DISCUSSION PAPER SERIES

IZA

IZA DP No. 9810

Heterogeneous Effects of Medical Interventions on the Health of Low-Risk Newborns

N. Meltem Daysal
Mircea Trandafir
Reyn van Ewijk

March 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Heterogeneous Effects of Medical Interventions on the Health of Low-Risk Newborns

N. Meltem Daysal
University of Southern Denmark and IZA

Mircea Trandafir
University of Southern Denmark and IZA

Reyn van Ewijk
University of Mainz

Discussion Paper No. 9810
March 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9810
March 2016

ABSTRACT

Heterogeneous Effects of Medical Interventions on the Health of Low-Risk Newborns*

We investigate the impact of early-life medical interventions on low-risk newborn health. A policy rule in The Netherlands creates large discontinuities in medical treatments at gestational week 37. Using a regression discontinuity design, we find no health benefits from additional treatments for average newborns. However, there is substantial heterogeneity in returns to treatments with significant health benefits for newborns in the lowest income quartile and no benefits in higher income quartiles. This seems due to increased maternal stress from referral to an obstetrician among higher-income mothers, heterogeneous effects of home births, and potential difficulties in risk screening among low-income women.

JEL Classification: I11, I12, I18, J13

Keywords: medical interventions, birth, heterogeneity, mortality

Corresponding author:

Mircea Trandafir
Department of Business and Economics
University of Southern Denmark
Campusvej 55
DK-5230 Odense M
Denmark
E-mail: mircea.trandafir@sam.sdu.dk

* Douglas Almond, Abby Alpert, Kitt Carpenter, John Cawley, Gordon Dahl, Hendrik Jürges, Amanda Kowalski, Amalia Miller, Hessel Oosterbeek, Martin Salm, Diane Whitmore Schanzenbach, Emilia Simeonova, Kosali Simon, Arthur van Soest, and participants at Aarhus University, Cornell University, Impaq International, Lund University, Stockholm School of Economics, Uppsala University, University of Copenhagen, University of Hamburg, University of Mainz, University of Munich, University of Southern Denmark, Universite de Sherbrooke, University of York, Tilburg University, and the meetings of the American Society of Health Economists, American Health Economics Workshop, European Economic Association, EuroEpi, European Workshop on Econometrics and Health Economics, iHEA-ECHE provided helpful comments and discussions. We thank Perined for making the data available. Tjeerd van Campen and Iris van Dam provided able research assistance. The authors bear sole responsibility for the content of this paper.

1 Introduction

It is a stylized fact that health care expenditures vary considerably throughout the developed world, both across and within countries (Skinner, 2011). These geographic variations in health care spending, however, are generally not associated with corresponding differences in health outcomes (e.g., Baicker and Chandra, 2004; Fuchs, 2004; Stukel et al., 2005) This pattern is in stark contrast to the well-documented health benefits of a range of medical technologies, broadly defined as pharmaceutical treatments, medical devices and procedures (e.g., McClellan and Newhouse, 1997; Almond et al., 2010; Garthwaite and Duggan, 2012; Daysal et al., 2015). Economists have recently emphasized the role of treatment heterogeneity in reconciling these seemingly contradictory findings. As the argument goes, cross-sectional studies identify the effects of incremental spending, which may be very different than the impact of treatments on the marginal patient. In this paper, we investigate the heterogeneity in the impact of early-life medical interventions on the short-term health of low-risk newborns using a unique confidential dataset from the Netherlands.

Focusing on early-life medical interventions is important for several reasons. First, spending for the very young increased substantially faster than spending for the average individual. For example, during the period 1960–1990, per capita spending in the US on infants under 1 year old increased by 9.8 percent per year whereas annual spending on individuals aged 1 to 64 increased by only 4.7 percent (Cutler and Meara, 1998). Second, it is widely accepted that changes in medical technologies are the main driver of medical cost growth, both in general and in the specific case of childbirth (Newhouse, 1992; Cutler and Meara, 1998). Third, any gains from survival are much larger in the case of newborns than for adults. Finally, understanding the heterogeneity in returns to medical interventions for low-risk births is especially important given the current policy debates on shifting these births from more costly to less costly childbirth technologies such as midwifery care and home births.

Empirical estimation of the returns to medical interventions is complicated by selection issues. Even among observably low-risk women, those with worse expected birth outcomes usually receive more (intensive) treatments, leading to biased estimates in simple regressions. In order to eliminate this bias, we exploit a policy rule in the Netherlands that provides exogenous variation in the medical treatments administered to low-risk births. The Dutch system is unique in its division between the primary care provided by midwives and the secondary care provided by obstetricians (OB/GYN). Low-risk women, i.e., women without known medical risk factors, start their pregnancy under the supervision of a midwife and stay under the supervision of a midwife as long as no risk factors appear. Their delivery is supervised by a midwife, who is prohibited by law from performing any medical intervention. The birth can take place either at home or in a hospital, and in both cases no OB/GYN is present. However, if labor is premature (i.e., before 37 completed gestational weeks), the woman should be referred to an obstetrician. In this case, the OB/GYN supervises the delivery, which always takes place in a hospital. Thus, the "week-37 rule" generates a discontinuity at 37 completed gestational weeks in three important medical inputs: the medical professional supervising the delivery (OB/GYN instead of midwife), the location of delivery (hospital versus home), and all the medical interventions that physicians are allowed to perform during and immediately after birth (e.g., use of forceps and vacuum, administration of antibiotics). This motivates the use of a regression discontinuity (RD) design.[1]

We start by investigating effects on the average low-risk newborn. We show that the week-37 rule generates substantial variation in all our measures of medical technologies/inputs. For example, the probability that a spontaneous low-risk birth is supervised by an obstetrician increases by 40 percentage points below the 37-week threshold. Similarly, newborns slightly below the week-37 cutoff are 26 percentage points more likely be delivered in a hospital and 12.5

[1]It is worth noting that the rate of planned C-sections is generally very low in the Netherlands and that planned C-sections do not occur among low-risk women. Only around 7 percent of all births are primary C-sections (i.e., planned before the start of delivery). Most of these are for medical reasons and among women not classified as low-risk. Elective C-sections for non-medical reasons are very rare and virtually non-existent around the 37-week cutoff. As detailed later in the paper, all planned C-sections are excluded from our analysis sample.

percentage points more likely to be admitted to a neonatal intensive care unit (NICU). These estimates are economically large and correspond to increases of 30–110% when compared to the mean above the cutoff. Despite the substantial variation in medical interventions, we do not find any significant differences in newborn health outcomes (7-day and 28-day mortality and likelihood of a low Apgar score) across the week-37 cutoff.

Average effects can mask significant variation in benefits across the population (Bitler et al., 2006). Therefore, we next turn to heterogeneity in returns to medical treatments. Given that previous literature documents large differences in infant mortality across different socio-economic groups (Case et al., 2002; Currie et al., 2007), we examine whether returns to medical treatments vary by socio-economic status as proxied by the average income in the postal code of residence of the mother.[2] We find that the discontinuities in medical treatments across the week-37 cutoff are similar across the income distribution. However, there are significant differences in the effects of these treatments on newborn health. Our results consistently indicate economically large health gains to preterm newborns in the lowest income quartile. In contrast, we find no significant health differences between preterm and at-term newborns in the other three income quartiles.

There are several channels that may explain the heterogeneity in the returns to childbirth technologies. First, Daysal et al. (2015) investigate the returns to one of the three medical inputs investigated in the present study: home versus hospital births. Using Dutch data, they show that home births increase newborn mortality rates and that these results seem driven by the poorest half of the population. We calculate that heterogeneity in health benefits of a hospital birth across the income distribution can only partly explain the heterogeneity in the returns to early-life medical treatments found in the present study. Second, the current risk selection system may be better suited to screening high-risk mothers among higher income women. We

[2]Postal codes in the Netherlands are much smaller than zip codes in the United States. We use 4-digit postal codes, which on average have 4,075 inhabitants and a land surface of 8.5 square kilometers (3.28 square miles). We do not have information on individual income or education.

provide anecdotal evidence that this may indeed be the case. Third, it is possible that low-income newborns are exposed to additional treatments in a timelier manner because their mothers tend to reside closer to a hospital. We rule out this explanation because we find no evidence of residential sorting by distance and income. Fourth, higher-income mothers may suffer from differentially higher maternal distress when referred to an obstetrician due to their preference for a midwife-supervised home birth (de Jonge et al., 2009). We find evidence in support of this explanation. Using meconium staining as a proxy, we find increased maternal distress among mothers of preterm newborns in higher-income quartiles, which may counteract any potential benefits from additional medical treatments. In lower-income quartiles, in contrast, we find no increased maternal distress as a result of referrals to an obstetrician.

Our study fits broadly in the previous economics research on returns to medical technologies. A large part of this literature investigates treatments for adults, such as heart attack (Cutler et al., 1998; Skinner et al., 2006) or HIV/AIDS patients (Duggan and Evans, 2008). More recently, a growing number of papers examine returns to early-life medical interventions, with a special focus on treatments for very low birth weight children. Increased treatments for this group are generally shown to reduce mortality (Cutler and Meara, 2000; Almond et al., 2010; Bharadwaj et al., 2013; Breining et al., 2015). Research on the returns to medical interventions for low-risk infants is limited with mixed results. Using state-level variation in mandated insurance coverage of midwifery services, Miller (2006) finds that midwifery-promoting public policies had no significant effect on maternal mortality or Apgar scores, but were associated with lower neonatal mortality. Almond and Doyle (2011) show that longer hospital stays do not affect infant health outcomes after uncomplicated deliveries. Our paper is most closely related to Daysal et al. (2015), who use an instrumental variables strategy to find that giving birth in a hospital (as opposed to home) leads to reductions in the mortality of low-risk newborns. Although both papers find health benefits from early-life medical interventions for low-income low-risk newborns, they are different along several important dimensions. First, our study examines a much wider range of medical technologies, extending the relevance of our findings to other countries where home births are not as common as in the Netherlands. Second, we rely on a regression discontinuity design generated by the week-37 rule. Thus, the marginal infant in the two studies is likely drawn from different parts of the distribution of low-risk newborns. Finally, the main focus of this paper is on the heterogeneity in the returns to early-life medical treatments, which was only briefly discussed in Daysal et al. (2015).

Our paper also contributes to the growing interest, both within economics and among medical researchers, in understanding the heterogeneity in returns to medical treatments (Kravitz et al., 2004; Chandra and Skinner, 2012). To the best of our knowledge, the only paper to explicitly examine heterogeneity in returns to early-life medical interventions is Evans and Garthwaite (2012). The authors use changes in minimum postpartum stay laws in California to investigate the impact of postpartum length of stay on newborn health. Their results point to modest reductions in the probability of readmission for the average newborn, but also to substantial heterogeneity in the effects across the distribution of medical need. In particular, the health gains are found to be largest for infants with high *a priori* likelihood of longer stay. To the extent that low-income newborns in the Netherlands also have higher medical need, our results are consistent with those of Evans and Garthwaite (2012). In addition, we document heterogeneity in the returns to early-life medical interventions along a richer set of treatments.

Our results are relevant to the ongoing policy debates on effective health policy. The fact that medical treatments improve newborn outcomes even among low-risk women living in a developed country cautions against designing "one-size-fits-all" policies. In addition, the heterogeneity in the returns to medical interventions suggests that a good understanding of the causes of this heterogeneity should go hand in hand with crafting policies about child birth technologies. This is especially important in light of the growing emphasis on cost reduction through increased use of physician extenders (Institute of Medicine, 2011).

2 The Dutch Obstetric System

Obstetric care in the Netherlands is guided by the principle that pregnancy and delivery are natural processes that do not require attendance by a (specialized) physician as long as there are no deviations from the perfectly normal course. The ability of midwives to fully provide care for uncomplicated pregnancies and deliveries was established as early as 1865 through the "Law of Medical Practice" and upheld in subsequent legislation. These laws also prohibit the use of any "obstetrical instruments" by midwives (Amelink-Verburg and Buitendijk, 2010). However, a clear separation between the roles of midwives and obstetricians was introduced only a century later. In 1958, with the clear goal of reducing medical expenditures, the Dutch National Health Insurance Board compiled a list of conditions that require a hospital admission in the area of maternity care. This list introduced the division between the primary care provided by midwives (or general practitioners in areas with no midwife practices) and the secondary care provided by specialized physicians such as obstetricians. It also set the foundation for risk selection, the principle that uncomplicated births should stay in primary care and that hospital admissions are necessary only in case of deviations from the normal course of pregnancy or labor. The list was updated over time and its use became explicit in 1973, when it was published as the "List of Obstetric Indications" (LOI) in the Dutch Textbook of Obstetrics and Gynecology (Amelink-Verburg and Buitendijk, 2010). Since then, the LOI is used to determine when referrals are made from primary to secondary care.

