Early childhood health inequalities and in-utero health interventions: evidence from the treatment of gestational diabetes*

Gabriella Conti[1] and Paul Rodríguez-Lesmes[2]

[1]Department of Economics and Department of Social Science, University College London
[2]School of Economics, Universidad del Rosario

February 15, 2021

Abstract

For years, diagnosis and treatment for gestational diabetes mellitus (GDM) has been an unresolved question. Current guidelines have established diagnosis and treatment standards, but the question remains on the potential impacts on child development via in-utero exposure to diabetes medications. We use uniquely rich data from the Born-in-Bradford cohort study, in which all pregnant women underwent universal screening for GDM based on clear thresholds over oral glucose tolerance tests (OGTT), rather than using pre-existing conditions as a screening criterion (as it is the standard practice in England and elsewhere). To evaluate the effects of GDM treatment, we use the OGTT cutoff within a sharp regression discontinuity design, so we compare children from pregnancies with a positive and negative GDM diagnosis near the cutoff. We find that diagnosis and subsequent treatment of GDM resulted in improved outcomes at birth, in particular a significant reduction in the odds of being a macrosomic newborn. We also detect child BMI differences postnatally, but these fade out after 3 years of age; additionally, at five years of age, children of mothers diagnosed with GDM in pregnancy have reduced blood pressure. At age five, however, children of non-diagnosed mothers outperform those of GDM mothers in the Early Years Foundation Stage Profiles standardised test, a nationally-administered measure of school readiness. Hence, we confirm that treatment for GDM, as shown in the medical literature, improves outcomes of children at birth. However, we present for the first time evidence of potential adverse outcomes of treating GDM on children academic achievement five years later. Therefore, studies on long-run consequences of these GDM treatments on child development are urgently required.

1 Introduction

Hyperglycaemia on pregnant women is associated with adverse perinatal outcomes. It is associated with higher odds of induction of labour and caesarean section delivery, and newborns to be large for gestational age, macrosomia, shoulder dystocia, birth injury or stillbirth (Farrar et al., 2016). The HAPO study (Metzger et al., 2008) challenged any doubt about the importance of screening and diagnosing diabetes during pregnancy. Since then, health agencies around the world have adjusted their pregnancy guidelines. Screening is typically based on an oral glucose tolerance test (OGTT) consisting of two blood samples, one at fasting and another two hours after taking a substance equivalent of 75 grams (g) of anhydrous glucose. If blood glucose concentration is above certain cutoffs, and the pregnant women have no pre-existing diabetes, she is diagnosed with gestational diabetes mellitus (GDM). However, evidence from Tarry-Adkins et al. (2019) shows that sugar deprivation in the blood for the fetus, which can happen with treatments for the condition such as prescription of metformin, resulted on the potential increase on the odds overweight in mid-childhood. Moreover, there is little evidence of the long-run impacts on children that resulted by treating the condition (Durnwald and Werner, 2018), and caution about treatment has been expressed due to the potential impacts of a specific type of treatments (Lindsay and Loeken, 2017; Barbour et al., 2018). This leads to the critical question on how to weight the gains of in-utero interventions with potential long-term benefits and side-effects.

*We thank Sofía Casabianca, Alejandro Estefan and Carolina Vélez for their valuable research assistance. We also express our gratitude to the attendants of the research seminar at Universidad del Rosario.

The Barker (1990) hypothesis states that the intrauterine environment has long-term effects on humans' metabolic characteristics, and it gave birth to the "fetal origins" literature. Almond and Currie (2011) present a review of the most prominent studies, which usually analyse *good* or *bad* interventions. In some cases, promising interventions result in good impacts during pregnancy or only for mothers. This is the case of Baranov et al. (2020) who found strong long-run impacts on women mental health of a psychotherapy intervention for maternal depression in Pakistan, but no evidence of impacts on children socio-emotional, physical or cognitive development. We contribute to this literature by analysing how treatment for GDM, which might have both positive and negative effects on child development. This is important as Persson et al. (2019) discusses how the rapid transition on nutrition patterns in developing countries exposes children to worse future health outcomes: obesity, diabetes, etc. However, it exposes babies to unknown long-term side effects of treatments aimed towards those conditions. For instance, Akhter et al. (2019) found that while bariatric surgery is a well-known strategy to treat severe obesity, it has a serious impact on the children of mothers who received such treatment.

The cohort study Born in Bradford (BiB), which covers all pregnancies in this English city's main hospital between 2007 and 2011, provides an adequate scenario for studying the condition. The study includes information on socio-economic characteristics, health and education outcomes for the first five years of age. The main advantage of BiB for studying GDM is the universal screening approach of the study. All pregnant women were offered a 75 g OGTT, which includes fasting and 2 hours post-load measures, and the diagnosis was carried on using clearly defined thresholds over them. This is uncommon as usually OGTT is requested based on risk classification, and for diagnosis (and subsequent treatment) until very recently there was no consensus on how to diagnose it. Moreover, the study included surveys for the mother and a wide range of clinical outcomes as ultrasound records, blood tests and anthropometrics for the first five years of age. That information includes both study-specific and administrative records from visits to the local family doctor and school nursery. Additionally, it is complemented with the Early Years Foundation Stage Profile (EYFS) and Key Stage 1 (KS1) school standardised test scores for children (ages 5 and 7). Thus, using a 'doughnut-hole' Regression Discontinuity Design (RDD), we can compare the growth of children born from mothers diagnosed with GDM (and treated for) with those who were not diagnosed, but with a small difference on underlying glucose concentration.

We found that babies born from treated women are born a few days earlier (more likely to be a programmed birth), shorter, and weigh less, as described in the clinical literature. Moreover, cord blood leptin is lower. Anthropometric differences disappear relatively fast with age. Also, there is no evidence of differential diets, energy intake, or weight perceptions by parents. However, there is evidence of worse performance on standardised tests at age 5, especially in mathematics topics, mostly because it became less likely that treated children attain exceeding results.

The rest of this paper is organised as follows. Section two presents a summary of GDM diagnosis, treatment, and clinical evidence. Section three presents the data and statistical methods. Lastly, section four presents the discussion and section five concludes.

2 Context: diagnosis of GDM

Until recently, there was no consensus on which glucose concentration levels should be used for diagnosing GDM. As described by Farrar et al. (2016), diagnosis rules for the International Association of Diabetes and Pregnancy Study Groups (IADPSG), the American College of Obstetricians and Gynecologists (ACOG), and the World Health Organization (WHO), use different OGTT tests 75g plasma glucose, (100g plasma or serum glucose) and consider it at different moments (fasting, 1, 2 or even 4-hour post-load). The authors highlight that this lack of consensus, partly because of the lack of cost-effectiveness analysis of the alternatives, resulted in arbitrary guidelines. Part of this heterogeneity on recommendations results from the lack of evidence of a "threshold effect", that is, damage on fetus grows notoriously at a particular point of glucose concentration (Lindsay et al., 2011). The UK, lead by the National Institute for Health and Clinical Excellence (NICE), adopted 1999 WHO standard in 2008 (75g OGTT with 7.0 mmol/L fasting, 7.8 mmol/L 2h) (Lindsay et al., 2011). NICE recommends screening for women in high risk (BMI above 30 kg/m2, previous macrosomic baby, previous gestational diabetes, first-degree relative with diabetes, minority ethnic family origin with a high prevalence of diabetes). In 2015, NICE adopted the 5.6 fastings, 7.8 2h cutoffs measured around 24 to 28 weeks of gestation (NICE, 2015). This last update resulted from the new 2013 WHO guidelines following the HAPO study and resulting IADPSG recommendations. Currently, public policy debate moves towards the idea of universal screening for GDM as suggested by IADPSG (2010), and whether the cutoffs for diagnosis

should be different for specific ethnic populations such as south Asian women (Farrar et al., 2015).

According to Durnwald and Werner (2018), about 30% of pregnant women diagnosed with GDM need pharmacological treatment and adequate nutritional therapy, exercise and gestational weight control to reduce risks of macrosomia and other conditions for the offspring. There are two pharmacologic options, insulin (and some insulin analogues) and selected oral antihyperglycemic agents such as metformin. Systematic reviews of studies of pregnancy outcome in women with GDM treated with oral antihyperglycemic agents or insulin have generally found that both approaches can be effective.

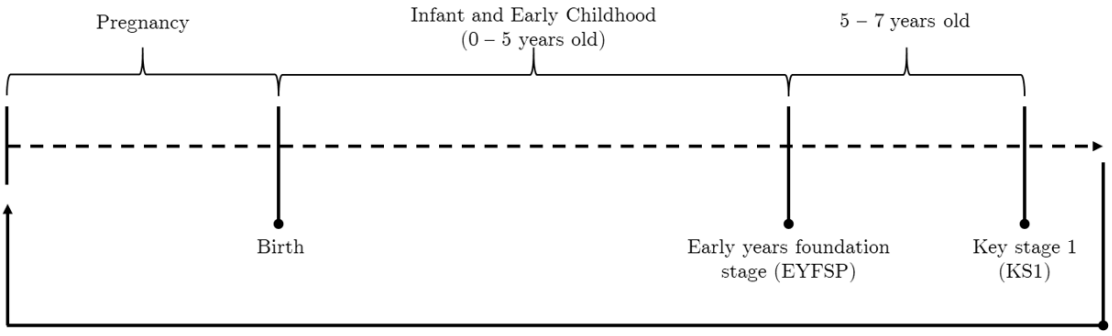
Insulin is a hormone made by the pancreas that allows our bodies to use sugar (glucose) from carbohydrates in the food ingested for energy or to store glucose for future use. The insulin dose varies between individuals due to varied rates of obesity, ethnic characteristics, degree of hyperglycemia, and other demographic criteria. Still, the majority of studies have reported a total insulin dose ranging from 0.7 to 2 units per kg (current pregnant weight) to achieve glucose control. On the other hand, metformin, a relatively cheap treatment and widely used globally, predominantly reduces hepatic glucose output and improves insulin sensitivity. A significant concern with metformin is that it crosses the placenta, and there is the uncertainty of its implications for both mother and child. So far, there is no clear evidence against the usage of metformin but the lack of studies that can assess long-run consequences has driven several countries to not recommend its use during pregnancy (Lindsay and Loeken, 2017). Institutions such as the American Diabetes Association (ADA) have a strong preference for insulin treatment due to metformin's potential secondary effects.

3 Methods

3.1 BiB study and sample selection

We use data from BiB, a birth cohort study based in a hospital in the North of England which serves individuals from a wide variety of ethnic origins (Wright et al., 2012). Participant women were at 26 to 28 weeks of gestation at recruitment between 2007 and 2011. As part of the measurement, a 75 g OGTT (fasting and 2h) was requested. Women with fasting glucose equal or above 6.1 mmol/L or a 2h post-load glucose equal or above 7.8 mmol/L were diagnosed with GDM, which corresponds to a modified WHO 1999 standard (Lawlor et al., 2014; Farrar et al., 2015). Besides, we have information regarding the children whose mothers were part of the BiB survey for the first two tests of the National Curriculum, to quantify the mother's effects for being treated for GDM. Nonetheless, the identification strategy will be explained in further detail below.

