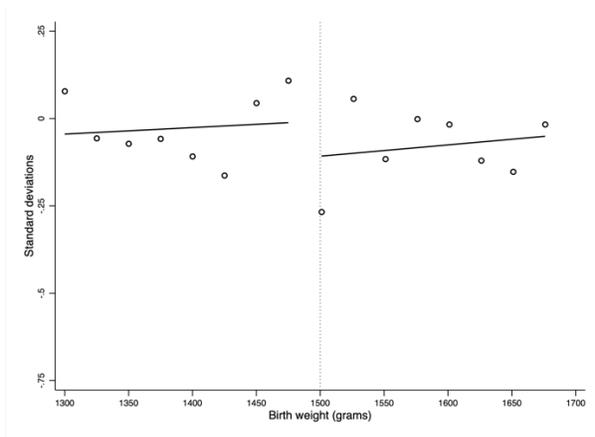
(c) Short-term income
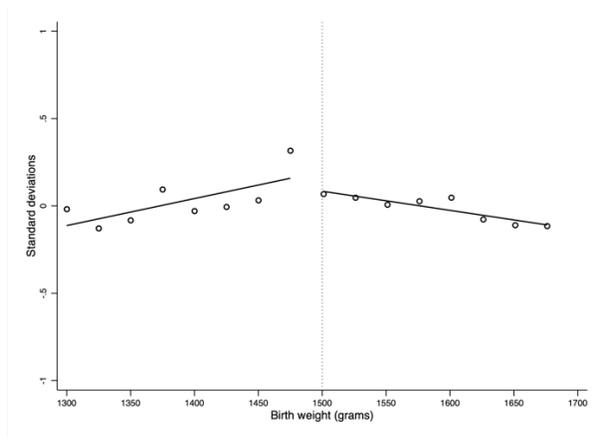
(d) Longer-term health
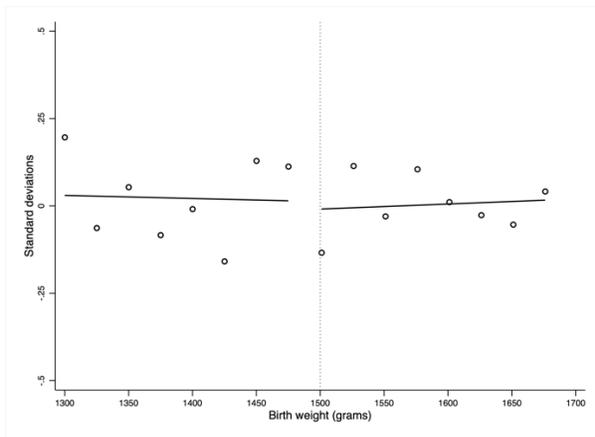
(e) Longer-term labor market
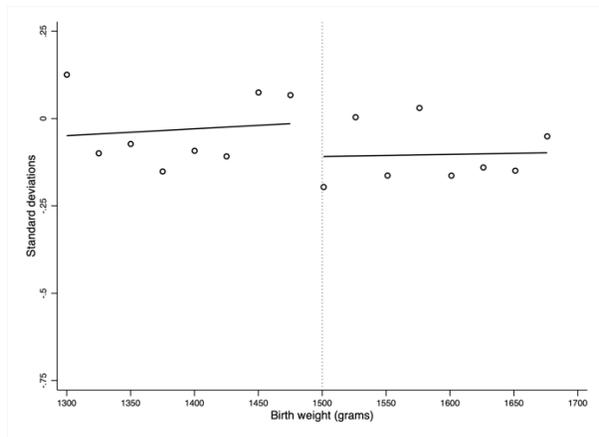
(f) Longer-term income

Notes: Sample of fathers of focal children with gestational age of at least 32 weeks. Each dot represents the average of the summary index indicated in the panel for a 40g bin. Fathers of focal children with birth weight of 1,500g are excluded. The lines plot a first-degree polynomial estimated separately on either side of the VLBW cutoff.

Figure 4: Evolution of summary indices of fathers of focal children around the VLBW cutoff