Currently, the Dutch maternity care system functions as follows. Pregnancies start under supervision of a midwife as long as none of the conditions described in the LOI are present. As long as no complications arise, midwives supervise the entire pregnancy, perform all checks, and attend the birth (Bais and Pel, 2006). If at least one condition in the LOI is found, then a referral to secondary care needs to be made at that point and the rest of the pregnancy and the birth is supervised by an OB/GYN. The LOI contains four types of criteria that lead to a referral: non-gynecological pre-existing conditions (e.g., diabetes, alcoholism or psychiatric disorders), gynecological pre-existing conditions, obstetric anamnesis (C-section, very premature births or severe complications during previous deliveries), and conditions arising or first diagnosed during pregnancy such as hyperemesis gravidarum, infections, plurality, gestational hypertension, or blood loss (CVZ, 2003). Referrals for reasons not listed in the LOI are not allowed and physician fees are not covered by insurance plans in such cases (CVZ, 2003). Finally, women are not allowed to directly contact an obstetrician.

This risk selection system divides delivering women into two groups. High-risk women are those referred to an OB/GYN at any point during pregnancy (before the onset of labor). Their prenatal care is provided by obstetricians from the moment of the referral and they are required to give birth in a hospital under the supervision of an OB/GYN. Low-risk women are those who do not have any LOI-listed conditions until the onset of labor. These women receive their prenatal care entirely from midwives and they can choose between a home and a hospital birth. In both cases, their deliveries are supervised by a midwife with no obstetrician present unless a complication arises during labor or during the delivery.

Among both high- and low-risk women, special medical guidelines exist in the case of prematurity, which is defined as the onset of labor before 37 completed gestational weeks from the last menstrual period. For example, many hospitals in the Netherlands regularly admit preterm infants for observation, and some hospitals administer antibiotics to women whose water breaks before week 37 in order to reduce the risk of infection (Schakel and Bekhof, 2010). In addition, in the case of low-risk women the LOI includes a rule (hereafter the "week-37 rule") requiring midwives to refer women whose labor starts or threatens to start prematurely to an obstetrician. These births then have to take place in a hospital under the supervision of the obstetrician, and both these women and their newborns have access to all the treatments that obstetricians can provide during and shortly after the birth.

To summarize, the week-37 policy rule generates plausibly exogenous variation in the medical professional attending the birth of low-risk women. This rule divides low-risk women into two groups, both of whom received their prenatal care from midwives: those delivering under the supervision of a midwife with no obstetrician present, and those who deliver under the supervision of an obstetrician. Given that obstetricians only deliver in hospitals, the rule also induces variation in the location of delivery. Finally, because midwives cannot perform any medical interventions, the week-37 rule also produces variation in the medical treatments available during and immediately after birth.

3 Empirical Strategy

We are interested in the heterogenous impact of early-life medical interventions on the health of low-risk newborns. To identify the effects, we exploit plausibly exogenous variation in early-life medical treatments due to the "week-37 rule" in a regression discontinuity (RD) design.

An RD design relies on the idea that if a policy requires a sharp and arbitrary cutoff for implementation and is based on a measure that is not perfectly controlled by the targeted individuals, then random variation around the cutoff will partly determine when the policy is implemented (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). The week-37 cutoff provides an ideal case for an RD design. It is based on an arbitrary threshold in the sense that there are no specific developmental changes that occur in the fetus or in the mother between day 258 and day 259. Kramer et al. (2012, p.111) note that "[i]nfants born before 20 weeks or at 37 or 38 weeks share many features with births at 20–36 weeks, including etiological and prognostic features," and thus conclude that the choice for the upper (37 weeks) and lower (20 or 22 weeks) bounds for defining a preterm birth are arbitrary. In addition, there is no evidence that any intervention (including hydration, antibiotics, or tocolytic therapy) can consistently delay delivery by more than 24–48 hours after the onset of labor (Norwitz and Caughey, 2011). This suggests that, in a sample of spontaneous births, expectant mothers cannot precisely manipulate the timing of their birth so as to control their assignment to different medical providers and treatments. As such, the variation in early-life medical interventions around the week-37 cutoff should be as good as random.

Our empirical strategy is described by the following local-linear regression:

$$Y_{iat} = f(a - 258) + \beta W37_a + u_{iat}, \tag{1}$$

where the unit of observation is infant i born in year t at gestational age a, Y_{iat} is a measure of infant health or of medical treatments, $W37_a$ is an indicator for prematurity (gestational age strictly below 37 completed weeks, or 259 days), and $f(\cdot)$ is a first-degree polynomial in normalized gestational age that is allowed to vary on both sides of the discontinuity. We normalize the running variable to zero at 258 gestational days because the treatments are applied to the left of the cutoff. With this definition of the running variable, the coefficient of interest β captures the change in low-risk newborns' outcomes and receipt of medical treatments as gestational age moves from 259 days (exactly 37 completed weeks) to 258 days.

Our baseline regressions use a rectangular kernel which places the same weight on all observations. This is equivalent to estimating OLS regressions within the chosen bandwidth (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). Since the running variable is discrete, we cluster the standard errors in all regressions at the gestational day level (Lee and Card, 2008).

Estimation in an RD framework is conducted within a small interval around the discontinuity. Larger bandwidths increase the degree of precision of the estimates, but also increase the risk of bias. We use a rule-of-thumb approach to select our bandwidth (Lee and Lemieux, 2010). For each health outcome and treatment measure, the optimal rule-of-thumb bandwidth is given by:

$$h_{ROT} = k \left[\frac{R\hat{\sigma}^2}{\sum_{i=1}^{n} (\hat{m}_i'')^2} \right]^{1/5},$$

where k is a parameter that depends on the kernel choice (2.702 for the rectangular kernel), R is the range of the running variable, n is the sample size, and $\hat{m}''(\cdot)$ and $\hat{\sigma}$ are the curvature and standard error of the regression of the health outcome on a fourth-degree polynomial in normalized gestational age, respectively. Appendix Table A1 lists the optimal bandwidths for our selected outcomes. Our baseline regressions use a bandwidth of 14 days to the left and right of gestational day 258.

4 Data

We use data from the Perinatal Registry of the Netherlands (Perinatale Registratie Nederland, Perined) for the years 2000–2008. Perined is an annual dataset covering approximately 99 percent of the primary care and 100 percent of the secondary care provided during pregnancy and delivery in the Netherlands (de Jonge et al., 2009). It is constructed by linking individual birth records submitted by midwifes (LVR-1), obstetricians/gynecologists (LVR-2) and paediatricians (LNR).[3]

The data include detailed information on the birth process. For each delivery, we observe the date and time of birth, type of birth attendant (midwife or OB/GYN), delivery location (home or hospital), method of delivery (vaginal, planned C-section, emergency C-section), use of interventions during vaginal delivery (labor augmentation, induction, use of forceps or vacuum), as well as the presence of complications during pregnancy or delivery. In the case of complications, we can observe the date and the reason for referral from midwife to an obstetrician. The data also provide rich background information on newborns (gender, gestational age in days, birth weight, parity, plurality) and basic demographic characteristics of mothers (age, ethnicity, 4-digit residential postal code). We complement the individual-level Perined data with a secondary postal code-level data set from Statistics Netherlands (Kerncijfers postcodegebieden 2004). These data provide a snapshot of average characteristics in the postal code of residence of the mother as of January 1, 2004, such as average monthly household income, average area density, and the share of residents 0-15 years old.[4]

[3]Perined data does not include information on births supervised by general practitioners, a very small share of all primary care deliveries (Amelink-Verburg and Buitendijk, 2010).

[4]Average area density is the average number of addresses per square kilometer in a circle with a radius of 1 km around each address in the postal code.

Our outcomes include a number of variables pertaining to medical treatments administered during or soon after birth as well as measures of short-term infant health. We start by examining the effect of the week-37 rule on medical interventions during and after delivery: obstetrician supervision of birth, delivery in a hospital, use of forceps or vacuum, and admission to a neonatal intensive care unit (NICU) within the first 7 days of life. We then examine effects on newborn short-term health outcomes as measured by 7-day mortality, 28-day mortality, and low Apgar score.[5]

A variable crucial to our identification strategy is gestational age. The week-37 rule states that women should be referred to secondary care if the onset of labor occurs before 37 completed gestational weeks. In our data, we do not observe the date and time of the onset of labor. Hence, we define the cutoff based on gestational age at birth, measured as the number of days between the date of the last menstrual period and the date of birth.[6]

Some of our robustness checks include additional covariates, which can be classified into four groups. The first group (time effects) includes fixed effects for the year, month and day of the week of the birth. The second group (maternal characteristics) includes mother's age and ethnicity.[7] The third group (infant characteristics) includes birth weight and indicators for gender, congenital anomalies and birth position.[8] The final group (postal code characteristics) includes the average characteristics of the postal code

[5]We do not have information on longer term mortality rates. Apgar is measured 5 minutes after birth and summarizes the health of newborns based on five criteria: appearance (skin color), pulse (heart rate), grimace response ("reflex irritability"), activity (muscle tone), and respiration (breathing rate and effort). The score ranges from 0 to 10 with higher scores indicating better health. Low Apgar score refers to an Apgar score below 7.

[6]Alternatively, we can define gestational age at the onset of labor as gestational age at birth shifted by an "average duration of labor" in hours, because we can observe the exact time of birth. Analyses using these alternative definitions (available upon request) yield results almost identical to our baseline results.

[7]We include indicators for six maternal age categories (less than 20, 20–24, 25–29, 30–34, 35–39, 40 and above) and three maternal ethnicity categories: Dutch, Mediterranean and others (Moroccans and Turks, commonly identified as "Mediterraneans," represent the majority of the immigrant population in the Netherlands).

[8]Specifically, we include birth weight in grams and indicators for very low birth weight (less than 1,500 grams), low birth weight (between 1,500 and 2,500 grams), gender, congenital anomalies (mild and severe) and birth position (breech birth and other).

of residence of the mother: monthly household income, area density and the fraction of residents 0–15 years old.[9]

Our analysis sample includes live deliveries by low-risk women with gestational age between 245 and 272 days, a 14-day interval on each side of day 259 (exactly 37 completed gestational weeks). We focus on low-risk women because the week-37 rule does not apply to high-risk women.[10] This has the added benefit that women in this category are homogenous in terms of their prenatal care. As a result, we are able to identify the effects of early-life medical interventions abstracting from the effects of prenatal care.

Low-risk women are defined as those under the care of a midwife at the onset of labor, that is when contractions start spontaneously or when membranes rupture spontaneously (Evers et al., 2010; van der Kooy et al., 2011). Referrals under the week-37 rule can be made because of premature onset of labor, but also because of the "threat of prematurity," which midwives can potentially assess before the actual onset of labor (e.g., due to cramping, increased pressure in pelvis or vagina, or vaginal bleeding). If midwives have a tendency to refer women in poorer health in these cases, then referral patterns on each side of the threshold may be different and a comparison of births right above and right below the threshold would be misleading. In order to eliminate this potential bias, we define low-risk women as women who were not referred to an obstetrician by gestational age of 238 days,[11] 7 days before the lower bound of our target interval.[12] We also restrict our sample to first

[9]Some of the control variables (newborn gender, birth weight, mother's age, and postal code characteristics) are missing for a very small number of observations (less than 0.03 percent for individual characteristics and less than 0.8 percent for postal code characteristics). We replace these missing values with sample averages and we include indicators for missing values for each variable as additional controls.

[10]The week-37 rule affects three important medical inputs: the medical professional supervising the birth, the location of delivery, and the medical treatments during and soon after birth. Among high-risk women, there is no change in the first two inputs across the prematurity cutoff and only a limited change in the third input.

[11]The date of referral is missing for about 4 percent of our analysis sample. We exclude these observations from the main analyses and we check the sensitivity of our results to their inclusion in section 5.4.

[12]We thank Gordon Dahl for this suggestion. Results from a sample including only women under the supervision of a midwife at the onset of labor (available upon request) are quali-

births because future fertility may be endogenous to experiences in previous deliveries (which we do not observe in the data). In addition, midwives and women may use information from previous pregnancies to determine if and potentially when referral to an obstetrician should be made. Finally, we exclude multiple births, which are automatically referred to obstetricians, and cases in which gestational age may be manipulated (planned C-sections, induced and stimulated births). This results in an analysis sample of 85,797 women who are under the care of a midwife until at least gestational day 238, and who give birth to their first child between gestational days 245 and 272. When investigating the heterogeneity in returns to medical interventions, we divide the sample into quartiles of the average monthly household income in the postal code of residence of the mother.[13]

5 Results

5.1 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 there are no medical tests which can accurately predict prematurity and our analysis sample consists of spontaneous births, the variation in receipt of medical treatments near the week-37 cutoff should be as good as random. However, the key identification assumption of the RD design could be violated if women (or midwives) strategically misreport gestational age at birth.

In order to test this, we examine in Figure 1 the frequency of births by gestational age within a 4-week interval around the cutoff. A discontinuity in the density of births around the week-37 cutoff would suggest manipulation of the running variable and thus invalidate our RD design (McCrary, 2008). Not surprisingly, the number of births is increasing in gestational age, with the vast majority of births occurring after 39–40 completed gestational weeks.

tatively similar.

[13] Income quartiles are defined using the entire population of births.