Figure 1: Data sample timeline

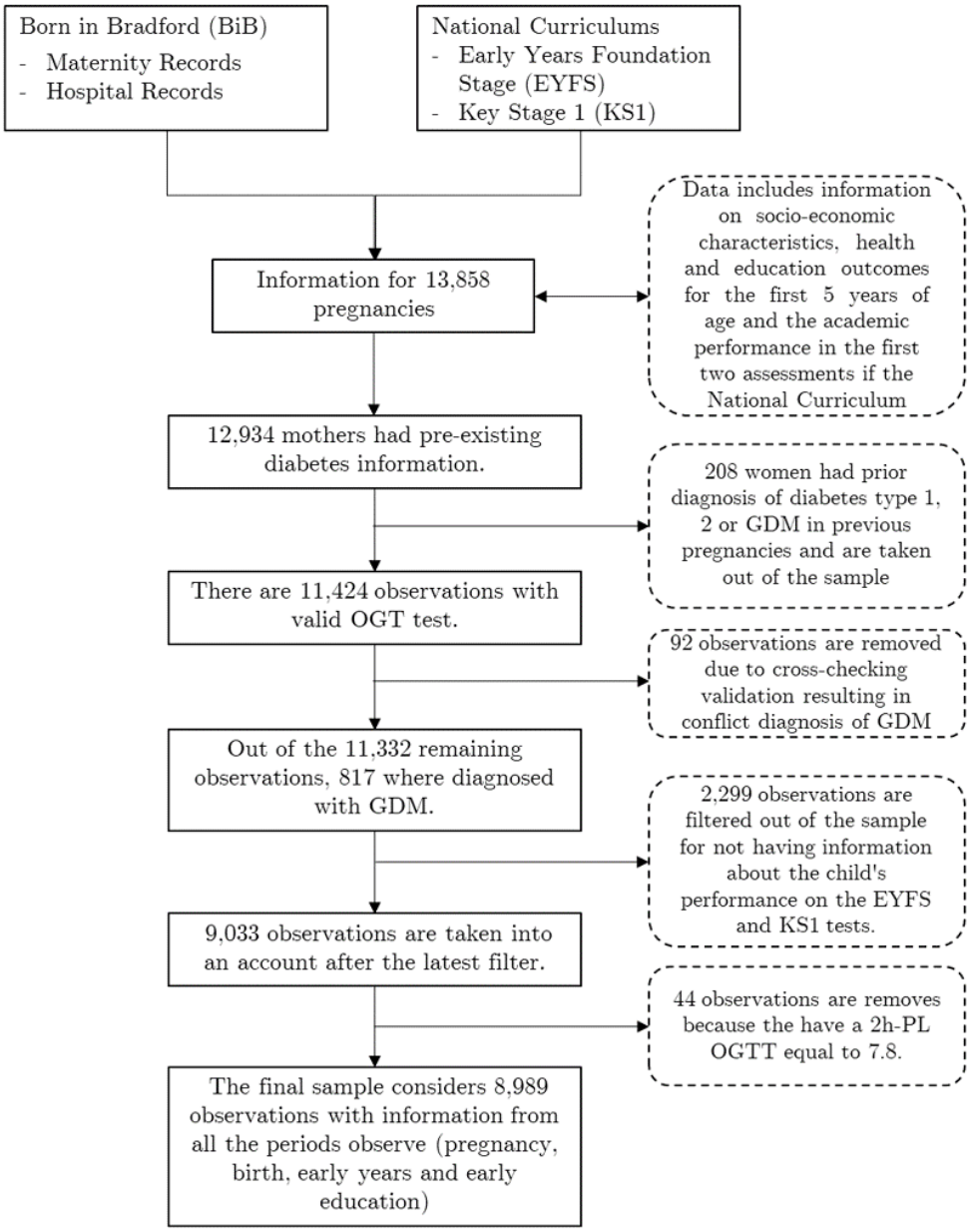


Map back children whose mothers were surveyed by the Born in Bradford survey during their pregnancy and that, thought a oral glucose tolerance test (OGT test), were diagnosed with Gestational Diabetes Mellitus

As is showed in figure 1 we have four primary sources of information. Information regarding the mother, pregnancy and early years is taken from the maternity and hospital administrative records, and the BiB study surveys and specific measurements. Third, the information form the Early Years Foundation Stage of the National Curriculum, which assesses children from 3 to 5 years old in the basic knowledge of literacy and mathematics. Lastly, the Key Stage 1 assessment for children between 5 and 7 years old. Figure 2 explains the processing of the data from the different information sources and the construction of the final

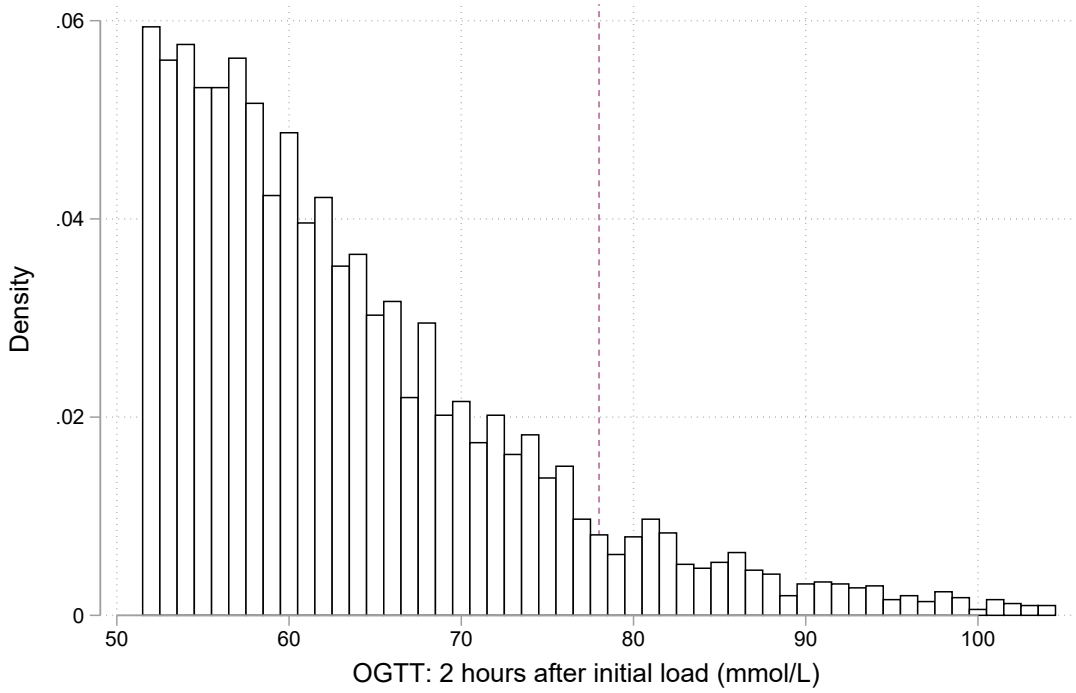
sample regarding the mothers whose children present the two tests proposed for this evaluation, along with other socioeconomic and demographic characteristics. In the sample, only 208 out of 12,934 pregnancies (with diabetes history information) were from mothers diagnosed either with diabetes type 1 or 2 before pregnancy, or with GDM in prior pregnancies. From the remaining, 11,424 had valid OGTT tests. There are 92 conflicting diagnosis reports according to the source (maternity records and hospital records) which were removed. From this group, 2999 observations are not considered because there is no available information about their education outcomes. The last filter, which corresponds to removing mothers who had a record of 2h PL OGTT strictly at 7.8 mmol/L, will be explained once the RDD method is presented.

Figure 2: Data processing and filters



Out of the 11,332 for whom there are OGTT records, 817 were diagnosed with GDM (7.4%). From these, 776 have a 2h PT glucose concentration above 7.8 mmol/L, and 41 only had fasting glucose above 6.1 mmol/L but a 2h post-load below 7.8 mmol/L. Because of this, we will centre our analysis on the cutoff of the 2h load

Figure 3: Distribution of PL 2h OGTT results



measure, which distribution is presented in figure 3.[1] Finally, 776 are under treatment (88%) which could be a mix between diet, lifestyle intervention, medication (metformin), or insulin.

3.2 Health records

BiB has rich information on biomarkers taken both by the study but also as part of regular health controls. There are regular ultrasound registers and anthropometric measurements on top of the information at birth and at booking (enrolment in the study). Panel A of figure 4 presents the distribution of observations in the ultrasounds records according to the mother's gestational week, and panel B shows the distribution on anthropometric information according to the age of children.

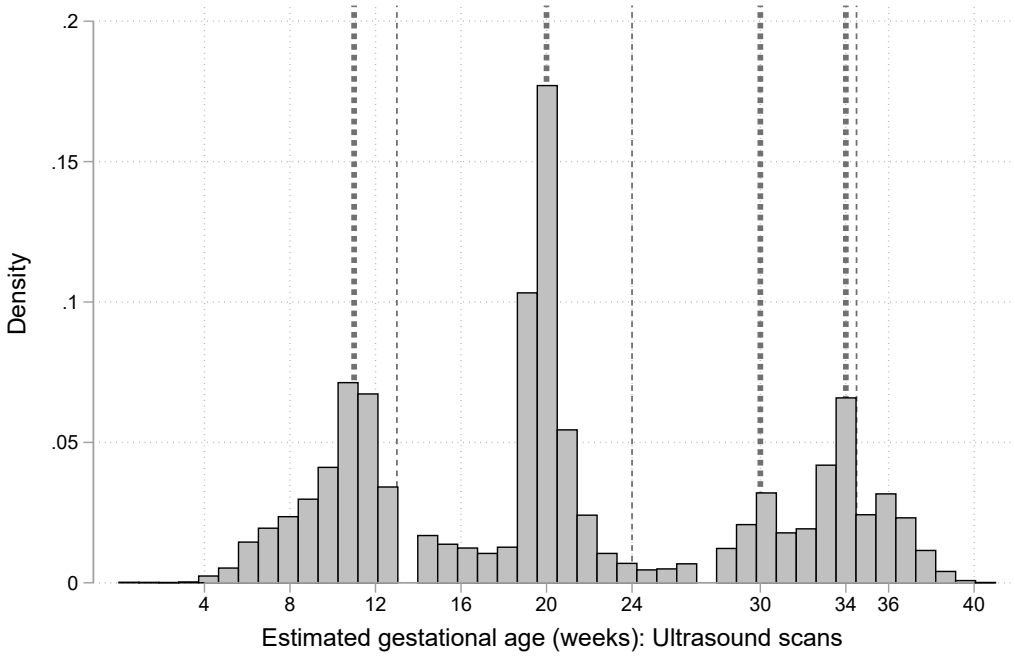
In ultrasounds data, fetuses are usually observed between 2 and 4 times, but it could be more if the attending physician required it in some cases. Measurements include femur length, abdominal circumference and head circumference, which we have standardised using the Intergrowth 21st tables (Papageorghiou et al., 2014). There are also twelve flags of physicians' problems based on the records, or abnormal indicators. We put them all together into a single dichotomic variable. We consider the abdominal circumference to head circumference ratio (AC/HC) and the femur length to head circumference ratio (FL/HC). These two indicators are known predictors of problems at birth for macrosomic babies. Duryea et al. (2017) presents cutoffs for the FL/AC. Hehir et al. (2019) suggest using the 90th percentile for the AC/HC. For our analysis, we will take only two measurements for each fetus, the last one before the OGTT, and the first one seven weeks after the OGTT. We selected this timing to allow for the treatment to be implemented but with enough observations before birth.

In the growth data, half of the children are observed three to five times in data, but it is not unusual to have up to ten measurements. This is because the sources included BiB specific measurements and health records and school nursery data. We consider only periods where there is a larger mass of respondents: 1-3 months, 4-13 months, 22-27 months, 54-60 months. If there is more than one record per children in each period, we take the average of all current records for the age-standardised measurements.

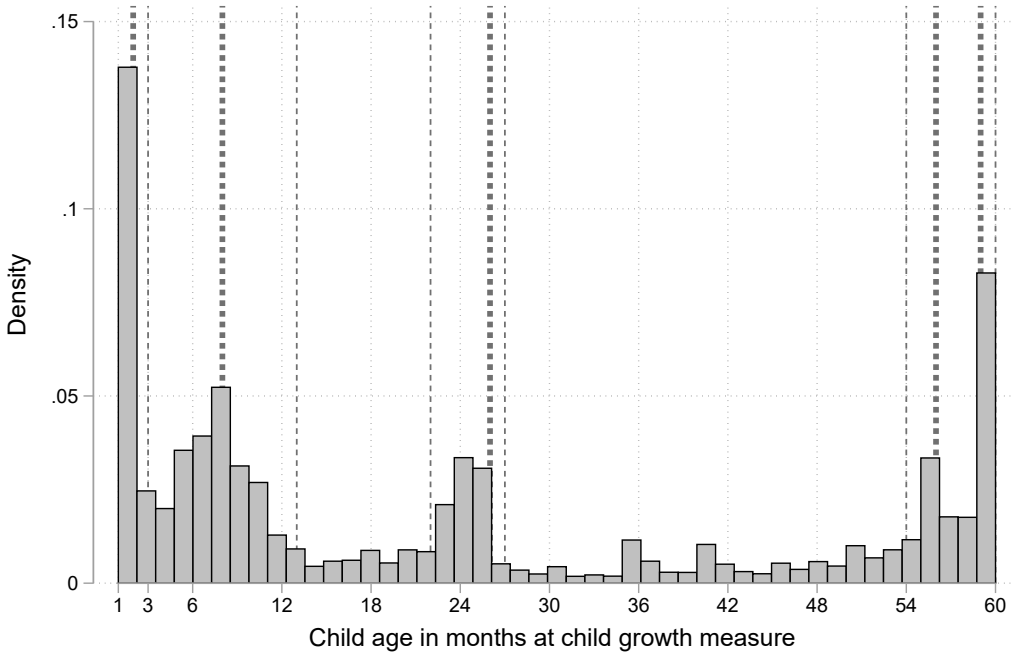
[1] Sample size above the 6.1 mmol/L fasting cutoff is too small for our type of analysis, as observed in Figure 10. Notice that this cutoff is well above current NICE guideline which uses 5.6 mmol/L instead.