Table 1. Distribution of covariates across the VLBW cutoff

	Estimate	Bias-corrected estimate	Robust standard error	Sharpened q-value	Mean of dependent variable
	(1)	(2)	(3)	(4)	(5)
A. Focal child characteristics (N = 2,156)					
Boy	-0.028	[-0.079]	(0.077)	{0.919}	0.456
Birth order	0.229	[0.166]	(0.173)	{0.919}	1.911
Multiple birth	0.065	[0.092]	(0.070)	{0.851}	0.208
Gestational age	-0.353**	[-0.508]	(0.258)	{0.576}	34.097
Family size	0.054	[0.009]	(0.160)	{1.000}	2.937
Has younger siblings	-0.066	[-0.020]	(0.078)	{1.000}	0.611
Number of younger siblings	-0.189	[-0.205]	(0.133)	{0.713}	0.899
B. Sibling characteristics (N = 3,311)					
Boy	-0.003	[-0.033]	(0.068)	{1.000}	0.520
Birth order	-0.115	[-0.154]	(0.147)	{0.919}	2.121
Multiple birth	0.026	[0.011]	(0.017)	{1.000}	0.023
Gestational age	-0.319	[-0.464]	(0.449)	{0.919}	38.209
Birth weight	-128.494*	[-188.938]	(105.751)	{0.618}	2,898.7
VLBW	0.012	[0.019]	(0.033)	{1.000}	0.046
Age difference - older sibling	-0.119	[-0.397]	(0.782)	{1.000}	6.586
Age difference - younger sibling	-0.400	[-0.691]	(0.449)	{0.713}	4.515
C. Mother's characteristics at the birth of the focal child (N = 2,156)					
Age	1.118	[1.040]	(0.800)	{0.851}	27.735
Education (years)	-0.246	[0.218]	(0.389)	{1.000}	11.239
Immigrant	-0.021**	[-0.052]	(0.027)	{0.576}	0.068
Married	0.047	[0.003]	(0.080)	{1.000}	0.535
D. Father's characteristics at the birth of the focal child (N = 2,116)					
Age	2.044**	[2.132]	(0.873)	{0.507}	30.547
Education (years)	0.172	[0.465]	(0.400)	{0.919}	11.608
Immigrant	0.013	[-0.006]	(0.039)	{1.000}	0.074
Not reported	-0.008	[-0.004]	(0.023)	{1.000}	0.021

Notes: Sample of (family members of) focal children with birth weight within a 200g bandwidth around the 1,500g cutoff and gestational age of at least 32 weeks. Column 1 reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the characteristic listed in the row for the family member indicated in the panel heading. Column 2 reports the corresponding bias-corrected estimate, Column 3 the robust standard error (Calonico et al., 2014, 2019), Column 4 the sharpened *q*-value for the set of tests included in the Table (Anderson, 2008), and Column 5 the mean of the variable in the row calculated among (family members of) focal children with birth weight above 1,500g. Stars indicate statistical significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on the bias-corrected estimates (for details, see Calonico et al., 2014, 2019).

Table 2. Short-term effects of VLBW classification.

	Focal child (1)	Siblings (2)	Mother (3)	Father (4)
Mortality	-0.508** [-1.011] (0.408) {0.065}	---	---	---
Mean outcome	0.726			
Observations	2,156			
Health	0.133 [0.161] (0.190) {0.415}	-0.049 [-0.041] (0.142) {0.471}	-0.347*** [-0.338] (0.124) {0.052}	-0.016 [0.098] (0.173) {0.415}
Mean outcome	0.190	0.103	0.042	0.023
Observations	1,978	3,311	689	669
Labor market outcomes	---	---	0.068 [0.211] (0.150) {0.293}	0.066 [0.205] (0.156) {0.303}
Mean outcome			-0.060	0.051
Observations			2,143	2,099
Income	---	---	0.109* [0.286] (0.153) {0.213}	0.112* [0.289] (0.171) {0.239}
Mean outcome			-0.070	-0.048
Observations			2,144	2,100

Notes: Sample of (family members of) focal children with birth weight within a 200g bandwidth around the 1,500g cutoff and gestational age of at least 32 weeks. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the summary index listed in the row for the family member indicated in the column (see Appendix Table A1 for details on the construction of the summary indices). All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets, robust standard errors in brackets, and sharpened *q*-values in curly braces below the coefficient estimates. The mean of the outcome is reported for (family members of) focal children with birth weight above 1,500g. Stars indicate statistical significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on the bias-corrected estimates (for details, see Calonico et al., 2014, 2019).

	Focal child	Siblings	Mother	Father
	(1)	(2)	(3)	(4)
Health	-0.324***	0.090	-0.168	0.233
	[-0.437]	[0.167]	[-0.199]	[0.427]
	(0.144)	(0.155)	(0.122)	(0.293)
	{0.052}	{0.332}	{0.239}	{0.293}
Mean outcome	0.039	0.115	-0.011	-0.026
Observations	1,960	3,311	2,155	2,116
Disability diagnosis by age 10	0.234	---	---	---
	[0.246]			
	(0.294)			
	{0.415}			
Mean outcome	0.249			
Observations	2,156			
9th grade test scores	0.314**	0.375***	---	---
	[0.564]	[0.524]		
	(0.234)	(0.193)		
	{0.065}	{0.052}		
Mean outcome	-0.257	-0.216		
Observations	952	1,539		
Higher education	0.044	0.048	---	---
	[0.146]	[0.086]		
	(0.180)	(0.141)		
	{0.415}	{0.415}		
Mean outcome	-0.380	-0.211		
Observations	2,156	2,759		
Labor market outcomes	---	---	0.093	0.080*
			[0.098]	[0.252]
			(0.153)	(0.152)
			{0.415}	{0.239}
Mean outcome			-0.028	0.037
Observations			2,124	2,074
Income	---	---	0.076	0.095
			[0.171]	[0.214]
			(0.139)	(0.149)
Mean outcome			{0.332}	{0.293}
Observations			-0.063	-0.068

Table 3. Long-term effects of VLBW classification.

Notes: See the notes in Table 2.