However, visually, there is no significant jump in the number of births between day 258, when the week-37 rule applies, and day 259, when it does not. More formally, we estimate a local-linear regression similar to equation (1), using the logarithm of the number of births at each gestational age as the dependent variable. We indeed do not find evidence of a statistically significant discontinuity in the number of births.[14]

Next, we check whether there are differences in observable characteristics across the week-37 cutoff. If the RD design is valid, then the observable characteristics should be locally balanced on both sides of the week-37 cutoff. Figure 2 presents the means of selected covariates by gestational age in a 4-week interval before and after the cutoff.[15] The Figure shows that the distribution of the covariates is smooth around the discontinuity. In order to examine this issue more formally, we also provide in Table 1 the means of covariates on either side of the discontinuity within our 14-day bandwidth after controlling for gestational age. The last column of the table provides the p-values for the test of equality of the means, clustered at the gestational day level.[16] The results reported in Table 1 confirm the visual evidence in Figure 2: observations just below the week-37 cutoff are similar to those just above the week-37 cutoff in terms of the majority of maternal characteristics, newborn characteristics (low birth weight, congenital anomalies, breech birth), and average characteristics in the postal code of the residence of the mother (density, share of 0-15 year olds, distance to the nearest hospital). It is worth noting that even in the few cases where we find statistically significant differences, the difference in

[14]The estimated discontinuity at the cutoff is -0.033 (s.e. 0.023). Appendix Figure A1 plots the frequency of births by gestational age for each income quartile. The corresponding results for the regression-based McCrary test for quartiles 1–4 are: -0.093 (s.e. 0.039), -0.054 (s.e. 0.042), 0.024 (s.e. 0.038), -0.01 (s.e. 0.044).

[15]For visual clarity, here and in the rest of the paper, we group the data in 4-day bins starting from the cutoff. Appendix Figures A2–A5 plot the corresponding distributions by income quartile.

[16]This analysis is equivalent to estimating a local-linear regression similar to equation (1) using the covariates as the dependent variable, with the difference in means below and above the cutoff (i.e., columns 1 and 2) representing the coefficient estimate for $W37_a$ and the corresponding p-value indicated in column 3. Appendix Table A2 provides a similar analysis by income quartile.

the magnitudes is very small with no clear pattern. For example, infants born before day 259 are on average only 32 grams lighter than those born after the cutoff.[17] On the other hand, mothers of preterm infants reside in postal codes where the average monthly household income is higher by €12.[18]

Overall, we find no evidence of manipulation of the running variable around the week-37 cutoff. In addition, we find no systematic evidence of discontinuities in the observable characteristics of newborns and their mothers. This lends support to the claim that the variation in medical treatments near the week-37 cutoff is as good as random.

5.2 The Discontinuity in Medical Interventions

If the Dutch institutional rule governing the supervision of premature births is binding, then we should observe a discontinuity in receipt of medical treatments at 37 completed gestational weeks. To examine this, in Figure 3 we plot several medical treatments for gestational ages within a 4-week interval around the cutoff. Visually, there is a substantial jump up at the week-37 cutoff for each of these medical treatments, with newborns below the cutoff having higher rates of obstetrician supervision, hospital births, NICU admissions and medical interventions during delivery (use of forceps and vacuum).[19]

[17]The small statistically significant jump in birth weight is not surprising because birth weight and gestational age are particularly related to each other. Almond et al. (2010) exploit the variation in medical inputs across the very low birth weight threshold to estimate the marginal returns to medical care and also find a statistically significant jump in gestational age at the very low birth weight cutoff.

[18]One potential explanation for the observed significant jumps is that our linear polynomial in gestational age may not fully explain the trends in some covariates by gestational age. We therefore additionally run regressions with a second-degree polynomial in gestational age, as well as regressions in which we exclude the observations very close to the cutoff (donut regressions), regressions with alternative bandwidths, and regressions with additional covariates. Our results prove robust across all specifications (see section 5.4).

[19]There are two reasons why the probability of obstetrician supervision does not "jump" from 1 to 0 when gestational age increases from just under to just over 37 weeks. First, the week-37 rule is not perfectly enforced, meaning that not all the infants born before 37 completed gestational weeks are referred to an OB/GYN. Second, low-risk women can be referred to an OB/GYN for reasons other than prematurity, including complications arising during delivery, slow progression, or the need for pain relief medication. As a result, some of the births with at least 37 completed gestational weeks are at least partially supervised

In the first column of Table 2, we examine whether the difference in receipt of medical treatments below and above the week-37 cutoff is statistically significant by estimating equation (1) within our baseline bandwidth of 14 days around the cutoff. Each cell reports the coefficient of $W37$ from a different regression. The results suggest that preterm newborns receive significantly more medical treatments: they are, on average, 40 percentage points more likely to be supervised by an obstetrician; 26 percentage points more likely to be delivered in a hospital; 12.5 percentage points more likely to be admitted to a NICU within the first seven days of life; and 4 percentage points more likely to be delivered by use of forceps and vacuum. These estimates are economically large and represent increases of 30–110% when compared to the mean of the outcomes above the cutoff.

Appendix Figures A6–A9 and columns 2–5 of Table 2 show that the week-37 rule leads to discontinuities in receipt of medical treatments across the income distribution. The estimated discontinuities are all statistically significant and point to an income gradient whereby higher-income mothers experience somewhat larger relative increases in treatments at the cutoff. For example, preterm newborns in the lowest income quartile are, on average, 11.36 percentage points more likely to be admitted to a NICU, which represents a 110% increase at the mean above the cutoff. Preterm newborns in the highest income quartile, on the other hand, are 12.91 percentage points more likely to be admitted to a NICU, a 132% increase at the mean above the cutoff. Similarly, the week-37 rule increases the probability of a hospital birth by 29% for preterm newborns in the lowest quartile and by 43% for preterm newborns in the highest income quartile. Overall, the evidence suggests that the Dutch institutional setup provides significant variation in receipt of medical treatments among low-risk first-time mothers, regardless of income.

by OB/GYNs.

5.3 Short-Term Newborn Health

In this section we present our estimates of the effects of early-life medical interventions on newborn health outcomes. Figure 4 plots the evolution of our three measures of newborn health as a function of gestational age within a 4-week window around the cutoff. The Figure indicates a smooth evolution of all health measures across the week-37 cutoff, suggesting no significant health differences between births slightly below and slightly above 258 completed gestational days.

The first column of Table 3 presents the regression estimates corresponding to the visual evidence from Figure 4. The results, although imprecise, confirm that there are no significant health differences between preterm newborns and those born after 37 completed gestational weeks. Since these coefficients represent an intention-to-treat effect of the week-37 rule, our estimates suggest that this rule yields no significant health benefits for the average low-risk newborn with gestational age close to 37 weeks.

The coefficient estimates in columns 2–5 of Table 3, however, suggest that the average effects mask substantial heterogeneity in the returns to medical treatments.[20] In particular, we find that preterm newborns in the lowest income quartile are significantly less likely to die and to have low Apgar scores when compared to low-income newborns who are slightly above the week-37 cutoff. When thinking about the magnitudes of the effects, it is worth emphasizing that these estimates have relatively wide confidence intervals that include much smaller but still economically important returns. For example, the lower bounds of a 95-percent confidence interval indicate 0.63 fewer infant deaths per 1,000 births for 7-day mortality and a 0.07 percentage point reduction in the probability of low Apgar score. In stark contrast to these findings, our results in columns 3–5 indicate no significant health differences between higher-income preterm newborns relative to higher-income at-term newborns. In all cases, we reject the equality of the estimates between the lowest-income quartile and the other income quartiles.

[20] Appendix Figures A10–A13 provide the corresponding visual evidence.

5.4 Robustness Checks

We next investigate the robustness of our results to several scenarios that could lead to biased estimates. If the key assumption in our RD design is satisfied (i.e., the variation in receipt of medical treatments is as good as random around the week-37 cutoff), then including additional covariates in our model should not change our conclusions. In panel A of Table 4 we present estimates from a specification that includes the full set of controls described in section 4. We again find statistically significant health benefits for preterm babies in the lowest income quartile, and no significant health differences between preterm and at-term newborns in the other three quartiles. The magnitudes of the estimated effects are very similar to the baseline results.

Next, we turn to the possibility that our results could be driven by heaping at the cutoff. In order to address this issue, Barreca et al. (2016) suggest estimating "donut" regressions that exclude the observations at the cutoff. Panel B of Table 4 shows results estimated on a sample excluding newborns with gestational ages of 258 and 259 days. This strategy does not alter the main conclusions from our results: the week-37 rule leads to health gains among the poorest quartile, but not among the other income quartiles.

In Panels C–E, we examine the sensitivity of our estimates to model specification. Panel C focuses on the choice of kernel and reports results based on a triangular kernel which places less weight on observations farther away from the cutoff. Our results again point to health gains for babies slightly below the week-37 cutoff in the lowest income quartile and generally no benefits for preterm babies in the higher income quartiles. In Panel D, we test the robustness of our results to different bandwidths using intervals of 7 and 21 days on either side of the cutoff. The estimated effects are again very similar to those obtained in the baseline model. Panel E tests the sensitivity of our results to the degree of the polynomial in gestational age. Our choice of a linear function in gestational age is motivated by the reduced form relationship between infant outcomes and gestational age plotted in Figure 4, which does not indicate nonlinearities within our bandwidth. When we reestimate our baseline regressions using a second degree polynomial (as before, allowed to vary on either side of the cutoff), the coefficients again indicate positive health returns for preterm babies in the lowest income quartile, although we lose some precision due to overfitting the data.

Panel F of Table 4 examines the sensitivity of our results to alternative sample selection criteria. Recall that our analysis sample includes women under the supervision of a midwife at least until gestational day 238. Since date of referral is missing for a small number of observations, we exclude these from the main analyses. In the first part of the Panel, we expand our analysis sample to include these women and show that the results are very similar to our baseline findings. In the remaining parts of the Panel, we change the sample to include women under the supervision of a midwife at least until gestational day 245 and 231, respectively. In both cases, we confirm our baseline results: preterm newborns from the lowest income families gain substantially from the medical treatments induced by the week-37 rule, but those in higher income quartiles do not seem to have any significant benefit from these additional treatments.

5.5 Potential Mechanisms

Our results consistently show that the week-37 rule generates health benefits even among an observably low-risk population, but only among newborns in the lowest income quartile. There are several mechanisms that may explain this heterogeneity in the returns to early-life medical interventions. To begin with, the current risk selection system may be better suited to screen for risks among higher income women. For example, communications about health-related issues between midwives and low-income (and thus lower educated) pregnant women may be more difficult. This may make the risk assessment among these women less precise. As a result, some high-risk low-income women may be incorrectly classified as low-risk. These low-income women and their infants may experience higher benefits from the additional medical treatments provided due to the week-37 rule.

In addition, low-income mothers may engage in unhealthy behaviors that are not included as reasons for referral to an obstetrician in the LOI. In this case, the pool of low-income mothers classified as low-risk may on average have worse (unobserved) underlying health than the higher-income mothers classified as low-risk. Medical interventions could then be more beneficial for infants in the low-income group.

Data limitations do not allow us to investigate the extent to which the current risk selection system may be less precise for low-income women. However, there is some anecdotal evidence suggesting that this channel may be important. A recent survey by the Royal Dutch Organisation of Midwives reports that midwives needed on average 23 percent extra time when caring for low-income women. The need for extra time was due to difficulties in collecting relevant (medical) data, additional education on prevention, lifestyles and risk, more frequent home visits, and consultations to exclude uncertainties. This led to a policy change in 2009 that increased reimbursements for midwives by 23 percent in selected very low-income postal codes (NZA, 2011).

In the remainder of this section, we investigate three other potential mechanisms for which we have reliable data. First, if lower income mothers reside closer to the hospital, their newborns may be exposed to additional treatments in a timelier manner and therefore profit more from such treatments. In the last row of Appendix Table A2, we check whether distance to the nearest hospital is larger among higher income mothers. We find no consistent pattern for such residential sorting by income. As such, this scenario is unlikely to drive our results.

Second, the observed heterogeneity in the impact of the week-37 rule may be driven by heterogeneity in the effect of location of birth, one of the medical inputs that change as a result of the week-37 rule. We find that infants slightly below the prematurity cutoff have a higher likelihood of a hospital birth (instead of a home birth). Using data from the Netherlands, Daysal et al. (2015) find that giving birth in a hospital leads to lower infant mortality among low-risk deliveries and that this effect is entirely driven by lower income mothers. If we assume that the causal impact of a hospital birth is the same in our sample, a 20 percentage point increase in the fraction of hospital births would be associated with a reduction of 2.5 deaths in 7-day infant mortality in the lowest-income quartile.[21] This suggests that the change in location of delivery can explain only part of the heterogeneity in returns to medical treatments.

Third, prematurity and referral to an obstetrician may increase unmet maternal expectations and distress during delivery differentially across the income distribution. Among low-risk Dutch women, middle and higher income mothers are more likely than lower income mothers to prefer midwife-supervised home births (de Jonge et al., 2009). The unexpected change in the medical professional supervising the birth and in location of delivery due to the week-37-rule may hence increase maternal distress in the middle and higher income groups. This may in turn counteract the potential gains from medical treatments: "[o]nce a woman is in labor, state-anxiety, due to loss of control and feelings of powerlessness, has been found to be positively related to 'abnormal' delivery, in particular prolonged labor" (Paarlberg et al., 2006). We examine effects on the presence of meconium in amniotic fluid, which is generally considered the most direct measure of fetal distress during labor and which can lead to severe complications if breathed in by the baby (meconium aspiration). Our results point to statistically significant increases in the likelihood of meconium staining among the preterm babies of higher-income mothers and no effects among lower-income mothers.[22] This suggests that increased maternal distress may also contribute to the observed differences in the returns to medical treatments by income.