Figure 4: Density of available health records

Panel A: Ultrasounds registers according to gestational week



Panel B: Child growth measurements according to age



Notes: distribution of individuals with anthropometric measures according to gestational week and age in months.

3.3 Education records

Public primary and secondary schools' curriculum in the United Kingdom includes the national curriculum in their primary curriculum. This curriculum is a set of subjects and standards used, so children learn the same topics (GOV.UK, 2014a). It covers subjects that are to be taught in each stage of education, and the standard knowledge children should reach in each of those subjects.

The national curriculum is divided into blocks of years called "key stages" (KS). This paper uses information from the first two assessments for children between 2 and 7 years. The first assessment is called the early years' foundation stage (EYFS), which sets the standards for the learning, development and care of children from birth to 5 years of age (GOV.UK, 2015). This assessment tests children in communication and language, literacy and mathematics, among other subjects. The second assessment taken into account for this investigation is the Key Stage 1 (KS1), which targets primary school children between 5 and 7 years and checks, as does EYFS, for proficiency in math, literacy and science (GOV.UK, 2014b).

EYFSP 2013 version includes 17 areas, and for each of them, the possible outcomes are (1) Emerging, (2) Expected, (3) Exceeding. The test is taken in year R, corresponding to years old children. For KS1, taken in year Y2 (around age 7), there are two versions, one prior o 2016 and other post-2016. We recoded the variables into the classification. (1) Emerging: Levels 1 and 2c in the pre-2016 version, and the below standard and pre-key stage foundations (BLW, PKF) in the post-2016 version. (2) Expected: levels 2B and 2A, and working towards and at the expected standard (WTS, EXS). (3) Exceeding: level 3, and working at greater depth (GDS).

Following the methodology for summary index used in Anderson (2008), we pooled multiple outcomes from the EYFS and the KS1 into a single measure for each assessment. A summary index or Anderson Index is a weighted mean of several standardised outcomes. The weights are calculated to maximise the amount of information captured in the index. This method is robust to over-testing because each index represents a single test; therefore, the probability of false rejection does not increase as additional outcomes are added to the index (Anderson, 2008). Another advantage of the summary index includes a statistical test for a program's "general effect" on a set of outcomes. Additionally, it is potentially more potent than individual tests at attaining statistical significance due to the multiple outcomes that approach marginal significance as they are being aggregated into a single index.

For the EYFSP assessment, the subjects included in the index where only mathematics and literacy, as for the KS1, science was also considered. On the other hand, the KS1 test literacy subject is divided into two sections, writing and reading.

3.4 Health records after birth

The BiB cohort includes an administrative link with general practitioners (GP) registers of both the children and mothers. We requested the Bradford team to assess if there were registers of appointments, procedures or test related to a list of ICD-10 codes linked in the literature with GDM, for both mothers and children.[2].

3.5 BiB1000 sample

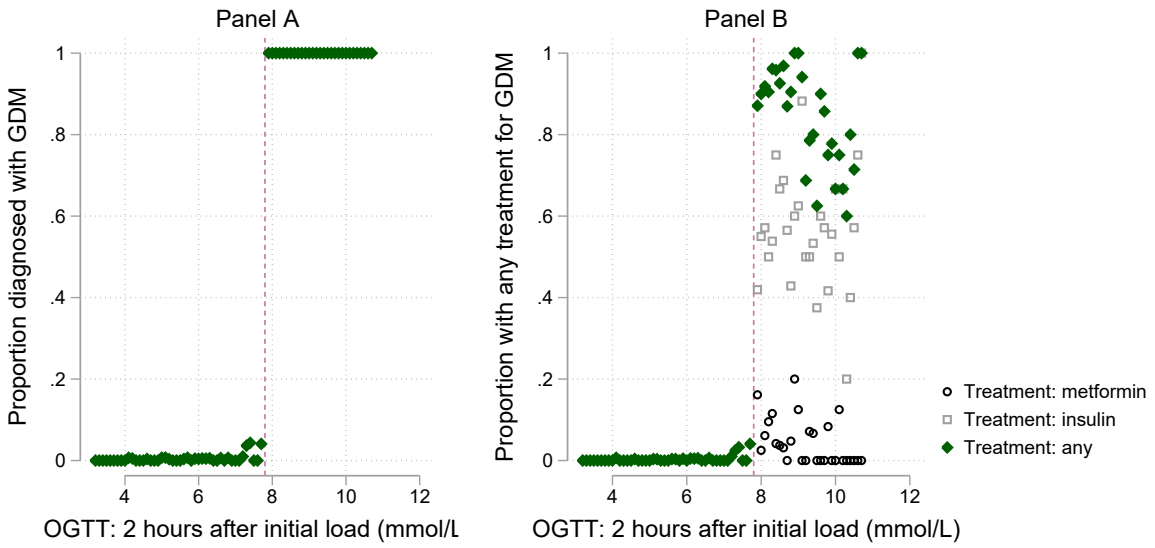
The BiB study includes a household survey conducted with a sample of 1000 mothers after they were discharged from the hospital (BiB1000). Surveys were done 6, 12, 18 and 36 months after birth. The instrument includes general health, socio-emotional skills, physical activity, parental practices, and nutrition. The study's goal was to collect data for designing preventive information to alter modifiable risk factors of healthy eating behaviour (Wright et al., 2016).

The study includes the Strengths and Difficulties Questionnaire (SDQ), an emotional and behavioural screening questionnaire for children aged 2 to 4 (Muris et al., 2003). Their survey includes five scales of 5 items: emotional symptoms, conduct problems, hyperactivity and inattention, peer relationships problems, and prosocial behaviour. With the first four scales, we derived a total difficulties index.

In addition to the SDQ, the survey includes a report on parenting measures for the BiB1000 cohort. Evidence to date from Longitudinal Study of Australian Children (LSAC) and Zubrick et al. (2014) confirm the important role of parenting in shaping children's behavioural and emotional adjustment, early literacy, and lifestyle-related health conditions such as obesity. Through Structural Equation Modelling (SEM), the authors examined the mother- and father's properties - reported parenting measures and derived recommended weighted composites.

[2]Bradford GPs use the CTV3 classification

Figure 5: Proportion of individuals diagnosed and treated for GDM, according to their 2h OGTT test



We group food intake into food groups, following Mahoney et al. (2018) and Wright et al. (2016). Respondents are asked about the frequency of consumption of several products classified into food groups. If the frequency was above the median, the group indicator is set as 1.

3.6 Regression Discontinuity Design

We implement a Sharp Regression Discontinuity Design (RDD). Figure 5 presents the central idea behind our identification strategy: we compare children whose mothers had an OGTT test at booking just above 7.8 mmol/L against those slightly below such record. This is because everyone above the threshold was diagnosed with GDM, and almost everyone below it was not diagnosed with the condition (see panel A). As a result, Panel B shows that there is also a jump on the probability to receive treatment for the condition. Panel B also presents two types of treatment; insulin, the most common, and metformin, the less common.

Following Imbens and Lemieux (2008), the impact of the diagnosis of GDM $W = 1(OGTT \geq 7.8)$ on outcome Y_t measured at period t (gestational weeks or age in semesters) was identified by the exploiting the discontinuity in its conditional expectation at the diagnosis cutoff $OGTT = 7.8$:

$$\begin{aligned} \delta_0^{Y,t} &= E[Y_{i,t}(W=1) - Y_{i,t}(W=0)|OGTT_i = 7.8] \\ &= \lim_{OGTT \downarrow 7.8} E[Y_{i,t}|OGTT_i = 7.8] - \lim_{OGTT \uparrow 7.8} E[Y_{i,t}|OGTT_i = 7.8] \end{aligned} \tag{1}$$

, where i indicates an individual child. Figure 5 presents the idea of the estimate: the OGTT value is presented in the x-axis, so each of the dots in the figure presents the proportion of individuals how are reported as diagnosed with GDM (Panel A) and given any specific treatment (Panel B). Given that the jump on the proportion is practically from 0 to 1, parameter $\delta_0^{Y,t}$ presented in equation 1 can be obtained by estimating parameter $\delta_{Y,t}$ from equation 2:

$$Y_{i,t} = \delta_{Y,t} W_i + \alpha_0 + f_l(\alpha_l, OGTT_i|W_i = 0) + f_r(\alpha_r, OGTT_i|W_i = 1) \quad , OGTT_i \in [-h_l, h_r] \tag{2}$$

where functions $f(\cdot)$ correspond to the function form chosen for the link between OGTT test and the probability of diagnosis. Our main results correspond to local linear regressions with triangular weights. The other main decision is the window size for the inclusion of individuals above and below the cutoff, which is

represented by the bandwidths h_l and h_r. Such values are usually obtained from minimising the mean square error of the non-linear approximation, which depends not only on the running variable distribution but also on the specific outcome one (Calonico et al., 2014). To maintain a consistent sample over the estimates, we compute optimal bandwidths for the education variables ($h_l = 1.8$, $h_r = 1.4$) and use them for all outcomes. In other words, estimates computed within the range (6 - 7.8 mmol/L) and (7.8 - 9.2 mmol/L). In the appendix, we provide estimates using alternative bandwidths, including the optimal one for each outcome. Finally, confidence intervals are robust to the bandwidth selection following Calonico et al. (2014) algorithm.[3] A fuzzy RDD is not considered for treatment as, conditional on diagnosis, this channel of action might be correlated with the measured outcomes.

The crucial identification assumption of this method is that individuals cannot decide whether they are on the left or the right of the cutoff. This is established with a continuity of the density function at the cutoff, shown in appendix C, and by showing no discontinuity on pre-intervention characteristics. Before exploring this assumption in detail, it is essential to consider a particularity at OGGT 2h=7.8 mmol/L. Individuals with precisely this record have notoriously different characteristics such as a notoriously low booking weight. As this is exactly at the cutoff, discontinuity analyses are sensitive to such particularities, as is discussed in depth in appendix B (Barreca et al., 2011). Hence, we removed those observations from the rest of the study, as presented in the diagram of figure 2. Thus, the technique is a 'doughnut-hole' RDD.

3.7 Pre-intervention characteristics

Table 1 presents the mothers' characteristics before and at booking. Column 2 presents the average of the rows variables for all mothers who meet the filters described above. Columns 3 to 5 do the same but according to the OGTT 2h post-load test. OGTT categories are based on the optimal bandwidth calculations described in the empirical section below. Effectively, our main estimates compare children born from mothers in columns 3 and 4. The mothers in these groups are slightly older (approx. 29 years old) than the characteristic mother of the study (approx. 27 years old), with a higher BMI (27 vs 26 of the general mother). This result is expected as it is closer to the risk profile for GDM.

Columns 6 and 7 present the balance test between groups using the same RDD strategy used for the primary outcomes. In almost all covariates, there are no statistically significant differences. The only case where this happens is on the proportion of respondents who have ever smoked cigarettes, where the null can be rejected at a 5% significance level. This assures us that the main assumption of our identification strategy, that near the 7.8 mmol/L threshold the GDM diagnosis could be considered as if random.

[3]This procedure was implemented using the `rdrobust` routine (Calonico et al., 2014).

Table 1: Mothers' characteristics at enrolment

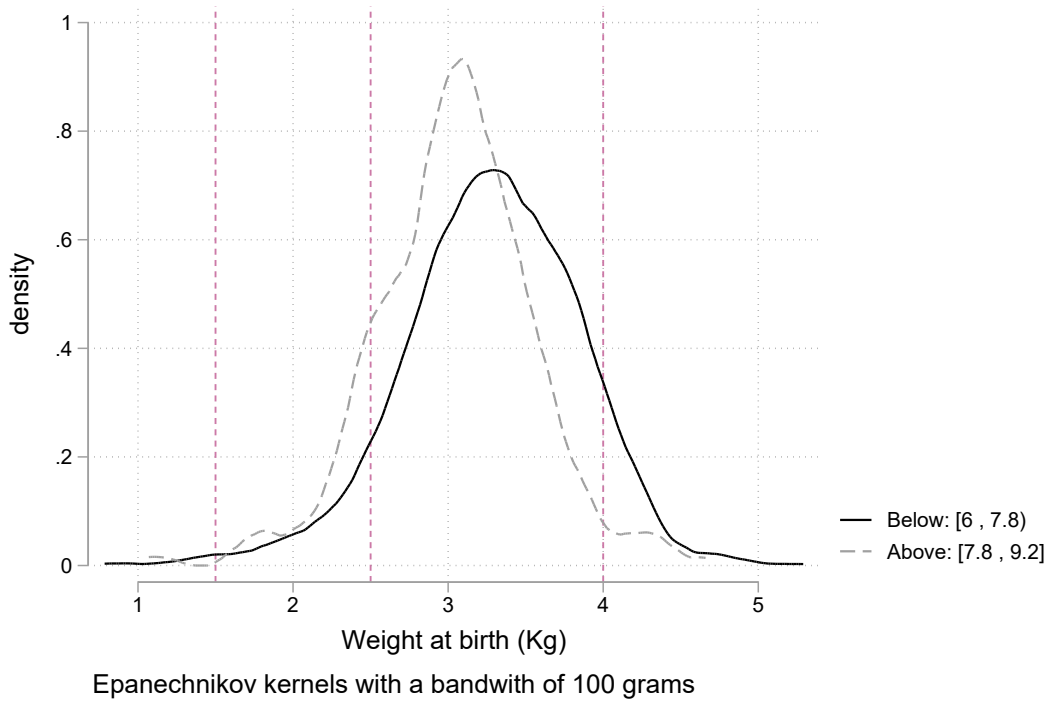
Variables	(1) **ALL**	(2) Below cutoff **[1.6,5.8)**	(3) **[5.8,7.8)**	(4) Above cutoff **[7.8,9.8)**	(5) **[9.8,27.7]**	(6) Balance test **Difference**	(7) **[p-val]**
Panel A. Socio-demographic							
Mother age at booking	27.28	26.41	28.16	29.80	32.83	0.159	[0.571]
Total number of members in household	4.20	4.14	4.25	4.25	4.95	-0.359	[0.684]
Father age (years): Father baseline questionnaire	29.95	29.13	30.70	32.83	35.28	-2.491	[0.125]
White British Mother	0.35	0.38	0.33	0.25	0.08	0.044	[0.580]
Mother's qualification below A-level	0.49	0.49	0.49	0.47	0.50	0.056	[0.220]
Panel B. Economic							
IMD Deciles (WITHIN Bradford)	3.84	3.81	3.88	3.91	3.56	0.329	[0.847]
Low decile within Bradford	0.45	0.45	0.45	0.45	0.43	-0.041	[0.838]
Mother currently employed	1.57	1.55	1.57	1.61	1.70	-0.028	[0.501]
GHQ factor score (derived)	0.07	0.06	0.09	0.08	0.10	0.020	[0.894]
Panel C. Health							
OGTT: Fasting glucose	4.52	4.39	4.60	4.94	6.16	-0.124	[0.919]
OGTT: 2-hour post load glucose	5.66	4.76	6.51	8.55	11.56		
BMI at booking	26.01	25.28	26.83	27.90	30.35	-0.551	[0.496]
Systolic blood pressure at booking	108.31	107.62	109.16	109.53	112.25	-0.456	[0.332]
Diastolic blood pressure at booking	64.20	63.60	64.96	64.96	67.56	-1.199	[0.852]
Mother has ever smoker	0.30	0.34	0.26	0.18	0.17	-0.042	[0.048]
Any pre-existing condition	0.34	0.34	0.33	0.36	0.36	0.013	[0.687]
Existing HT	0.01	0.01	0.01	0.01	0.04	-0.001	[0.738]
Pre-eclampsia	0.03	0.02	0.03	0.03	0.05	0.026	[0.328]
Family History of diabetes	0.30	0.27	0.32	0.39	0.57	-0.068	[0.672]
Family History of high blood pressure	0.31	0.30	0.32	0.36	0.32	-0.007	[0.942]
Gestation (wks) at booking appointment	12.43	12.48	12.34	12.26	12.95	0.156	[0.727]
Gestational age at OGTT (completed weeks)	26.29	26.27	26.29	26.28	27.06	0.054	[0.488]
Number of observations	8307	4913	2839	428	127		

Notes: Simple averages using BiB data. It includes only those pregnancies for women without pre-existing diabates and for which there is an OGTT test. For columns 2 to 5, the sample was restricted according to the average OGTT 2h load measure in mmol/L. Column 6 estimates the jump in the running variable at $OGTT2h = 7.8$ using a sharp RDD, and column 7 presents the robust p-value for the null of the jump being zero.

4 Results

4.1 Children health outcomes

Figure 6: Weight-at-birth below and above 2h PT OGTT = 7.8 mmol/L



Notes: vertical lines present the standard cutoffs for classifying newborns according to their weight at birth. Very low birth weight (VLBW): Below 1.5 Kg; Low birth weight (LBW): below 2.5 Kg; and macrosomic: above 4 Kg.

Figure 6, which compare weight distributions below and above the 7.8 cutoff, summarises the main result at birth: treated babies are less likely to be macrosomic. Panel A of table 2 presents the main differences on outcomes at birth between children whose mother had a 2h PL OGTT slightly above and below the 7.8 mmol/L threshold. Columns 2 and 3 present the effective number of observations included to the left and the right given the chose bandwidths. For some variables, there is no information for all members of the study, and for this reason, it differs from outcome to outcome. Column 1 presents the mean of the variable below the threshold, and column 4 the estimated size of the jump for children who pass that threshold, which we interpret as the causal effect of GDM diagnosis (and treatment). Column 6 presents the associated standard error for the jump of column 5, and column 7 the corresponding p-value for the null test for the estimate. The first row shows that newborns from mothers who receive diagnosis and treatment, just because their 2h PT OGTT was slightly above the cutoff, weigh 242 grams less than those who did not receive the same treatment but had a slightly lower OGTT result (their weight was on average of 3309 grams). An important question is how it impacts weight distribution and other well-known indicators of child growth. The following rows show the proportion of children classified as very low birth weight (VLBW), low birth weight (LBW), or macrosomic. The result is that the only significant impact is on the macrosomic category, as shown in the literature. Apart from this, all other results are similar irrespective of the bandwidth.[4]

Apart from weight and size at birth, a second result is that gestational time is shorter for those diagnosed with GDM in around 6 days (0.53 standard deviations). This is because it is more than twice as common that the onset of labour was not spontaneous (from 30% below to 70% above). It does not seem to be related to

[4]The difference in size at birth is found in other child-growth indicators such as abdominal circumference, mid-arm circumference and sub-scapular skin-fold thickness.

particular issues at the end of pregnancy, and it does not imply more pre-term births. This can be observed because c-sections rates (elective or not) and APGAR tests are not different.

The last result comes from the analysis of blood samples from the cord. There is no difference for insulin, but there is an important one for the concentration of leptin. This, the 'satiety hormone', is responsible for telling the body that no more food is required and related to cognition and other functions (Dalamaga et al., 2013). The more body fat, the more leptin is expected. It was 13.8 ng/ml for mothers undiagnosed with GDM and dropped to 8.3 ng/ml if GDM was diagnosed.

Panel B of table 2 presents two of the main outcomes but under the BiB1000 sample. This involves a drastic reduction on the sample size. The jump estimate for weight at birth is almost the same, but its standard error is more than twice the one of the main sample. The same happens for gestational days. As we will present several outcomes in this sample, we should bear in mind the reduction in detection power.

Fetal ultrasound data are currently being analysed, and the results will be added in the next version of the paper.

Table 2: Impact on children and mother characteristics at birth

Variable	(1) Mean	(2) N left	(3) N right	(4) Jump	(5) SE	(6) p-val
Panel A: Health records at birth						
Weight at birth (Kg)	3.309	2157	365	-0.242	0.064	0.0002
Weight at birth $<$ 2.5 Kg (LBW)	0.069	2157	365	0.058	0.038	0.1257
Weight at birth $\geq$ 4 Kg (Macrosomic)	0.102	2157	365	-0.055	0.028	0.0529
Gestation in days	276.353	2158	365	-6.230	1.390	0.0000
Gestational weeks $<$ 37 (pre-term)	0.071	2158	365	0.010	0.032	0.7570
Gestational weeks $\geq$ 42 (post-term)	0.013	2158	365	-0.007	0.006	0.2093
Non spontaneous onset of labour	0.302	2157	365	0.411	0.054	0.0000
C-section	0.259	2158	365	0.029	0.053	0.5917
APGAR at 1 minute	8.387	2139	364	0.204	0.153	0.1836
APGAR at 5 minutes	9.079	2138	364	0.026	0.058	0.6536
Insulin	6.155	1291	225	-1.310	0.815	0.1082
Cord blood leptin	13.820	1306	228	-5.074	1.409	0.0003
Panel B: BiB1000 subsample						
Weight at birth (Kg)	3.283	320	57	-0.232	0.149	0.1204
Gestation in days	277.119	320	57	-3.848	2.170	0.0761

Notes: Lower BW= 1.8 mmol/L, upper BW = 1.4 mmol/L. Variable names in bold correspond to cases where the estimated impact coefficient (jump) is significant at 10% level.

4.2 Mother health outcomes

As for the mother, table 3 presents outcomes related to the mother during and after pregnancy. There is no evidence of differences in BMI or systolic blood pressure. However, diastolic BP has an interesting pattern: due to the diagnosis and subsequent treatment, it was larger at 28 weeks, but then it became lower at birth. Nevertheless, there is no evidence of differences of high blood pressure rates (systolic BP $\geq$ 140 or diastolic BP $\geq$90). For lipid results, there are differences. Triglyceride and insulin levels of the mother are lower and depending on the bandwidth, the Cholesterol HDL ratio as well. Panel B shows whether a health record at the GP or a hospital, in the five years after birth, is related to some of the most common health conditions associated with GDM. There is no evidence of differences with respect to diabetes, angina or heart disease. Finally, panel C presents BMI and subjective health results for those women in the BiB100 subsample. The BMI for treated mothers appears to be smaller, especially 6 months after pregnancy. However, the difference is not significant. Additionally, six months after birth, mothers general health appear to be better according to the general health questionnaire.

Table 3: Impact on mother characteristics

Variable	(1) Mean	(2) N left	(3) N right	(4) Jump	(5) SE	(6) p-val
Panel A: Health records at birth						
Any problem at birth	0.243	2142	363	-0.062	0.050	0.2146
Systolic blood pressure at birth	129.395	1686	255	-0.732	2.239	0.7438
Diastolic blood pressure at birth	76.689	1686	255	-2.437	1.276	0.0561
High blood pressure at birth	0.284	1686	255	-0.087	0.059	0.1428
Cholesterol HDL Ratio	3.371	1988	334	-0.215	0.109	0.0472
Cholesterol	6.264	1988	334	-0.134	0.151	0.3752
HDL	1.930	1988	334	0.090	0.054	0.0939
LDL	3.371	1964	330	-0.186	0.138	0.1767
Triglyceride	2.255	1988	334	-0.185	0.102	0.0684
Panel B: Health records after birth related to						
Diabetes mellitus	0.118	2119	354	-0.093	0.110	0.3999
Hypertensive heart disease	0.044	2119	354	0.035	0.065	0.5874
Angina pectoris	0.116	2119	354	-0.026	0.045	0.5648
Panel C: BiB1000 subsample						
BMI at 6 months after birth	26.887	260	49	-2.261	1.740	0.1937
BMI at 12 months after birth	26.446	253	51	-0.584	1.705	0.7319
BMI at 24 months after birth	26.206	239	48	-0.704	1.930	0.7153
BMI at 36 months after birth	26.503	239	42	0.256	1.813	0.8878
Mother general (bad) health: 6 m.	1.679	244	47	-0.215	0.086	0.0122
Mother general (bad) health: 18 m.	1.615	263	50	-0.054	0.129	0.6769

Notes: Lower BW= 1.8 mmol/L, upper BW = 1.4 mmol/L. Aby problem is a dummy that takes the value of one if there is a record for any of the following complications: vaginal bleeding, vaginal discharge/infection, urinary tract infection, threatened to be preterm birth, hospital admissions postpartum, suspected fetal growth retardation, oligohydramnios, gestational hypertension or preeclampsia. High blood pressure: systolic BP $\geq$ 140 or diastolic BP $\geq$90. Variable names in bold correspond to cases where the estimated impact coefficient (jump) is significant at 10% level.