[21] Daysal et al. (2015) exploit the exogenous variation between a mother's residence and the nearest obstetric ward to estimate the casual impact of a hospital birth. Their instrumental variables strategy indicates that giving birth in a hospital leads to 12.65 fewer infant deaths per 1,000 births among lower income mothers, with no significant effect on infant mortality in higher income areas.

[22] The coefficient estimates for $W37$ are -0.487 (s.e. 0.597) in the lowest income quartile, -0.089 (s.e. 0.427) in the second quartile, 1.444 (s.e. 0.511) in the third quartile and 1.508 (s.e. 0.748) in the highest income quartile. The results for the highest two income quartiles are statistically significant and represent increases of 22–25% when compared to the mean to the right of the cutoff.

6 Conclusions

In this paper, we examine the impact of early-life medical interventions on the health outcomes of low-risk newborns. In order to address the endogeneity in receipt of medical treatments, we exploit the exogenous variation generated by a policy rule in the Netherlands. The policy rule requires that low-risk women give birth under the supervision of a midwife unless the birth occurs before 37 completed gestational weeks, generating variation in the medical professional supervising the birth. Given that obstetricians only deliver in hospitals and that midwives cannot perform any medical interventions, the week-37 rule also induces variation in the location of delivery and in the medical treatments administered during and immediately after birth.

Using data from the Netherlands for the period 2000–2008, we find that the week-37 rule leads to statistically and economically significant increases in all of our measures of early-life medical treatments. Despite the substantial variation in medical inputs, our results indicate that average newborn health outcomes are similar across the week-37 cutoff. However, the average effects mask substantial heterogeneity in the returns to medical treatments along the income distribution. Our results indicate that preterm newborns in the lowest income quartile are significantly less likely to die and to have low Apgar scores when compared to low-income newborns who are slightly above the week-37 cutoff. The heterogeneity in the returns to early-life medical interventions may be caused by various channels. While we are not able to investigate some interesting pathways, we provide evidence suggesting that potential difficulties in risk screening, the change in delivery location, and maternal distress during delivery could be important factors.

One policy implication of our work is that ensuring access to medical treatments may improve newborn outcomes even among low-risk women living in a developed country. The Netherlands is a country where maternity care is provided using a rigorous process of risk selection based on both past medical history and the current health status and development of the mother and the fetus. The Dutch maternity system is explicitly geared toward midwife-supervised home births. Yet, even with a relatively sophisticated model of risk selection, we find that the babies of some women classified as low-risk benefit from the additional medical treatments provided by obstetricians in a hospital. These are women in the lowest income quartile, suggesting perhaps that risk selection is more difficult and less precise for women with lower socioeconomic status. Having a good understanding of risk selection must go hand in hand with crafting policies about childbirth technologies.

A second policy implication of our work is that programs aimed at reducing maternal distress due to changes in medical professionals or location of delivery may improve the health outcomes of some newborns. In our context, higher-income mothers exhibit signs of maternal distress as a result of the week-37 rule, suggesting that income is not protective in this regard.

References

Almond, Douglas, and Joseph Doyle (2011), "After midnight: A regression discontinuity design in length of postpartum hospital stays," *American Economic Journal: Economic Policy* 3(3), 1–34.

Almond, Douglas, Joseph Doyle, Amanda Kowalski, and Heidi Williams (2010), "Estimating marginal returns to medical care: Evidence from at-risk newborns," *Quarterly Journal of Economics* 125(2), 591–634.

Amelink-Verburg, Marianne, and Simone Buitendijk (2010), "Pregnancy and labour in the Dutch maternity care system: What is normal? The role division between midwives and obstetricians," *Journal of Midwifery & Women's Health* 55(3), 216–225.

Baicker, Katherine, and Amitabh Chandra (2004), "Medicare spending, the physician workforce, and beneficiaries' quality of care." *Health Affairs (Project Hope)* Suppl Web Exclusives, W4—-184–97.

Bais, Joke, and Maria Pel (2006), "The basis of the Dutch obstetric system: risk selection," *European Clinics in Obstetrics and Gynaecology* 2(4), 209–212.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell (2016), "Heaping-induced bias in regression-discontinuity designs," *Economic Inquiry* 54(1), 268–293.

Bharadwaj, Prashant, Katrine Løken, and Christopher Neilson (2013), "Early life health interventions and academic achievement," *American Economic Review* 103(5), 1862–1891.

Bitler, Marianne P, Jonah B Gelbach, and Hilary W Hoynes (2006), "What mean impacts miss: Distributional effects of welfare reform experiments," *American Economic Review* 96(4), 988–1012.

Breining, Sanni, N. Meltem Daysal, Marianne Simonsen, and Mircea Trandafir (2015), "Spillover effects of early-life medical interventions," IZA Discussion Paper No. 9086.

Case, Anne, Darren Lubotsky, and Christina Paxson (2002), "Economic status and health in childhood: The origins of the gradient," *American Economic Review* 92(5), 1308–1334.

Chandra, Amitabh, and Jonathan Skinner (2012), "Technology growth and expenditure growth in health care," *Journal of Economic Literature* 50(3), 645–680.

Currie, Alison, Michael A. Shields, and Stephen Wheatley Price (2007), "The child health/family income gradient: Evidence from england," *Journal of Health Economics* 26(2), 213–232.

Cutler, David, Mark McClellan, Joseph Newhouse, and Dahlia Remler (1998), "Are medical prices declining? Evidence from heart attack treatments," *Quarterly Journal of Economics* 113(4), 991–1024.

Cutler, David, and Ellen Meara (1998), "The medical costs of the young and old: A forty-year perspective," in David Wise, ed., *Frontiers in the Economics of Aging*, 215–246, University of Chicago Press.

——— (2000), "The technology of birth: Is it worth it?" *NBER/Frontiers in Health Policy Research* 3(1), 33–67.

CVZ (2003), *Verloskundig Vademecum*, Diemen, The Netherlands.

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), 28–50.

de Jonge, A., B.Y. van der Goes, A.C.J. Ravelli, M.P. Amelink-Verburg, B.W. Mol, J.G. Nijhuis, J. Bennebroek Gravenhorst, and S.E. Buitendijk (2009), "Perinatal mortality and morbidity in a nationwide cohort of 529,688 low-risk planned home and hospital births," *BJOG: An International Journal Of Obstetrics And Gynaecology* 116(9), 1177–84.

Duggan, Mark G, and William N Evans (2008), "Estimating the impact of medical innovation: A case study of HIV antiretroviral treatments," *Forum for Health Economics & Policy* 11(2).

Evans, William N., and Craig Garthwaite (2012), "Estimating heterogeneity in the benefits of medical treatment intensity," *Review of Economics and Statistics* 94(3), 635–649.

Evers, Annemieke, Hens Brouwers, Chantal Hukkelhoven, Peter Nikkels, Janine Boon, Anneke van Egmond-Linden, Jacqueline Hillegersberg, Yvette Snuif, Sietske Sterken-Hooisma, Hein Bruinse, and Anneke Kwee (2010), “Perinatal mortality and severe morbidity in low and high risk term pregnancies in the Netherlands: Prospective cohort study,” *BMJ (Clinical research ed.)* 341, c5639.

Fuchs, Victor R (2004), “More variation in use of care, more flat-of-the-curve medicine.” *Health Affairs (Project Hope)* Suppl Variation, VAR104—-VAR107.

Garthwaite, Craig, and Mark Duggan (2012), “Empirical evidence on the value of pharmaceuticals,” in Patricia M. Danzon and Sean Nicholson, eds., *The Oxford Handbook of the Economics of the Biopharmaceutical Industry*, chap. 15, 463–492, Oxford University Press, 1 ed.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw (2001), “Identification and estimation of treatment effects with a regression-discontinuity design,” *Econometrica* 69(1), 201–209.

Imbens, Guido, and Thomas Lemieux (2008), “Regression discontinuity designs: A guide to practice,” *Journal of Econometrics* 142(2), 615–635.

Institute of Medicine (2011), *The Future of Nursing: Leading Change, Advancing Health*, Washington, D.C.: The National Academies Press.

Kramer, Michael, Aris Papageorghiou, Jennifer Culhane, Zulfiqar Bhutta, Robert Goldenberg, Michael Gravett, Jay Iams, Agustin Conde-Agudelo, Sarah Waller, Fernando Barros, Hannah Knight, and Jose Villar (2012), “Challenges in defining and classifying the preterm birth syndrome,” *American Journal of Obstetrics and Gynecology* 206(2), 108–112.

Kravitz, Richard L, Naihua Duan, and Joel Braslow (2004), “Evidence-Based medicine, heterogeneity of treatment effects, and the trouble with averages,” *The Milbank Quarterly* 82(4), 661–687, PMID: 15595946 PMCID: PMC2690188.

Lee, David, and David Card (2008), “Regression discontinuity inference with specification error,” *Journal of Econometrics* 142(2), 655–674.

Lee, David, and Thomas Lemieux (2010), “Regression discontinuity designs in economics,” *Journal of Economic Literature* 48(2), 281–355.

McClellan, Mark, and Joseph Newhouse (1997), “The marginal cost-effectiveness of medical technology: A panel instrumental-variables approach,” *Journal of Econometrics* 77(1), 39–64.

McCrary, Justin (2008), "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics* 142(2), 698–714.

Miller, Amalia (2006), "The impact of midwifery-promoting public policies on medical interventions and health outcomes," *B.E. Journal of Economic Analysis and Policy: Advances in Economic Analysis and Policy* 6(1), 1–34.

Newhouse, Joseph (1992), "Medical care costs: How much welfare loss?" *Journal of Economic Perspectives* 6(3), 3–21.

Norwitz, Errol R, and Aaron B Caughey (2011), "Progesterone supplementation and the prevention of preterm birth," *Reviews in Obstetrics and Gynecology* 4(2), 60–72.

NZA (2011), "Tariefbeschikking TB/CU-7007-01," Available online at http://www.nza.nl/regelgeving/tarieven/, accessed Sept. 4, 2012.

Paarlberg, K. Marieke, Ad J.J.M. Vingerhoets, and Herman P. van Geijn (2006), "Maternal stress and labor," in Asim Kurjak and Frank A. Chervenak, eds., *Textbook of Perinatal Medicine*, CRC Press, second ed.

Schakel, W, and J Bekhof (2010), "Prematuren geboren na 36 weken zwangerschapsduur - 48 uur observatie op de kraamafdeling is voldoende [prematures born after 36 weeks of gestation − 48 hours of observation in the obstetric ward is sufficient]," *Tijdschrift voor kindergeneeskunde* 78, 3–6.

Skinner, Jonathan (2011), "Causes and consequences of regional variations in health care," in Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, eds., *Handbook of Health Economics*, vol. 2 of *Handbook of Health Economics*, chap. 2, 45–93, Elsevier.

Skinner, Jonathan, Douglas Staiger, and Elliott Fisher (2006), "Is technological change in medicine always worth it? The case of acute myocardial infarction," *Health Affairs* 25(2), w34–w47.

Stukel, Therese A, F Lee Lucas, and David E Wennberg (2005), "Long-term outcomes of regional variations in intensity of invasive vs medical management of Medicare Patients with acute myocardial infarction." *{JAMA:} The Journal Of The American Medical Association* 293(11), 1329–1337.

van der Kooy, Jacoba, Jashvant Poeran, Johanna de Graaf, Erwin Birnie, Semiha Denktaş, Eric Steegers, and Gouke Bonsel (2011), "Planned home compared with planned hospital births in the Netherlands," *Obstetrics & Gynecology* 118, 1037–1046.