Table 4 presents results on the last fetal measurement before the OGTT, and the first after the OGTT. In general, OGTT treatment was expected to be carried on at week 24, but there is some variation. We do not find differences before the OGTT on any of the measures, before or after the intervention. This indicates that differences are likely to appear only in the last gestational weeks, for which fewer observations are available.

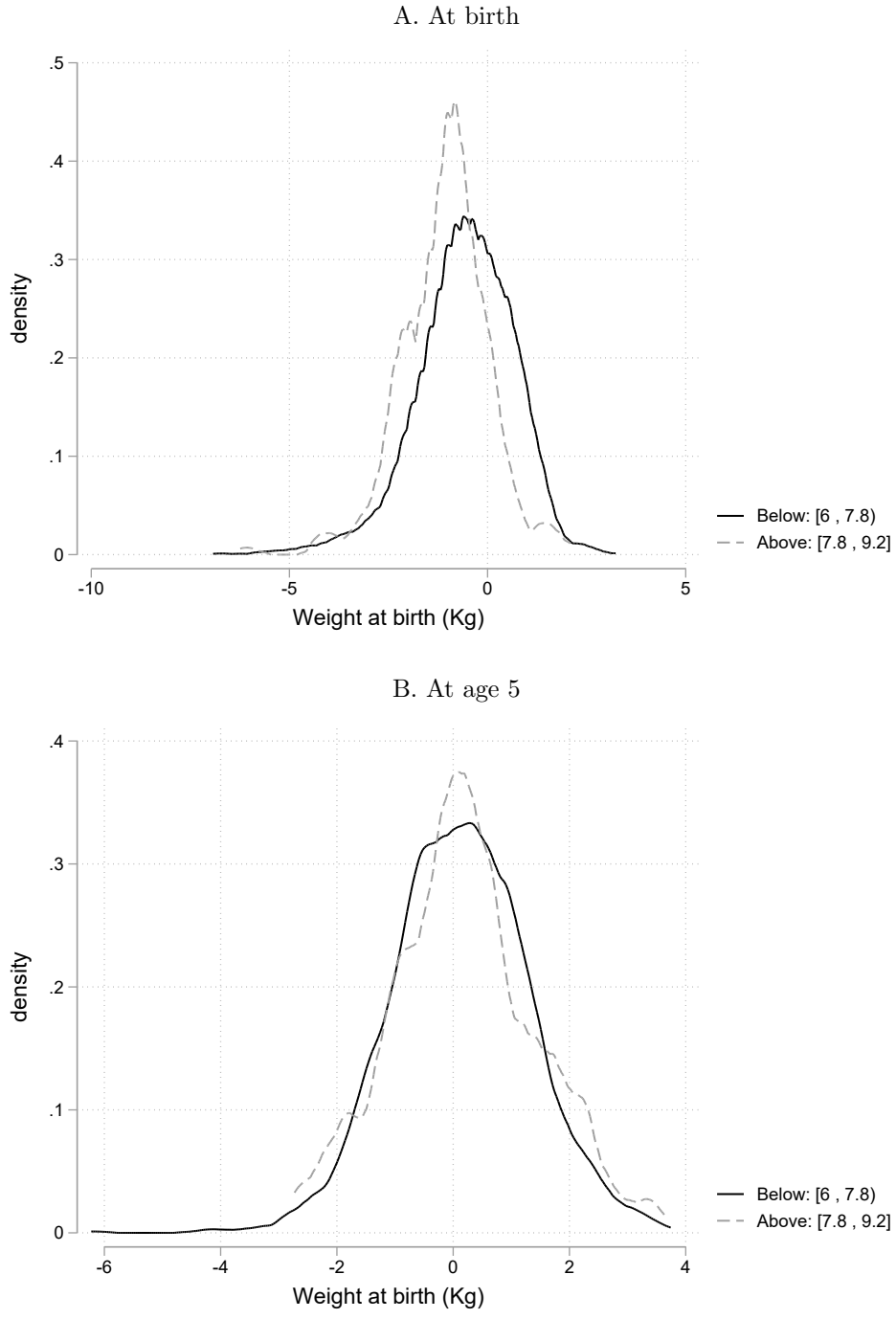
Table 4: Impact on fetal growth

Variable	(1) Age	(2) Mean	(3) N left	(4) N right	(5) Jump	(6) SE	(7) p-val
Femur length Z-score	Before OGTT	1.028	1885	312	-0.071	0.142	0.6181
Femur length Z-score	7 wks after OGTT	0.921	788	319	-0.153	0.148	0.3015
Abdominal cirfunference Z-score	Before OGTT	0.831	1451	244	-0.006	0.164	0.9716
Abdominal cirfunference Z-score	7 wks after OGTT	0.264	789	319	-0.140	0.165	0.3941
Head cirfunference Z-score	Before OGTT	0.896	1893	316	-0.074	0.143	0.6062
Head cirfunference Z-score	7 wks after OGTT	0.734	747	289	-0.014	0.154	0.9264
FL/AC ratio	Before OGTT	0.216	1445	242	-0.001	0.002	0.4992
FL/AC ratio	7 wks after OGTT	0.219	787	318	-0.001	0.002	0.4012
AC/HC ratio	Before OGTT	0.861	1435	244	0.006	0.006	0.2940
AC/HC ratio	7 wks after OGTT	0.957	747	288	0.008	0.007	0.2419
Adnormal Indicator (12 variables)	Before OGTT	0.001	2074	346	0.000	0.001	0.7744
Adnormal Indicator (12 variables)	7 wks after OGTT	0.009	806	325	0.002	0.003	0.6064
FL/AC $\leq$ 0.19	Before OGTT	0.011	1445	242	-0.008	0.007	0.2832
FL/AC $\leq$ 0.19	7 wks after OGTT	0.010	787	318	0.013	0.011	0.2261
FL/AC $\leq$ 0.20	Before OGTT	0.060	1445	242	0.048	0.041	0.2446
FL/AC $\leq$ 0.20	7 wks after OGTT	0.057	787	318	-0.000	0.035	0.9984
AC/HC ratio > p90	Before OGTT	0.111	1435	244	0.020	0.045	0.6507
AC/HC ratio > p90	7 wks after OGTT	0.139	747	288	0.040	0.061	0.5105

Notes: Lower BW= 1.8 mmol/L, upper BW = 1.4 mmol/L. Ultrasound standards per gestational week were taken from the Intergrowth 21st project. Estimates use the last available measurement before the OGTT test, and the last before pregnancy which took place at least 7 weeks after the test. Multiple measurements for the same gestational week per children were averaged, but the flags for abnormal indicators are for any measurement in the given week. Abdominal circumference to head circumference ratio (AC/HC) reference is the 90th percentile, following Hehir et al. (2019). Femur length to head circumference ratio (FL/HC) cutoffs from Duryea et al. (2017). Variable names in bold correspond to cases where the estimated impact coefficient (jump) is significant at 10% level.

4.3 Growth during early childhood

Figure 7: Weight-for-age z-score (UK1990) below and above 2h PT OGTT = 7.8 mmol/L



Notes: Epanechnikov kernels with a bandwidth of 0.2 standard deviations

Figure 7 summarises the main result of this section: differences in child growth at birth (Panel A) are no longer present at age 5 (Panel B). Those differences disappear by three months of age. Panel A of table 5 presents anthropometrics' progression for children since born until they are five years old. Following the previous section results, two of the three basic measures (BMI-for-age, height-for-age, and weight-for-age)

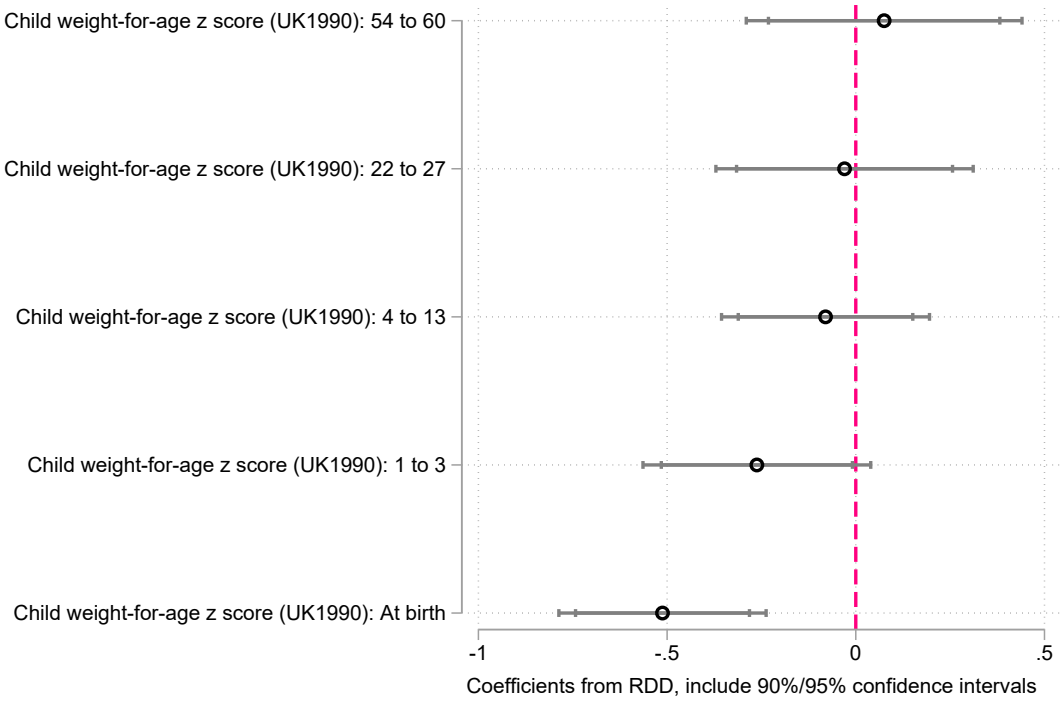
are significantly smaller for the GDM group. However, these differences disappear after the first year of age. Figure 8 presents the declining pattern graphically on mean weight-for-age difference. A second important observation is that these differences are not reflected in obesity indicators (defined over BMI-for-age). A third observation is related to blood pressure levels at age 4 and 5, measured by school nurses. There is evidence of lower systolic blood pressure (3.38 mmHg). These child growth differences might mask long-term impacts on health. Panel B shows no difference in health records related to GDM in the literature, apart from overweight and obesity.

Table 5: Impact on children growth

Variable	(1) Age in months	(2) Mean	(3) N left	(4) N right	(5) Jump	(6) SE	(7) p-val
Panel A: Anthropometrics							
Child BMI-for-age z score (UK1990)	1 to 3	-0.617	1661	266	-0.327	0.167	0.0500
Child BMI-for-age z score (UK1990)	4 to 13	-0.653	1951	338	-0.006	0.149	0.9670
Child BMI-for-age z score (UK1990)	22 to 27	0.058	1000	180	-0.243	0.196	0.2164
Child BMI-for-age z score (UK1990)	54 to 60	0.203	1441	231	0.078	0.177	0.6604
Obesity Indicator	1 to 3	0.014	1807	289	0.009	0.017	0.6192
Obesity Indicator	4 to 13	0.023	2014	351	0.016	0.018	0.3703
Obesity Indicator	22 to 27	0.081	1345	228	0.016	0.042	0.7002
Obesity Indicator	54 to 60	0.052	1738	276	0.013	0.025	0.5960
Diastolic BP (by school nurse)	54 to 60	63.093	1358	215	-1.806	1.624	0.2661
Systolic BP (by school nurse)	54 to 60	99.508	1358	215	-3.386	1.665	0.0419
Panel B: Health records after birth related to							
Obesity	0 to 60	0.024	2119	354	-0.030	0.016	0.0518
Lactose intolerance	0 to 60	0.024	2119	354	0.093	0.090	0.2970
Disorders of galactose metabolisme	0 to 60	0.011	2119	354	0.020	0.018	0.2681
Gilbert and Crigler-Najjar syndrome	0 to 60	0.026	2119	354	0.025	0.034	0.4711
Cystic fibrosis	0 to 60	0.013	2119	354	-0.000	0.015	0.9983
Developmental disorder of scholastic	0 to 60	0.015	2119	354	0.006	0.024	0.7943

Notes: Lower BW= 1.8 mmol/L, upper BW = 1.4 mmol/L. Variable names in bold correspond to cases where the estimated impact coefficient (jump) is significant at 10% level.

Figure 8: Impacts on children growth according to age in months



4.4 Education

Table 6 presents the estimated impact of GDM diagnosis on EYFSP and KS1 tests.