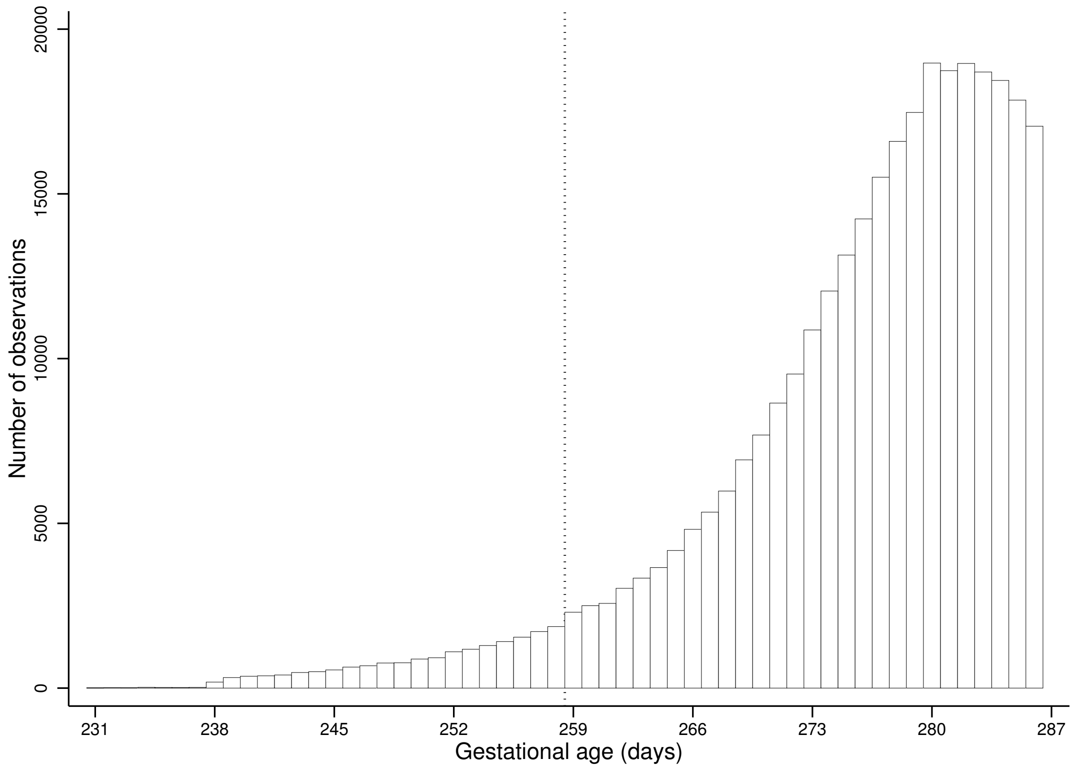
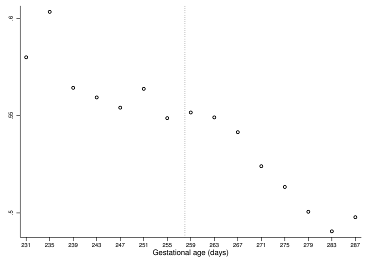
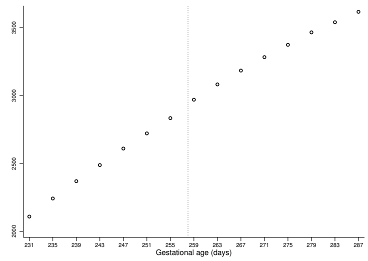


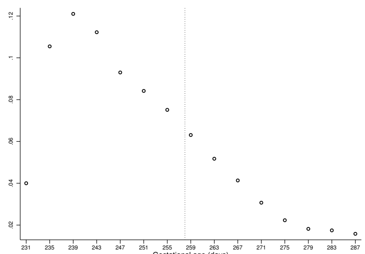
Figure 1: Frequency of births around the week-37 cutoff



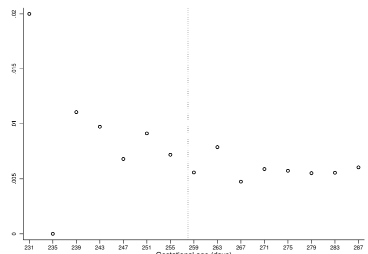
(a) Gender: male



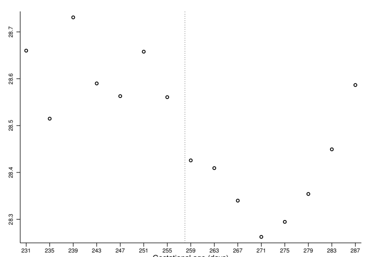
(b) Birth weight



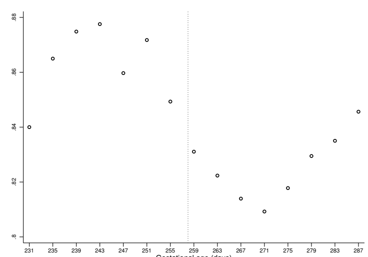
(c) Breech birth



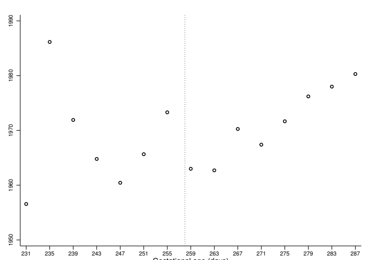
(d) Mild congenital anomaly



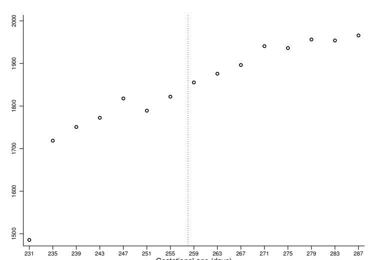
(e) Mother's age



(f) Mother's ethnicity: Dutch

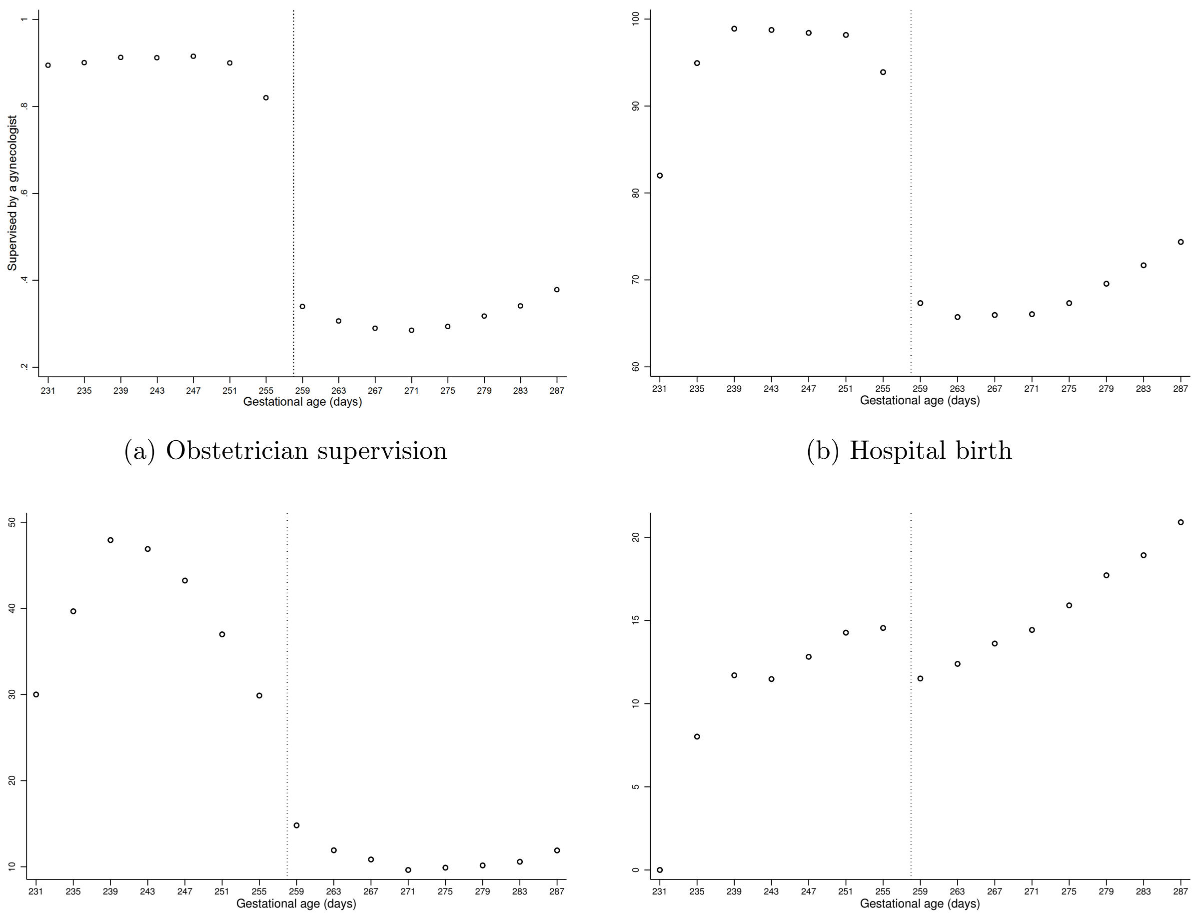


(g) Average household income



(h) Average density

Figure 2: Distribution of selected covariates around the week-37 cutoff



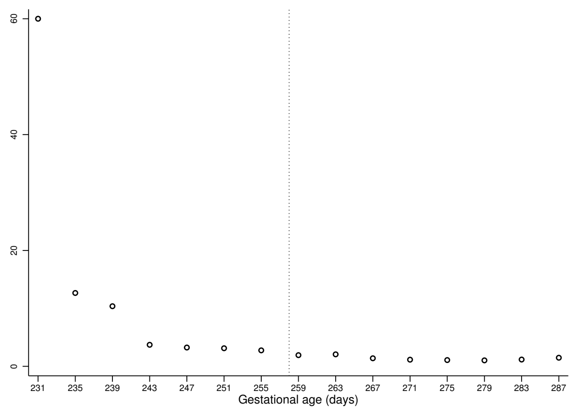
(a) Obstetrician supervision

(b) Hospital birth

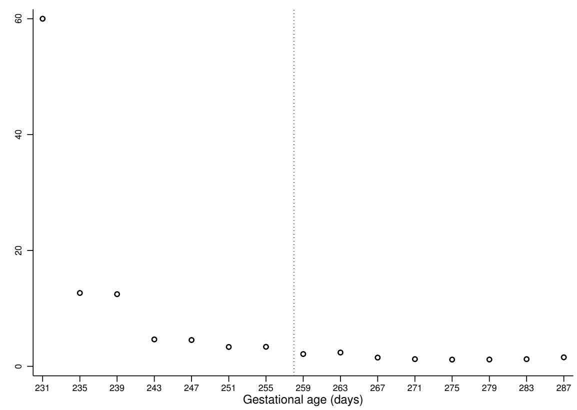
(c) NICU admission

(d) Use of forceps/vacuum

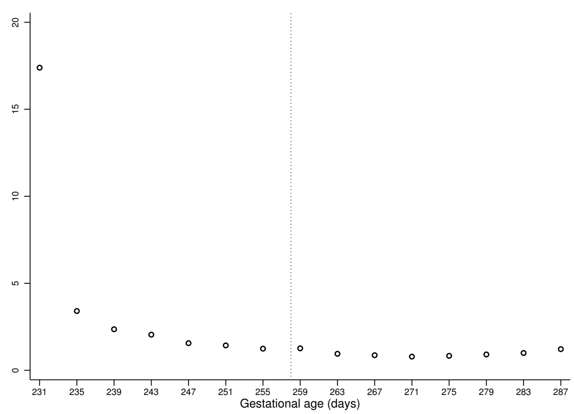
Figure 3: Medical treatments around the week-37 cutoff



(a) 7-day mortality



(b) 28-day mortality



(c) Low Apgar score

Figure 4: Newborn health around the week-37 cutoff

Table 1: Comparison of selected characteristics around the discontinuity

	Gestational age		Clustered
	Less than 37 completed weeks (1)	More than 37 completed weeks (2)	p-value for differences (3)
A. Maternal characteristics			
Age	28.610	28.474	0.033
20–24	0.141	0.156	0.000
25–29	0.384	0.381	0.554
30–34	0.349	0.334	0.010
35–39	0.084	0.087	0.478
40 and above	0.007	0.007	0.780
Ethnicity			
Dutch	0.849	0.838	0.032
Mediterranean	0.055	0.054	0.901
B. Newborn characteristics			
Male	0.548	0.559	0.112
Birth weight	2,871	2,903	0.000
Low birth weight ($leq 2,500$g)	0.124	0.118	0.501
Congenital anomaly			
Mild	0.008	0.007	0.547
Severe	0.013	0.012	0.503
Breech birth	0.071	0.069	0.678
C. Residential characteristics			
Average household income	1,973	1,961	0.000
Average density	1,821	1,835	0.543
Percent 0–15 year-old	18.843	18.858	0.833
Distance to nearest hospital (km)	4.911	4.849	0.264
Number of observations	15,302	70,495	

Notes: Each cell represents the mean of the corresponding variable in the row after controlling for gestational age. The last column presents the p-value for differences in means clustered at the gestational day level.