First, for the EYFSP, out of 17 areas, we centre our analysis on literacy and writing, which are usually discussed in the literature. We assess them separately but also with the derived single index. We found that in all areas, there is a negative coefficient associated with having a mother who was diagnosed (and treated) with GDM. The negative difference is significant for the mathematics area and the four core indicators' compound index. It is important to realise that the results per area are categorical; therefore the exact coefficient does not provide a clear interpretation of the magnitude. Alternatively, we derive another measure: the number of areas where respondents attained at least the "expected" classification, and the number of "exceeding". With this last variable, those below the threshold attain at least 2.5 exceeding, but this number is just 1.6 for those diagnosed with GDM. Table 11 considers the discrete version of the results for each category (emerging, expected, exceeding). There we confirm that there is a reduction in the probability of obtaining exceeding results in the mathematics areas.

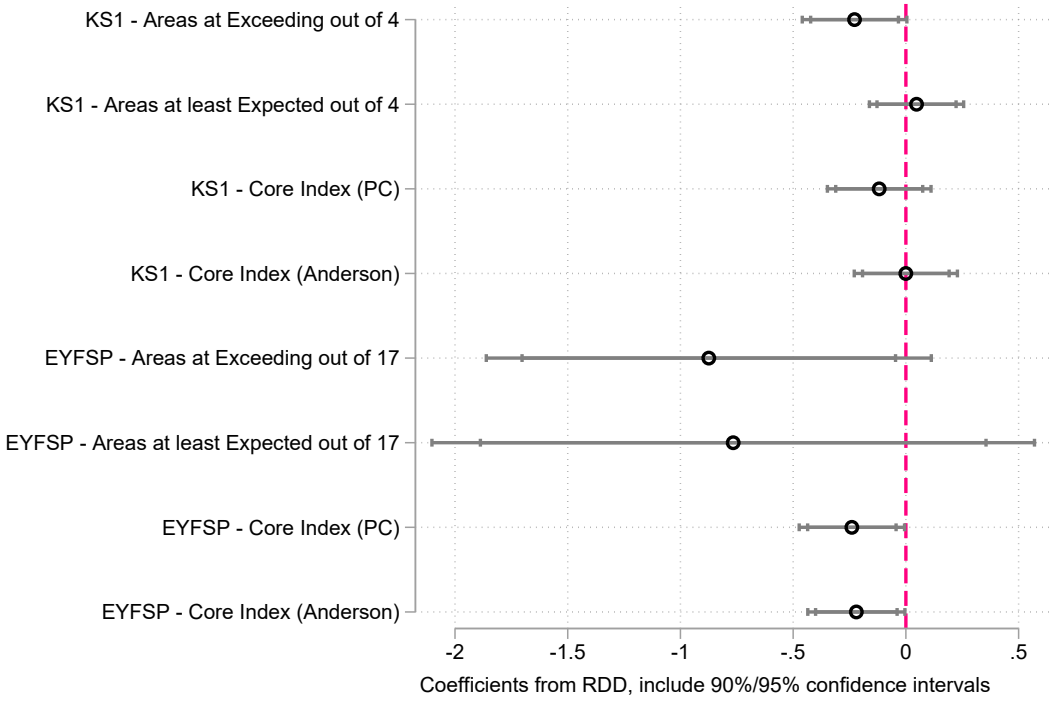
For Key Stage 1, we consider reading, writing, mathematics and science areas (the 2016 *English* area is not taken into account). While negative, differences are neither significant per area nor with the aggregate index. However, just as with EYFSP, the central difference is the number of areas in which an exceeding result is attained. This is a bit more than half (0.66) for the student born from a GDM undiagnosed mother, but these chances are reduced in 34% (to 0.43) if the mother was diagnosed. As with the EYFSP, table 11 considers the discrete version of each category's results (emerging, expected, exceeding). There we observe a reduction in the probability of exceeding for mathematics and writing.

Table 6: Impact on children standardised test results

Variable	(1) Mean	(2) N left	(3) N right	(4) Jump	(5) SE	(6) p-val
EYFSP - Literacy: Reading	1.888	2012	348	-0.127	0.079	0.1081
EYFSP - Literacy: Writing	1.793	2012	348	-0.100	0.079	0.2079
EYFSP- Mathematics: Numbers	1.869	2012	348	-0.164	0.071	0.0220
EYFSP- Mathematics: Shapes, space and measures	1.853	2012	348	-0.155	0.068	0.0233
EYFSP - Core Index (Anderson)	0.091	2012	348	-0.220	0.110	0.0455
EYFSP - Core Index (PC)	0.100	2012	348	-0.240	0.119	0.0444
EYFSP - Core Areas at least Expected out of 4	2.679	2158	365	-0.215	0.202	0.2874
EYFSP - Core Areas at Exceeding out of 4	0.488	2158	365	-0.170	0.129	0.1872
EYFSP - Areas at least Expected out of 17	13.434	2012	348	-0.766	0.682	0.2612
EYFSP - Areas at Exceeding out of 17	2.562	2012	348	-0.874	0.504	0.0826
KS1 - Reading	2.143	2142	364	-0.029	0.064	0.6539
KS1 - Writing	2.057	2138	363	-0.104	0.064	0.1029
KS1 - Mathematics	2.082	2144	364	-0.036	0.064	0.5741
KS1 - Science	1.060	2158	365	0.010	0.032	0.7532
KS1 - Core Index (Anderson)	0.053	2158	365	-0.000	0.117	0.9987
KS1 - Core Index (PC)	0.094	2138	363	-0.118	0.117	0.3130
KS1 - Areas at least Expected out of 4	2.681	2138	363	0.047	0.107	0.6578
KS1 - Areas at Exceeding out of 4	0.658	2138	363	-0.228	0.118	0.0544

Notes: Lower BW= 1.8 mmol/L, upper BW = 1.4 mmol/L. EYFSP core areas are literacy (reading; writing) and mathematics (numbers; shapes, space and measures). Variable names in bold correspond to cases where the estimated impact coefficient (jump) is significant at 10% level.

Figure 9: Impacts on education



4.5 Nutrition and child skills

Panel A of table 7 presents results on the SDQ instrument's subscales. More emotional symptoms are reported for those above the cutoff with the corresponding bandwidths. If we consider the optimal bandwidth, there is evidence of more conduct problems as well.

Panel B of the table is about nutrition: if the respondent is classified above the median frequency consumption of a particular food group. We observe less consumption of high-fat chips and more vegetables. If we consider the optimal bandwidth, evidence for more consumption of less high-fat chips and more non-processed meat is also obtained.

Table 7: Impact on children nutrition and behaviours age 6 months to 3 years. BiB1000 sample.

Variable	(1) Mean	(2) N left	(3) N right	(4) Jump	(5) SE	(6) p-val
Panel A: Strengths and Difficulties (SDQ, 3 years)						
Emotional symptoms	1.608	269	48	0.909	0.492	0.0647
Conduct problems	2.922	269	48	0.834	0.596	0.1617
Hyperactivity/inattention	4.765	269	48	-0.366	0.801	0.6479
Peer problems	2.196	269	48	-0.254	0.544	0.6407
Total difficulties	11.490	269	48	1.123	1.694	0.5074
High or above total difficult	0.255	269	48	0.133	0.127	0.2983
Prosocial	7.373	269	48	0.258	0.596	0.6658
Panel B: Food groups (3 years)						
High fat milk	0.725	264	48	-0.000	0.181	0.9981
Low fat milk	0.392	264	48	-0.187	0.165	0.2574
High fat chips	0.549	264	48	-0.320	0.151	0.0343
Low fat chips	0.471	264	48	0.217	0.159	0.1717
Non-processed meat	0.569	264	48	0.154	0.160	0.3362
Processed meat	0.431	264	48	0.082	0.146	0.5758
Vegetables	0.843	269	48	0.197	0.093	0.0333
Fruit	0.510	269	48	0.116	0.168	0.4902
Water	1.000	264	48	0.000	0.000	0.5746
Low sugar cereals	0.529	264	48	-0.036	0.172	0.8356

Notes: Lower BW= 1.8 mmol/L, upper BW = 1.4 mmol/L. Variable names in bold correspond to cases where the estimated impact coefficient (jump) is significant at 10% level.

5 Discussion and conclusions

We explored the consequences of diagnosis and treatment for GDM in early childhood. We did so by comparing children born from mothers with an OGTT slightly above and below the standard diagnostic threshold, applied to data from the Born in Bradford study.

Results of this analysis are restricted to individuals near the diagnosis cutoff. This means that we cannot extrapolate results to mothers with high levels of the OGTT, but it is clear that treatment is a requirement in those cases. Our aim is precisely to consider those cases where the diagnosis and subsequent treatment is less clear.

Concerning clinical outcomes, we found a significant reduction in the probability of having macrosomic babies (-5 pp.). Treated babies also had six days less of total gestational time, due to a higher probability of programmed labour (+ 41 pp.), but without an increase on the odds of preterm births or in c-sections. At birth, mothers had a reduced cardiovascular risk reflected in lower diastolic blood pressure at labour, triglycerides and cholesterol HDL ratio. Although children of treated mothers have a lower BMI soon after birth, these differences disappear after age 3; nonetheless, diagnosed children have reduced systolic blood pressure (- 3.4 mmHg) by age 5. Results above might not be driven only by the clinical treatments at

pregnancy, but also on sustained dietary changes. In the reduced BiB1000 sample, families of treated children consume less often high-fat chips and consume more vegetables.

However, when studying educational outcomes, we found that - for children of diagnosed mothers - there is an average reduction in the achievement level as measured by the Early Years Stage Foundation Profiles (EYFSP) at age 5, in areas related to mathematics. These differences seem to disappear when the children are administered the Key Stage 1 tests; however, we cannot rule out that differences in the tests' content drive the results. Additionally, although in a reduced subsample (BiB1000), we found evidence of emotional symptoms and conduct problems.

From the results above, we conclude that GDM diagnosis and treatment have a clear benefit on outcomes at birth, and there is some evidence of benefits by the age of 5. However, results on educational outcomes let us think that the treatment might have an adverse impact on school performance that requires further attention.

References

Akhter, Z., J. Rankin, D. Ceulemans, L. Ngongalah, R. Ackroyd, R. Devlieger, R. Vieira, and N. Heslehurst (2019). Pregnancy after bariatric surgery and adverse perinatal outcomes: A systematic review and meta-analysis. *PLoS medicine 16*(8).

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

Anderson, M. 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).

Baranov, V., S. Bhalotra, P. Biroli, and J. Maselko (2020). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American Economic Review 110*(3), 824–59.

Barbour, L. A., C. Scifres, A. M. Valent, J. E. Friedman, T. A. Buchanan, D. Coustan, K. Aagaard, K. L. Thornburg, P. M. Catalano, H. L. Galan, et al. (2018). A cautionary response to smfm statement: pharmacological treatment of gestational diabetes. *American journal of obstetrics and gynecology 219*(4), 367–e1.

Barker, D. J. (1990). The fetal and infant origins of adult disease. *BMJ: British Medical Journal 301*(6761), 1111.

Barreca, A. I., M. Guldi, J. M. Lindo, and G. R. Waddell (2011). Saving babies? revisiting the effect of very low birth weight classification. *The Quarterly Journal of Economics 126*(4), 2117–2123.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Calonico, S., M. D. Cattaneo, R. Titiunik, et al. (2014). Robust data-driven inference in the regression-discontinuity design. *Stata Journal 14*(4), 909–946.

Cattaneo, M. D., M. Jansson, and X. Ma (2016). rddensity: Manipulation testing based on density discontinuity. *The Stata Journal (ii)*, 1–18.

Cattaneo, M. D., M. Jansson, and X. Ma (2017). Simple local polynomial density estimators. *University of Michigan, Working Paper*.

Dalamaga, M., S. H. Chou, K. Shields, P. Papageorgiou, S. A. Polyzos, and C. S. Mantzoros (2013). Leptin at the intersection of neuroendocrinology and metabolism: current evidence and therapeutic perspectives. *Cell metabolism 18*(1), 29–42.

Durnwald, C. and E. F. Werner (2018). Gestational diabetes mellitus: Glycemic control and maternal prognosis. *UpToDate: https://www. uptodate. com/contents/gestational-diabetes-mellitus-glycemic-control-and-maternal-prognosis*.