Table 2: Medical treatments around the week-37 cutoff

	All	Quartile of average household income in postal code			
		First	Second	Third	Fourth
	(1)	(2)	(3)	(4)	(5)
Obstetrician supervision	0.403***	0.379***	0.398***	0.428***	0.408***
Mean outcome	0.451	0.458	0.449	0.446	0.449
Observations	85,797	21,384	22,206	21,265	20,942
Hospital birth	0.261***	0.208***	0.268***	0.293***	0.280***
	(0.018)	(0.014)	(0.023)	(0.027)	(0.019)
Mean outcome	0.660	0.714	0.647	0.628	0.649
Observations	85,797	21,384	22,206	21,265	20,942
NICU admission	0.125***	0.114***	0.123***	0.132***	0.129***
	(0.009)	(0.018)	(0.017)	(0.009)	(0.015)
Mean outcome	0.114	0.112	0.125	0.119	0.098
Observations	85,797	21,384	22,206	21,265	20,942
Use of forceps or vacuum	0.040***	0.031**	0.033***	0.042**	0.051***
	(0.009)	(0.013)	(0.009)	(0.018)	(0.015)
Mean outcome	0.131	0.122	0.127	0.134	0.142
Observations	85,682	21,355	22,178	21,232	20,917

Notes: Each cell reports the estimated discontinuity at the week-37 cutoff from a different regression. The dependent variable is listed in the row heading and the sample in the column heading. All treatment variables represent the fraction of births receiving that treatment. All specifications include a first-degree polynomial in normalized gestational age, allowed to vary on each side of the cutoff, and are estimated by OLS. Samples restricted to observations with gestational age within a 14-day bandwidth around day 258. Mean outcome refers to observations to the right of the cutoff. Robust standard errors clustered at the gestational day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Newborn health around the week-37 cutoff

	All	Quartile of average household income in postal code			
		First	Second	Third	Fourth
	(1)	(2)	(3)	(4)	(5)
7-day mortality	0.373	-3.428^{**}	3.069	0.737	0.920
	(0.480)	(1.424)	(2.270)	(1.160)	(1.320)
Mean outcome	1.518	1.696	1.315	1.328	1.740
Observations	85,797	21,384	22,206	21,265	20,942
28-day mortality	0.508	-3.548^{*}	3.155	0.934	1.300
	(0.750)	(1.898)	(2.712)	(1.116)	(0.914)
Mean outcome	1.702	1.866	1.425	1.501	2.030
Observations	85,797	21,384	22,206	21,265	20,942
Low apgar score	-0.149	-0.684^{**}	0.061	0.352	-0.331
	(0.136)	(0.311)	(0.350)	(0.305)	(0.318)
Mean outcome	0.919	1.161	0.824	0.827	0.866
Observations	85,652	21,348	22,164	21,229	20,911

Notes: Each cell reports the estimated discontinuity at the week-37 cutoff from a different regression. The dependent variable is listed in the row heading and the sample in the column heading. 7-day and 28-day mortality are measured per 1,000 births, while Low apgar score is the percentage of births with a score below 7. All specifications include a first-degree polynomial in normalized gestational age, allowed to vary on each side of the cutoff, and are estimated by OLS. Samples restricted to observations with gestational age within a 14-day bandwidth around day 258. Mean outcome refers to observations to the right of the cutoff. Robust standard errors clustered at the gestational day level. $^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$

Table 4: Robustness checks, by quartile of average household income

	First quartile (1)	Second quartile (2)	Third quartile (3)	Fourth quartile (4)
A. Including controls				
7-day mortality	-3.511^{**}	2.950	0.640	0.423
	(1.465)	(2.107)	(1.196)	(1.322)
Mean outcome	1.696	1.315	1.328	1.740
28-day mortality	-3.616^{*}	3.067	0.813	0.774
	(1.964)	(2.560)	(1.140)	(0.883)
Mean outcome	1.866	1.425	1.501	2.030
Low apgar score	-0.697^{**}	0.065	0.322	-0.409
	(0.312)	(0.368)	(0.307)	(0.336)
Mean outcome	1.161	0.824	0.827	0.866
Observations	21,384	22,206	21,265	20,942
B. Donut regressions				
7-day mortality	-4.188^{*}	4.911^{**}	-0.729	0.420
	(2.106)	(2.157)	(1.205)	(1.907)
Mean outcome	1.696	1.245	1.373	1.680
28-day mortality	-5.384^{**}	5.085^{*}	-0.457	1.200
	(2.477)	(2.878)	(1.145)	(1.177)
Mean outcome	1.871	1.358	1.552	1.980
Low apgar score	-1.019^{**}	0.416	0.367	-0.126
	(0.370)	(0.413)	(0.414)	(0.405)
Mean outcome	1.159	0.805	0.831	0.817
Observations	20,358	21,130	20,246	19,895
C. Triangular kernel				
7-day mortality	-2.971^{***}	3.613	0.866	0.430
	(0.758)	(2.592)	(1.240)	(1.339)
Mean outcome	1.696	1.315	1.328	1.740
28-day mortality	-2.598^{**}	4.136	0.858	0.863
	(1.254)	(3.138)	(1.214)	(0.684)
Mean outcome	1.866	1.425	1.501	2.030
Low apgar score	-0.572^{**}	0.103	0.200	-0.748^{**}
	(0.254)	(0.364)	(0.285)	(0.322)
Mean outcome	1.161	0.824	0.827	0.866
Observations	21,384	22,206	21,265	20,942

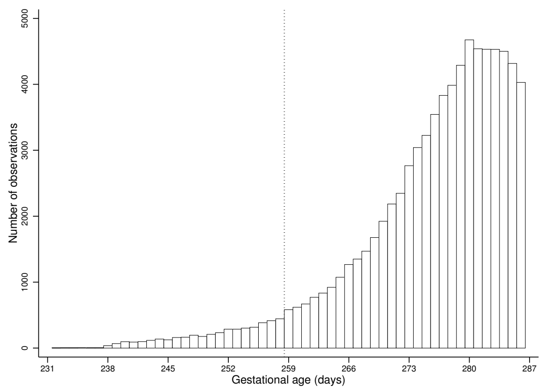
Table 4: Robustness checks, by quartile of average household income (cont'd)

	First quartile (1)	Second quartile (2)	Third quartile (3)	Fourth quartile (4)
D. Choice of bandwidth				
7-day bandwidth				
7-day mortality	−3.190***	5.315	1.796	0.215
	(0.733)	(3.458)	(1.341)	(1.971)
Mean outcome	2.194	2.149	1.319	2.688
28-day mortality	−2.116*	6.254	1.493	1.287
	(1.172)	(4.087)	(1.455)	(0.876)
Mean outcome	2.377	2.328	1.507	3.264
Low apgar score	−0.485	0.074	0.185	−1.237**
	(0.316)	(0.451)	(0.329)	(0.433)
Mean outcome	1.317	0.969	0.944	1.230
Observations	7,905	8,177	7,922	7,673
21-day bandwidth				
7-day mortality	−2.863***	2.888	−0.718	−0.272
	(0.918)	(1.938)	(1.487)	(1.252)
Mean outcome	1.369	1.096	1.186	1.356
28-day mortality	−3.420**	3.143	−0.618	0.443
	(1.566)	(2.308)	(1.417)	(0.983)
Mean outcome	1.534	1.165	1.281	1.522
Low apgar score	−0.557*	0.412	0.161	0.085
	(0.290)	(0.326)	(0.298)	(0.261)
Mean outcome	1.036	0.803	0.855	0.819
Observations	46,711	48,388	46,770	46,394
E. Choice of polynomial degree: Second degree polynomial				
7-day mortality	−2.225*	4.487	0.983	−0.275
	(1.141)	(3.395)	(1.693)	(1.848)
Mean outcome	1.696	1.315	1.328	1.740
28-day mortality	−1.079	5.657	0.616	0.162
	(1.514)	(4.207)	(1.779)	(0.947)
Mean outcome	1.866	1.425	1.501	2.030
Low apgar score	−0.417	0.152	−0.042	−1.435***
	(0.346)	(0.557)	(0.340)	(0.386)
Mean outcome	1.161	0.824	0.827	0.866
Observations	21,384	22,206	21,265	20,942

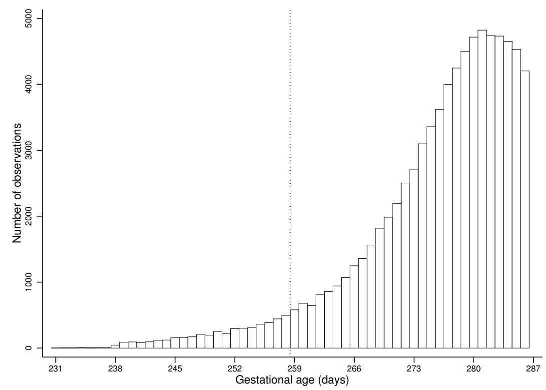
Table 4: Robustness checks, by quartile of average household income (cont'd)

	First quartile (1)	Second quartile (2)	Third quartile (3)	Fourth quartile (4)
F. Definition of analysis sample				
Including births with unknown referral date				
7-day mortality	−3.538**	2.986	0.705	1.161
	(1.360)	(2.239)	(1.126)	(1.314)
Mean outcome	1.673	1.425	1.391	1.659
28-day mortality	−3.666*	2.735	0.650	1.511
	(1.811)	(2.675)	(1.186)	(0.904)
Mean outcome	1.835	1.584	1.670	1.935
Low apgar score	−0.813**	0.047	0.390	−0.428
	(0.330)	(0.348)	(0.326)	(0.345)
Mean outcome	1.168	0.857	0.825	0.880
Observations	22,459	23,123	22,116	22,039
Under midwife supervision at least until gestational day 245				
7-day mortality	−2.936***	3.211	0.706	1.136
	(1.010)	(2.309)	(1.188)	(1.359)
Mean outcome	1.717	1.273	1.341	1.697
28-day mortality	−3.150**	3.733	0.647	1.482
	(1.499)	(2.792)	(1.167)	(0.887)
Mean outcome	1.889	1.384	1.516	1.989
Low apgar score	−0.648**	0.296	0.374	−0.366
	(0.265)	(0.369)	(0.307)	(0.315)
Mean outcome	1.169	0.821	0.812	0.867
Observations	20,959	21,815	20,868	20,590
Under midwife supervision at least until gestational day 231				
7-day mortality	−3.366**	3.041	1.161	0.810
	(1.395)	(2.239)	(1.194)	(1.312)
Mean outcome	1.685	1.307	1.319	1.787
28-day mortality	−3.480*	3.130	1.357	1.189
	(1.857)	(2.675)	(1.154)	(0.915)
Mean outcome	1.854	1.416	1.491	2.076
Low apgar score	−0.759**	0.033	0.394	−0.262
	(0.341)	(0.370)	(0.332)	(0.318)
Mean outcome	1.165	0.835	0.822	0.866
Observations	21,587	22,418	21,451	21,114

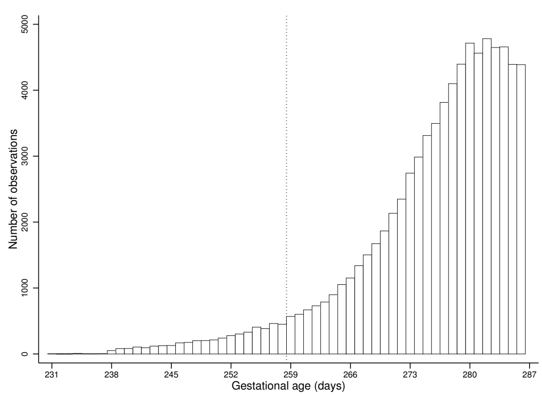
Notes: Each cell reports the estimated discontinuity at the week-37 cutoff from a different regression. The dependent variable is listed in the row heading and the sample in the column heading. 7-day and 28-day mortality are measured per 1,000 births, while Low apgar score is the percentage of births with a score below 7. All specifications include a first-degree polynomial in normalized gestational age, allowed to vary on each side of the cutoff, and are estimated by OLS, unless otherwise mentioned. Samples restricted to observations with gestational age within a 14-day bandwidth around day 258, unless otherwise mentioned. Mean outcome refers to observations to the right of the cutoff. Robust standard errors clustered at the gestational day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$



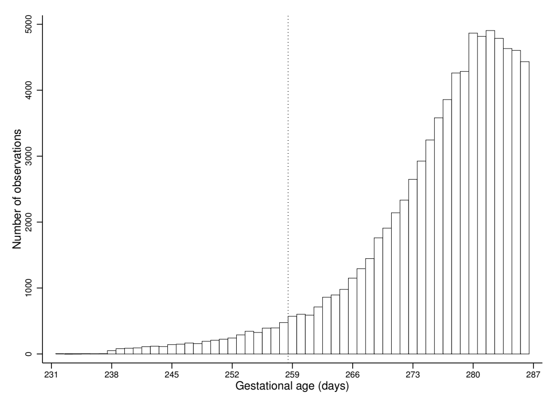
(a) First quartile



(b) Second quartile

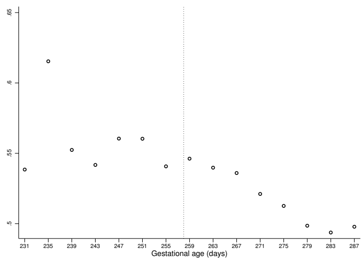


(c) Third quartile

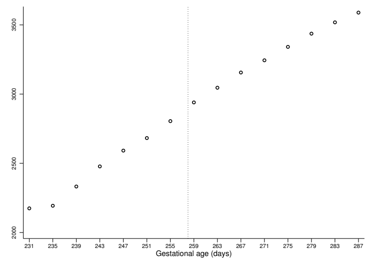


(d) Fourth quartile

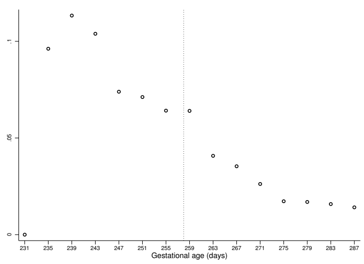
Figure A1: Frequency of births around the week-37 cutoff, by income quartile



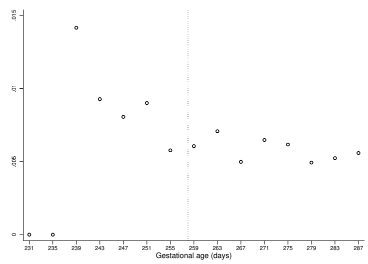
(a) Gender: male



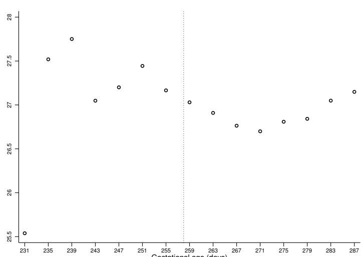
(b) Birth weight



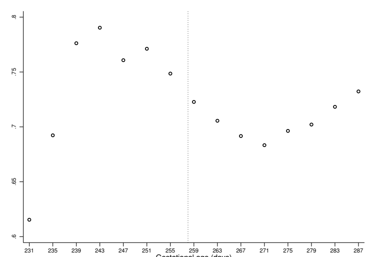
(c) Breech birth



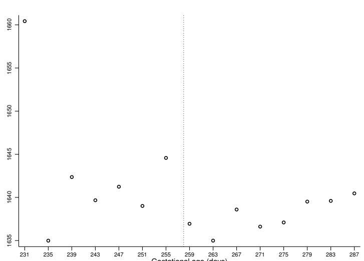
(d) Mild congenital anomaly



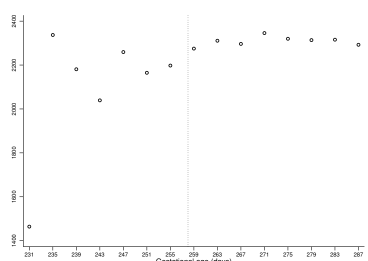
(e) Mother's age



(f) Mother's ethnicity: Dutch

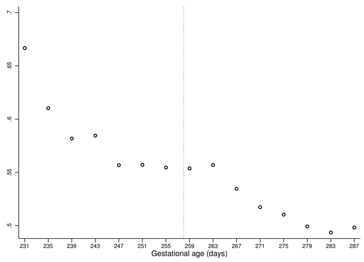


(g) Average household income

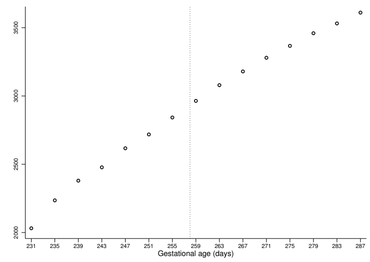


(h) Average density

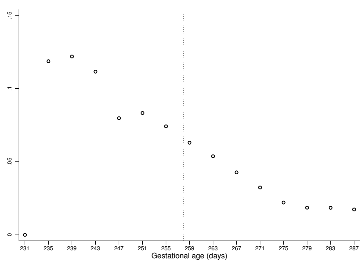
Figure A2: Distribution of selected covariates around the week-37 cutoff, first income quartile



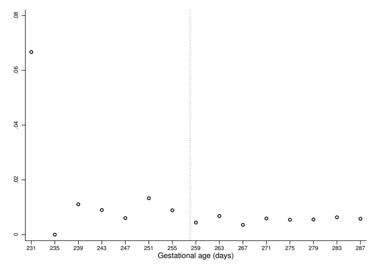
(a) Gender: male



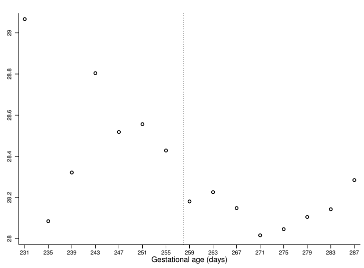
(b) Birth weight



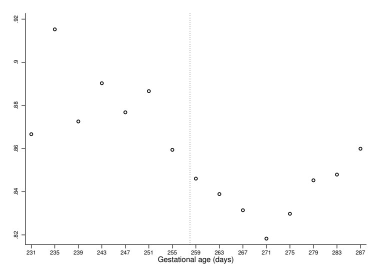
(c) Breech birth



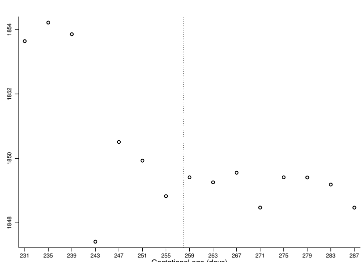
(d) Mild congenital anomaly



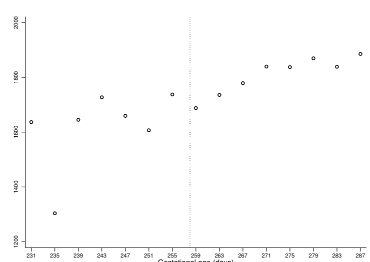
(e) Mother's age



(f) Mother's ethnicity: Dutch

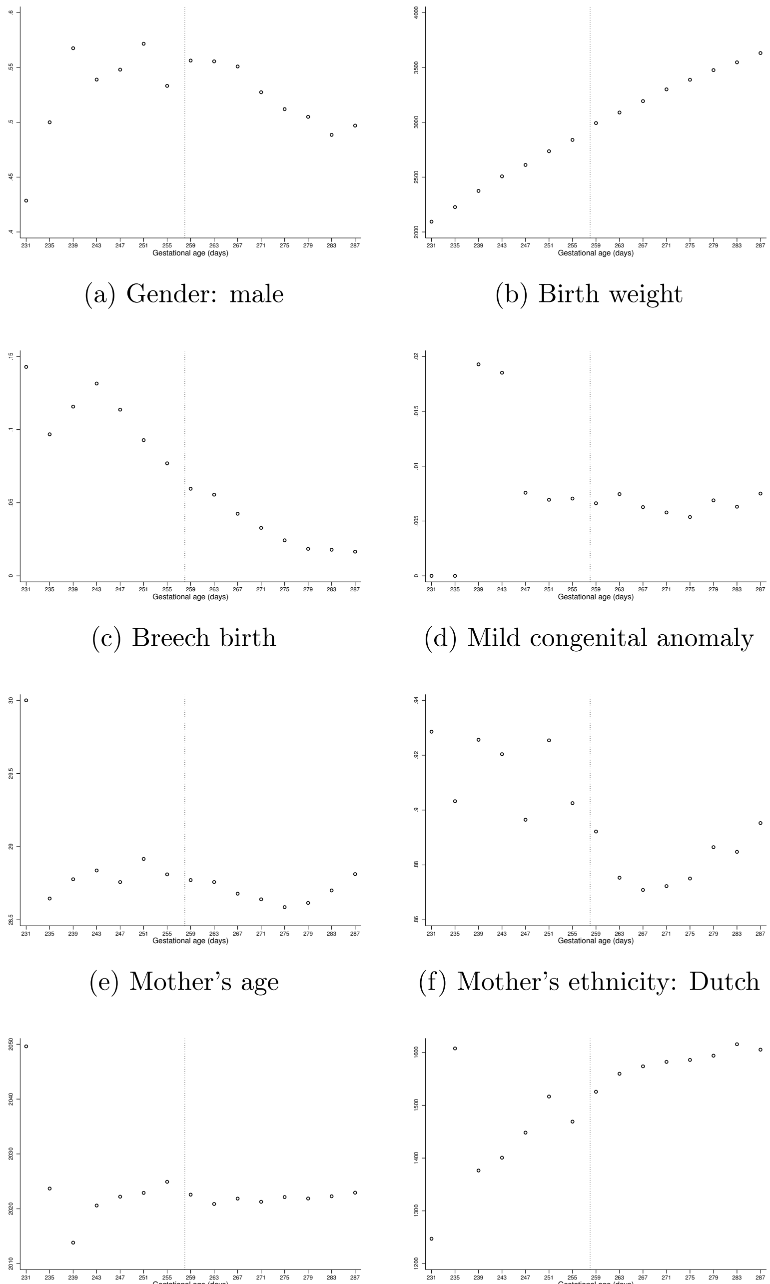


(g) Average household income



(h) Average density

Figure A3: Distribution of selected covariates around the week-37 cutoff, second income quartile



(a) Gender: male

(b) Birth weight

(c) Breech birth

(d) Mild congenital anomaly

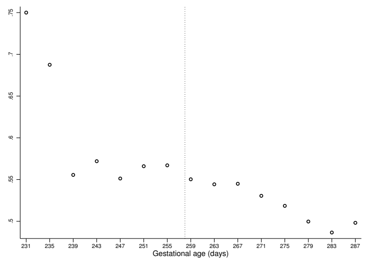
(e) Mother's age

(f) Mother's ethnicity: Dutch

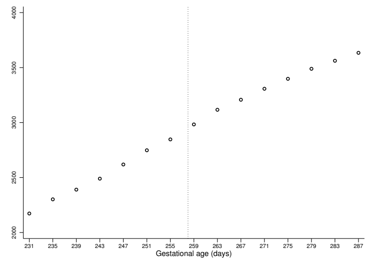
(g) Average household income

(h) Average density

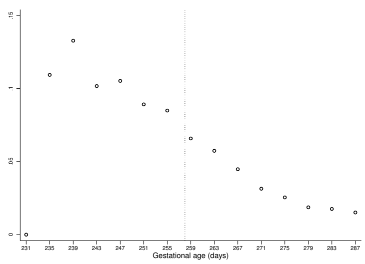
Figure A4: Distribution of selected covariates around the week-37 cutoff, third income quartile



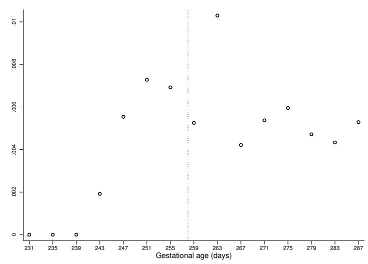
(a) Gender: male



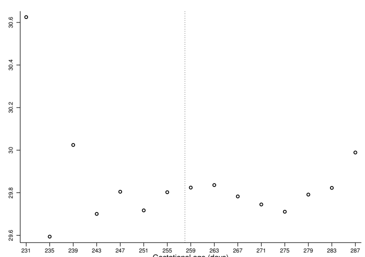
(b) Birth weight



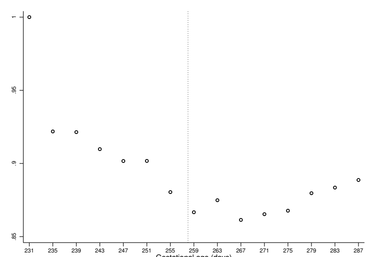
(c) Breech birth



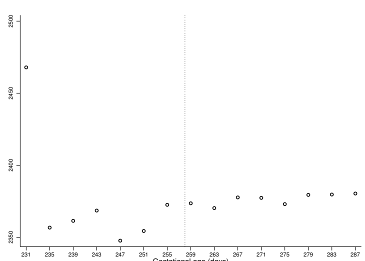
(d) Mild congenital anomaly



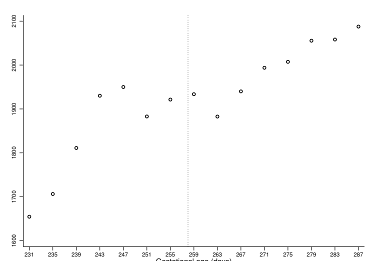
(e) Mother's age



(f) Mother's ethnicity: Dutch

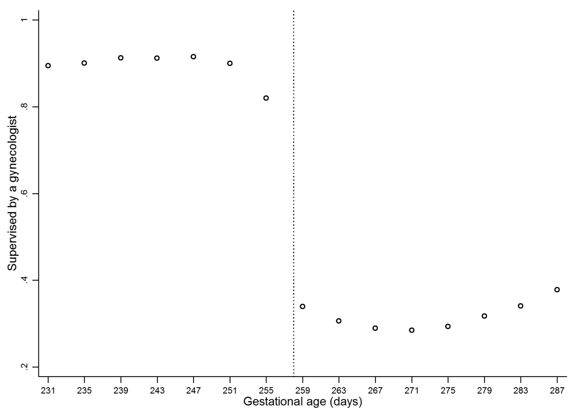


(g) Average household income

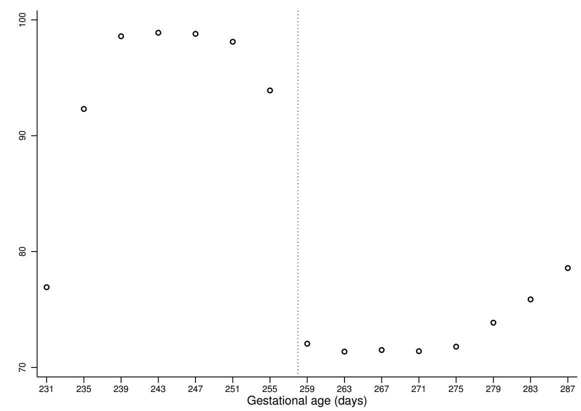


(h) Average density

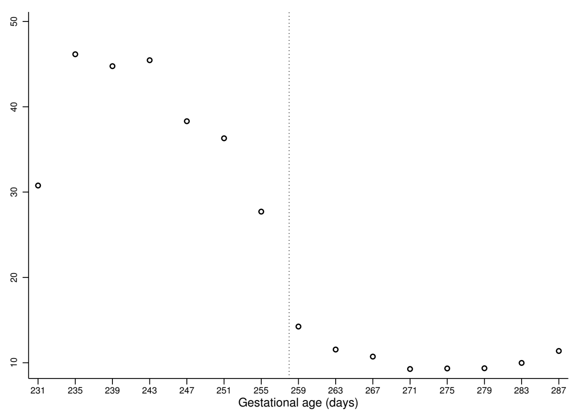
Figure A5: Distribution of selected covariates around the week-37 cutoff, fourth income quartile