Duryea, E. L., B. M. Casey, D. D. McIntire, and D. M. Twickler (2017). The fl/ac ratio for prediction of shoulder dystocia in women with gestational diabetes. *The Journal of Maternal-Fetal & Neonatal Medicine 30*(19), 2378–2381.

Farrar, D., L. Fairley, G. Santorelli, D. Tuffnell, T. A. Sheldon, J. Wright, L. van Overveld, and D. A. Lawlor (2015). Association between hyperglycaemia and adverse perinatal outcomes in south asian and white british women: analysis of data from the born in bradford cohort. *The lancet Diabetes & endocrinology 3*(10), 795–804.

Farrar, D., M. Simmonds, M. Bryant, T. A. Sheldon, D. Tuffnell, S. Golder, F. Dunne, and D. A. Lawlor (2016). Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis. *bmj 354*, i4694.

GOV.UK (2014a, Dec). The national curriculum.

GOV.UK (2014b, Dec). The national curriculum: Key stage 1 and 2.

GOV.UK (2015, May). Early years foundation stage.

Hehir, M. P., N. Burke, G. Burke, F. M. Breathnach, F. M. McAuliffe, J. J. Morrison, M. Turner, S. Dornan, J. R. Higgins, A. Cotter, et al. (2019). 1047: Increased abdominal circumference to head circumference ratio in late pregnancy is predictive of shoulder dystocia. *American Journal of Obstetrics & Gynecology 220*(1), S671–S672.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

IADPSG (2010). International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes care 33*(3), 676–682.

Lawlor, D. A., J. West, L. Fairley, S. M. Nelson, R. S. Bhopal, D. Tuffnell, D. J. Freeman, J. Wright, D. C. Whitelaw, and N. Sattar (2014). Pregnancy glycaemia and cord-blood levels of insulin and leptin in pakistani and white british mother–offspring pairs: findings from a prospective pregnancy cohort. *Diabetologia 57*(12), 2492–2500.

Lindsay, R., P. Catalano, and S. Nelson (2011). Diagnosis and treatment of gestational diabetes. *Royal College of Obstetricians and Gynaecologists Scientific Impact Paper No. 23*.

Lindsay, R. S. and M. R. Loeken (2017). Metformin use in pregnancy: promises and uncertainties. *Diabetologia 60*(9), 1612–1619.

Mahoney, S., M. Bryant, P. Sahota, and S. Barber (2018). Dietary intake in the early years and its relationship to bmi in a bi-ethnic group: the born in bradford 1000 study. *Public Health Nutrition 21*(12), 2242–2254.

Metzger, B. E., M. Contreras, D. Sacks, W. Watson, S. L. Dooley, M. Foderaro, C. Niznik, J. Bjaloncik, P. Catalano, L. Dierker, et al. (2008). Hyperglycemia and adverse pregnancy outcomes. *New England Journal of Medicine 358*(19), 1991–2002.

Muris, P., C. Meesters, and F. van den Berg (2003). The strengths and difficulties questionnaire (sdq). *European child & adolescent psychiatry 12*(1), 1–8.

NICE (2015). Diabetes in pregnancy: management from preconception to the postnatal period (nice guideline ng3).

Papageorghiou, A. T., E. O. Ohuma, D. G. Altman, T. Todros, L. C. Ismail, A. Lambert, Y. A. Jaffer, E. Bertino, M. G. Gravett, M. Purwar, et al. (2014). International standards for fetal growth based on serial ultrasound measurements: the fetal growth longitudinal study of the intergrowth-21st project. *The Lancet 384*(9946), 869–879.

Persson, L. Å., K. M. Rasmussen, and H. Yang (2019). Nutrition for women and children—are we doing the right things in the right way? *PLoS medicine 16*(8).

Tarry-Adkins, J. L., C. E. Aiken, and S. E. Ozanne (2019). Neonatal, infant, and childhood growth following metformin versus insulin treatment for gestational diabetes: A systematic review and meta-analysis. *PLoS medicine 16*(8).

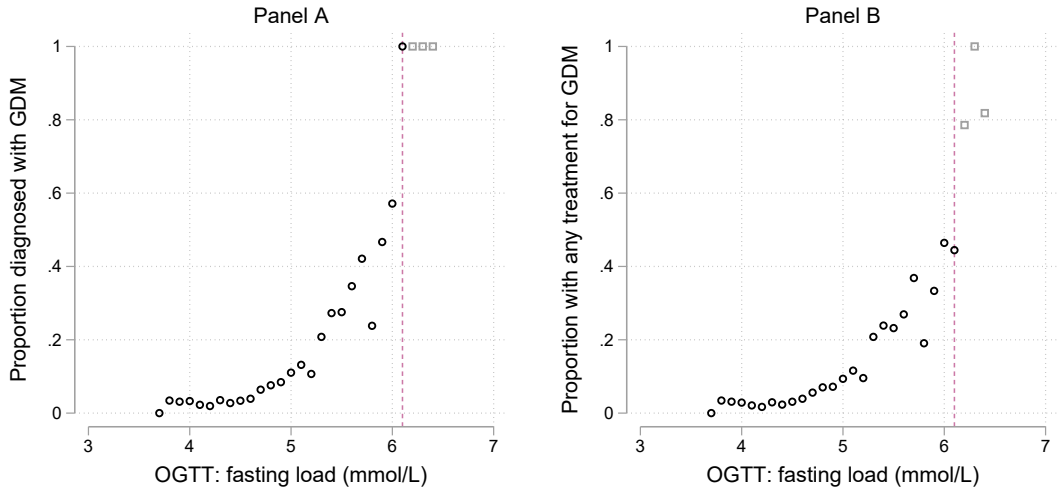
Wright, J., L. Fairley, R. McEachan, M. Bryant, E. Petherick, P. Sahota, G. Santorelli, S. Barber, D. A. Lawlor, N. Taylor, R. Bhopal, N. Cameron, J. West, A. Hill, C. Summerbell, A. Farrin, H. Ball, T. Brown, D. Farrar, and N. Small (2016). *Development and evaluation of an intervention for the prevention of childhood obesity in a multiethnic population: the Born in Bradford applied research programme*, Volume 4. Perth, Scotland: NIHR Journals Library, Programme Grants Appl Res.

Wright, J., N. Small, P. Raynor, D. Tuffnell, R. Bhopal, N. Cameron, L. Fairley, D. A. Lawlor, R. Parslow, E. S. Petherick, et al. (2012). Cohort profile: the born in bradford multi-ethnic family cohort study. *International journal of epidemiology 42*(4), 978–991.

Zubrick, S., N. Lucas, E. Westrupp, and J. Nicholson (2014). Parenting measures in the longitudinal study of australian children: construct validity and measurement quality, waves 1 to 4. Technical report.

A Fasting load

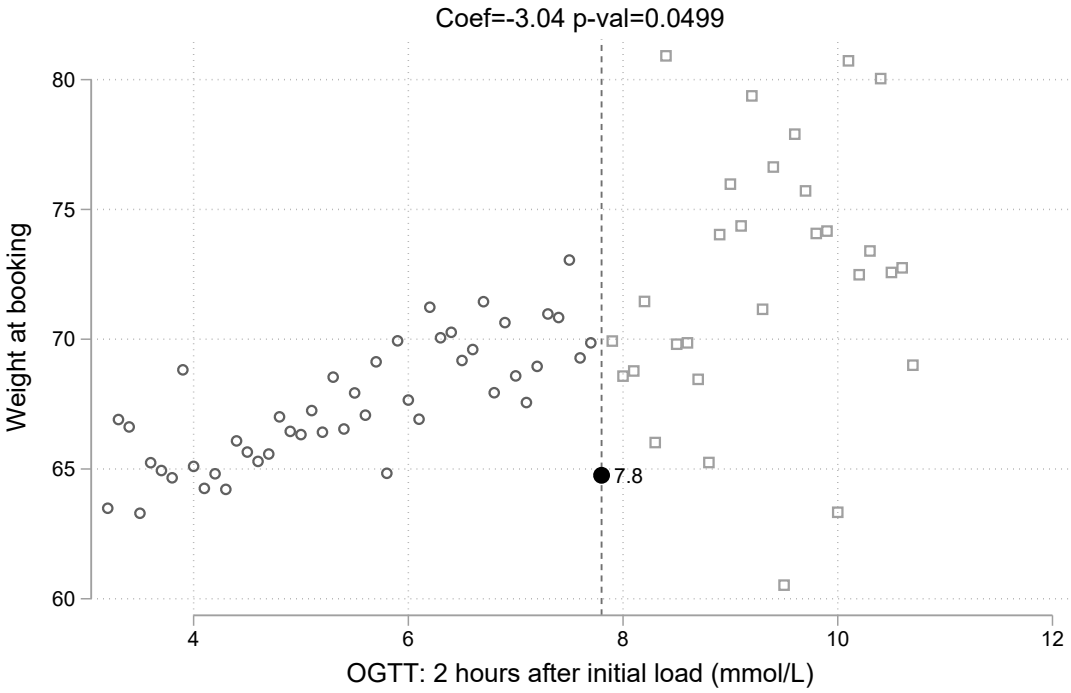
Figure 10: Proportion of individuals diagnosed and treated from GDM, according to their fasting OGTT test



B Removing observations exactly at the OGTT 2h post equal to 7.8 mmol/L

A particular characteristic of pregnancies was the 2h PL OGTT test was exactly 7.8 mmol/L is the low average weight at booking. This characteristic does not seem to be related to any particular rounding as there is no accumulation of the distribution of OGTT at such point (see figure 3). The implication is that this particular point for low bandwidths drives estimates towards a particular dimension, which is reflected in the unbalance of pre-intervention characteristics. To avoid this bias, we implemented the doughnut-hole RDD regression presented in the main text.

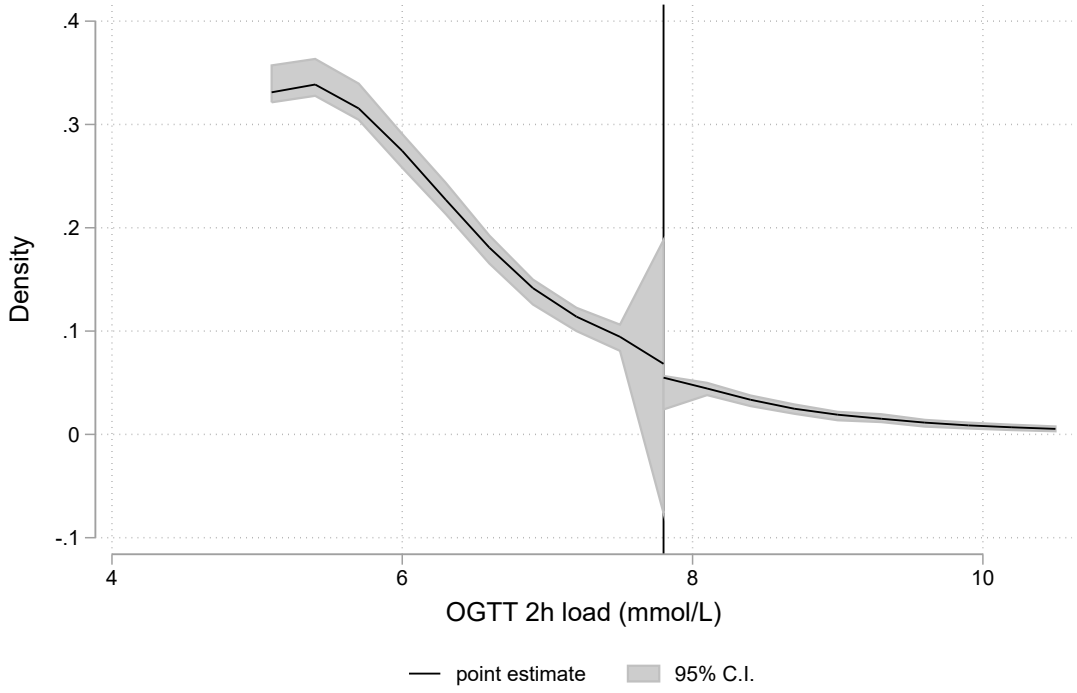
Figure 11: Average weight at booking according to the 2h post OGTT test



C Density test

The RDD design's main assumption is that it is random whether a respondent is just above or below the cutoff. We can assess this is a plausible assumption with a standard density test. We use the version designed by Cattaneo et al. (2017), who constructed a non-parametric density estimator using local polynomials which is valid at boundary points. Figure 12 shows a graphic version of this test (implemented using the Cattaneo et al. (2016) routine), where the difference in the densities is -0.2287 (bandwidths of 0.608 and 0.663 to the left and right), which cannot be rejected to be zero [pval=0.8191]. Here our 2H PL OGTT measure is the transformed version where the 7.8 mmol/L observations were removed, and the entire measure is moved left, so there is no "hole" on the distribution.

Figure 12: Density discontinuity test



D Alternative Bandwidths

Table 8: Impact on children characteristics at birth (Alternative Bandwidths)

		Optimal BW			$h_l = 1.8$, $h_r = 1.4$	$h = 1$	$h = 2$
Variable	Mean	h_l^*	h_r^*	Jump	Jump	Jump	Jump
Weight at birth (Kg)	3.309	1.829	1.497	-0.249 ***	-0.242 ***	-0.160 *	-0.279 ***
Weight at birth < 2.5 Kg (LBW)	0.069	2.459	2.090	0.063 *	0.058	0.030	0.063 *
Weight at birth ≥ 4 Kg (Macrosomic)	0.102	2.627	1.109	-0.050 *	-0.055 *	-0.005	-0.067 ***
Gestation in days	276.353	1.323	2.344	-5.789 ***	-6.230 ***	-3.951 **	-6.602 ***
Gestational weeks < 37 (pre-term)	0.071	2.765	1.943	0.007	0.010	-0.006	0.007
Gestational weeks ≥ 42 (post-term)	0.013	.	.		-0.007	0.005	-0.009
Non spontaneous onset of labour	0.302	1.983	2.074	0.394 ***	0.411 ***	0.411 ***	0.395 ***
C-section	0.259	2.526	1.390	0.040	0.029	-0.028	0.028
APGAR at 1 minute	8.387	2.466	1.665	0.187	0.204	0.142	0.245 *
APGAR at 5 minutes	9.079	2.187	1.277	0.018	0.026	0.032	0.012
Insulin	6.155	2.169	1.390	-1.144	-1.310	-1.006	-1.257 *
Cord blood leptin	13.820	1.351	1.120	-5.113 ***	-5.074 ***	-3.821 *	-5.203 ***

Notes: Optimal bandwidth from Calonico et al (2014) algorithm.

Table 9: Impact on maternity outcomes (Alternative Bandwidths)

		Optimal BW			$h_l = 1.8$, $h_r = 1.4$	$h = 1$	$h = 2$
Variable	Mean	h_l^*	h_r^*	Jump	Jump	Jump	Jump
Any problem at birth	0.243	2.262	1.889	-0.059	-0.062	-0.082	-0.060
Systolic blood pressure at birth	129.395	1.887	1.667	-0.682	-0.732	1.357	-0.899
Diastolic blood pressure at birth	76.689	1.629	1.464	-2.194 *	-2.437 *	-0.537	-2.578 **
High blood pressure at birth	0.284	2.238	1.811	-0.084	-0.087	-0.005	-0.087 *
Cholesterol HDL Ratio	3.371	1.810	2.082	-0.207 **	-0.215 **	-0.239	-0.193 **
Cholesterol	6.264	2.197	1.764	-0.090	-0.134	-0.259	-0.113
HDL	1.930	1.673	1.685	0.087 *	0.090 *	0.067	0.081 *
LDL	3.371	2.102	1.875	-0.137	-0.186	-0.270	-0.147
Triglyceride	2.255	2.231	2.065	-0.170 *	-0.185 *	-0.244 *	-0.177 *

Notes: Optimal bandwidth from Calonico et al (2014) algorithm.

Table 10: Impact on child growth (Alternative Bandwidths)

			Optimal BW			$h_l = 1.8$, $h_r = 1.4$	$h = 1$	$h = 2$
Variable	Age	Mean	h_l^*	h_r^*	Jump	Jump	Jump	Jump
Child BMI-for-age z score (UK1990)	.	At birth	.	.				
Child BMI-for-age z score (UK1990)	-0.653	4 to 13	1.516	2.073	-0.117	-0.006	-0.002	-0.063
Child BMI-for-age z score (UK1990)	0.058	22 to 27	2.402	1.237	-0.130	-0.243	-0.308	-0.275
Child BMI-for-age z score (UK1990)	0.203	54 to 60	2.953	1.213	0.120	0.078	0.004	0.040
Obesity Indicator	0.014	1 to 3	1.985	1.530	0.009	0.009	0.012	0.005
Obesity Indicator	0.023	4 to 13	1.529	2.840	0.015	0.016	0.028	0.016
Obesity Indicator	0.081	22 to 27	2.180	1.882	0.006	0.016	0.061	0.005
Obesity Indicator	0.052	54 to 60	2.362	1.471	0.011	0.013	-0.002	0.010
Diastolic blood pressure (school nurse)	63.093	54 to 60	2.088	1.328	-1.581	-1.806	-1.581	-1.958
Systolic blood pressure (school nurse)	99.508	54 to 60	1.314	2.069	-4.712 ***	-3.386 **	-5.457 **	-3.148 **

Notes: Optimal bandwidth from Calonico et al (2014) algorithm.

Table 11: Impact on children standardised test results (Alternative Bandwidths)

Variable	Mean	Optimal BW h_l^*	Optimal BW h_r^*	Optimal BW Jump	$h_l = 1.8$, $h_r = 1.4$ Jump	$h = 1$ Jump	$h = 2$ Jump
EYFSP - Literacy: Reading	1.888	1.599	1.716	-0.134 *	-0.127	-0.219 **	-0.099
EYFSP - Literacy: Writing	1.793	2.047	1.829	-0.080	-0.100	-0.101	-0.077
EYFSP- Mathematics: Numbers	1.869	1.606	1.279	-0.185 **	-0.164 **	-0.200 **	-0.130 **
EYFSP- Mathematics: Shapes, space and measures	1.853	2.162	1.316	-0.147 **	-0.155 **	-0.201 **	-0.131 **
EYFSP - Core Index (Anderson)	0.091	1.832	1.412	-0.215 **	-0.220 **	-0.287 *	-0.177 *
EYFSP - Core Index (PC)	0.100	1.778	1.409	-0.240 **	-0.240 **	-0.315 *	-0.193 *
EYFSP - Core Areas at least Expected out of 4	2.679	3.156	1.356	-0.122	-0.215	-0.297	-0.165
EYFSP - Core Areas at Exceeding out of 4	0.488	1.554	1.317	-0.216	-0.170	-0.393 **	-0.111
EYFSP - Areas at least Expected out of 17	13.434	2.002	1.580	-0.685	-0.766	-0.425	-0.652
EYFSP - Areas at Exceeding out of 17	2.562	1.577	1.669	-1.002 *	-0.874 *	-1.406 **	-0.753
KS1 - Reading	2.143	2.223	2.384	-0.052	-0.029	0.006	-0.041
KS1 - Writing	2.057	2.507	1.965	-0.078	-0.104	-0.050	-0.087
KS1 - Mathematics	2.082	2.516	1.881	-0.042	-0.036	-0.002	-0.045
KS1 - Science	1.060	1.942	1.990	0.003	0.010	0.051	0.002
KS1 - Core Index (Anderson)	0.053	2.235	1.631	-0.026	-0.000	0.165	-0.031
KS1 - Core Index (PC)	0.094	2.403	1.983	-0.115	-0.118	-0.025	-0.119
KS1 - Areas at least Expected out of 4	2.681	1.585	2.276	0.047	0.047	0.143	0.037
KS1 - Areas at Exceeding out of 4	0.658	1.680	1.799	-0.230 **	-0.228 *	-0.184	-0.220 **

Notes: Optimal bandwidth from Calonico et al (2014) algorithm.

Table 12: Results based on BiB1000 study (Alternative Bandwidths)

Variable	Mean	Optimal BW h_l^*	Optimal BW h_r^*	Optimal BW Jump	$h_l = 1.8$, $h_r = 1.4$ Jump	$h = 1$ Jump	$h = 2$ Jump
Weight at birth (Kg)	3.283	1.853	1.269	-0.204	-0.232	-0.043	-0.257 **
Gestation in days	277.119	1.703	1.492	-3.314	-3.848 *	1.577	-3.993 **
EYFSP - Core Index (Anderson)	-0.058	1.113	1.732	-0.201	-0.029	-0.186	0.046
BMI at 6 months	26.887	1.614	1.052	-2.086	-2.261	-3.865 *	-2.487
BMI at 12 months	26.446	1.246	1.019	-0.659	-0.584	-1.564	-1.194
BMI at 24 months	26.206	1.454	1.612	-0.693	-0.704	-2.344	-1.175
BMI at 36 months	26.503	1.558	1.024	0.693	0.256	-1.015	-0.562
Mother general health index (6 months)	1.679	1.027	0.677	-0.304 **	-0.215 **	-0.261 *	-0.151 *
Mother general health index (18 months)	1.615	1.544	2.408	-0.072	-0.054	-0.034	-0.036
Parent practices index: Affection (6 months)	3.263	1.538	0.690	0.171	0.198	0.296 *	0.162
Parent practices index: Perception (6 months)	8.943	2.346	1.101	-0.870 *	-0.750	-0.883	-0.518
Parent practices index: Affection (24 months)	3.500	1.538	1.029	-0.109	-0.142	-0.001	-0.158
Parent practices index: Perception (24 months)	5.022	1.537	1.858	-0.407	-0.470	-0.368	-0.455
Infant characteristics index (6 months)	2.364	1.611	1.071	0.189	0.198	-0.003	0.198
SDQ 3y: Emotional symptoms	1.608	0.916	0.958	1.222 *	0.909 *	1.249 **	0.688
SDQ 3y: Conduct problems	2.922	1.152	1.098	1.537 **	0.834	1.694 **	0.832
SDQ 3y: Hyperactivity/inattention	4.765	1.255	0.918	-0.440	-0.366	-0.456	-0.302
SDQ 3y: Peer problems	2.196	1.829	1.109	-0.236	-0.254	-0.866	-0.162
SDQ 3y: Total difficulties	11.490	1.177	1.028	2.070	1.123	1.622	1.055
SQD 3y: High or above total difficult.	0.255	1.257	0.871	0.212	0.133	0.246	0.119
SDQ 3y: Prosocial	7.373	2.014	0.838	-0.030	0.258	1.122	0.044
Food group: High fat milk	0.725	1.644	0.933	0.033	-0.000	0.224	-0.049
Food group: Low fat milk	0.392	1.400	1.150	-0.216	-0.187	-0.245	-0.107
Food group: High fat chips	0.549	1.593	1.031	-0.325 **	-0.320 **	-0.339	-0.275 **
Food group: Low fat chips	0.471	1.507	0.982	0.296 *	0.217	0.406 *	0.179
Food group: Non-processed meat	0.569	1.210	0.824	0.514 ***	0.154	0.641 ***	0.033
Food group: Processed meat	0.431	2.163	0.572	0.035	0.082	0.292 *	0.066
Food group: Vegetables	0.843	1.599	0.458	0.230 **	0.197 **	0.264 *	0.147 *
Food group: Fruit	0.510	1.375	0.838	0.324	0.116	0.311	0.047
Food group: Water	1.000	.	.		0.000	-0.000	-0.000
Food group: Low sugar cereals	0.529	2.043	1.032	-0.002	-0.036	-0.119	-0.082

Notes: Optimal bandwidth from Calonico et al (2014) algorithm.

Table 13: Health records after birth (Alternative Bandwidths)

Variable	Mean	Optimal BW h_l^*	Optimal BW h_r^*	Optimal BW Jump	$h_l = 1.8$, $h_r = 1.4$ Jump	$h = 1$ Jump	$h = 2$ Jump
Mother: Diabetes Mellitus	0.118	1.982	1.108	-0.069	-0.093	-0.232	-0.095
Mother: Hypertensive heart disease	0.044	0.423	1.648	0.160 **	0.035	0.024	0.007
Mother: Angina pectoris	0.116	1.937	0.992	-0.036	-0.026	-0.039	-0.018
Obesity	0.024	.	.		-0.030 *	-0.021	-0.031 **
Lactose intolerance	0.024	.	.		0.093	0.127	0.076
Disorders of galactose metabolisme	0.011	0.913	1.951	0.025	0.020	0.023	0.018
Gilbert and Crigler-Najjar syndrome	0.026	0.983	1.768	0.053	0.025	0.049	0.018
Cystic fibrosis	0.013	1.246	1.503	0.012	-0.000	0.011	0.003
Developmental disorder of scholastic	0.015	1.212	2.300	0.011	0.006	0.025	0.004

Notes: Optimal bandwidth from Calonico et al (2014) algorithm.