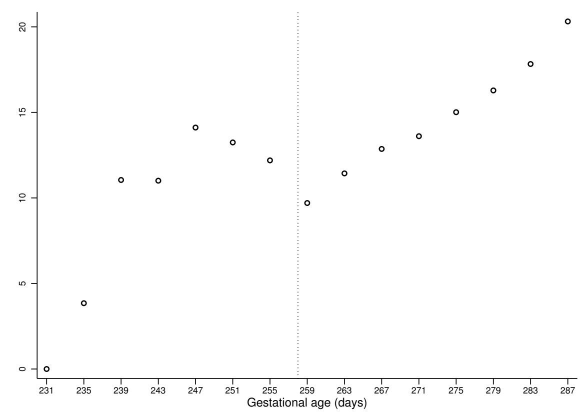
(a) Obstetrician supervision



(b) Hospital birth

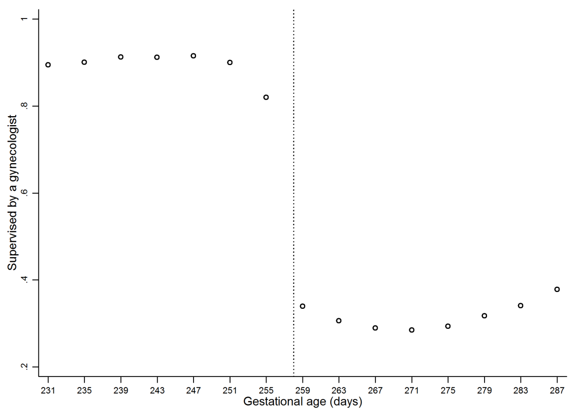


(c) NICU admission

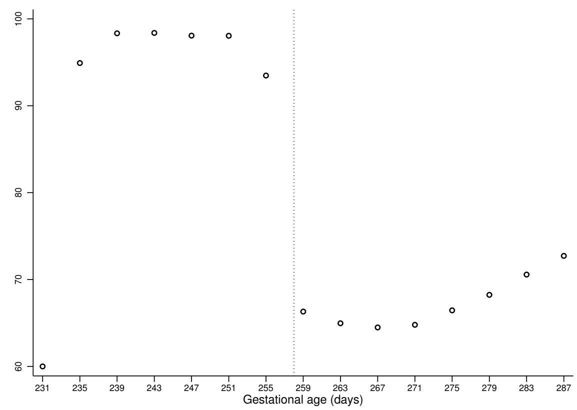


(d) Use of forceps/vacuum

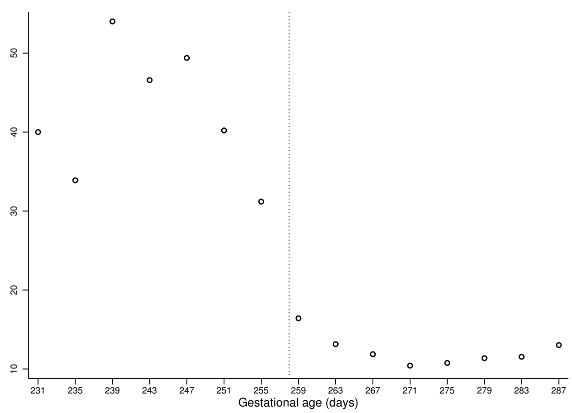
Figure A6: Medical treatments around the week-37 cutoff, first income quartile



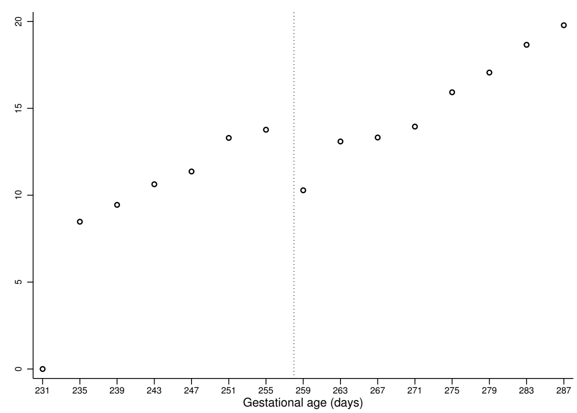
(a) Obstetrician supervision



(b) Hospital birth

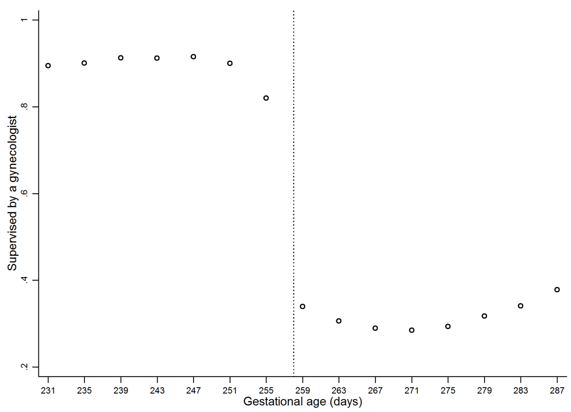


(c) NICU admission

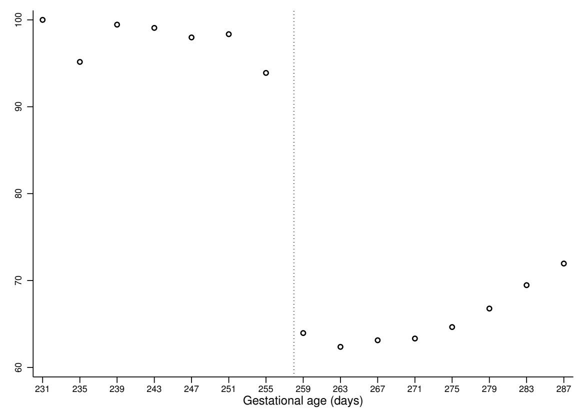


(d) Use of forceps/vacuum

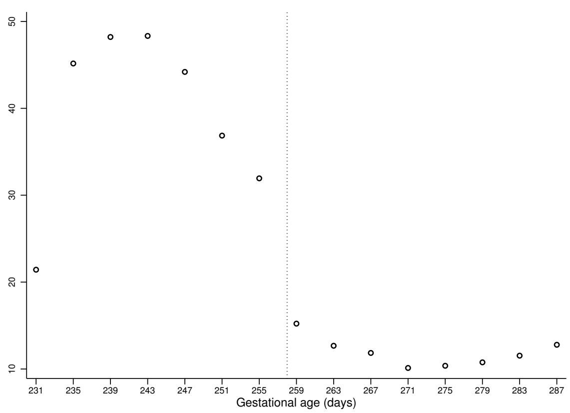
Figure A7: Medical treatments around the week-37 cutoff, second income quartile



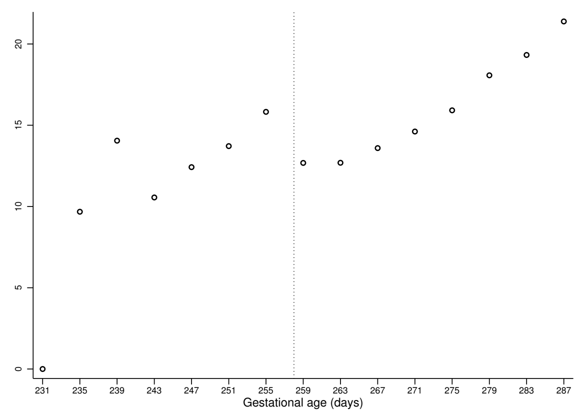
(a) Obstetrician supervision



(b) Hospital birth

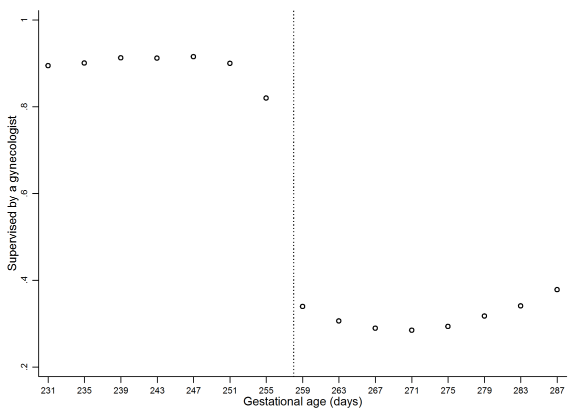


(c) NICU admission

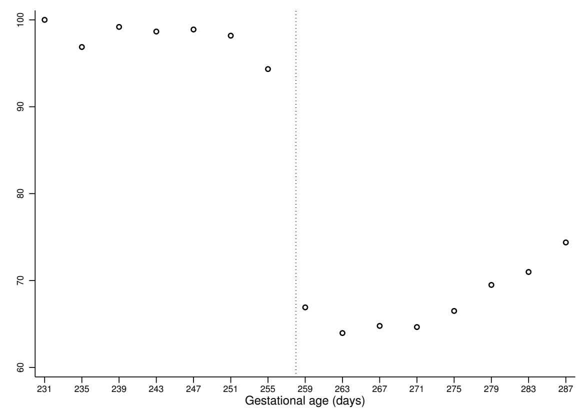


(d) Use of forceps/vacuum

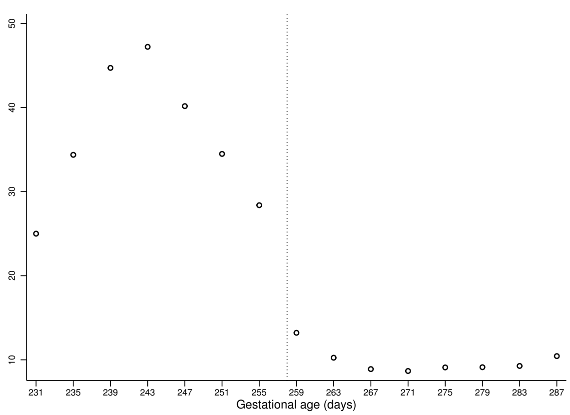
Figure A8: Medical treatments around the week-37 cutoff, third income quartile



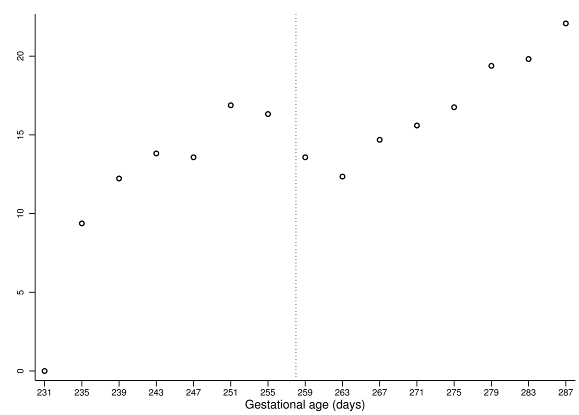
(a) Obstetrician supervision



(b) Hospital birth

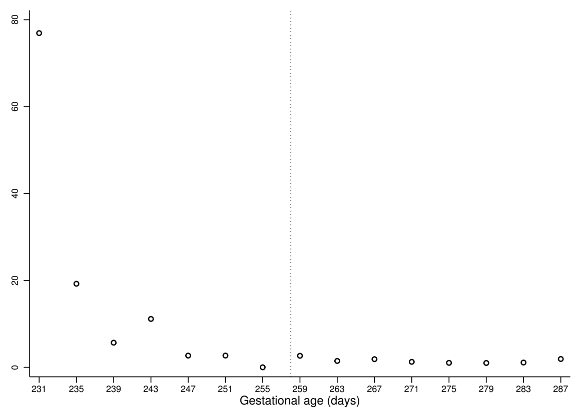


(c) NICU admission

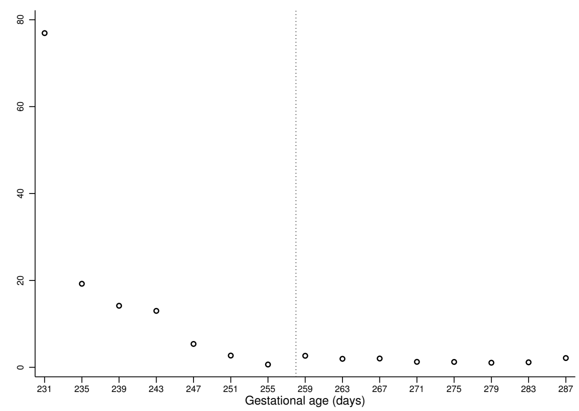


(d) Use of forceps/vacuum

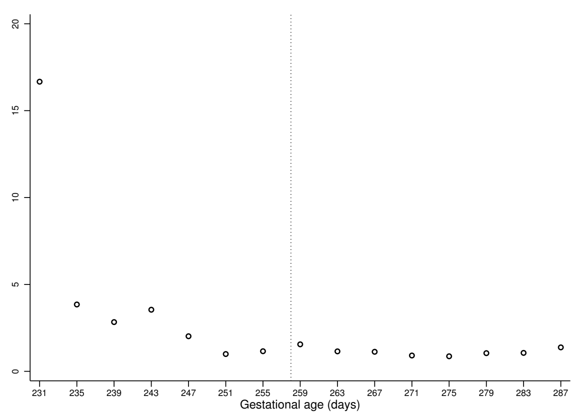
Figure A9: Medical treatments around the week-37 cutoff, fourth income quartile



(a) 7-day mortality

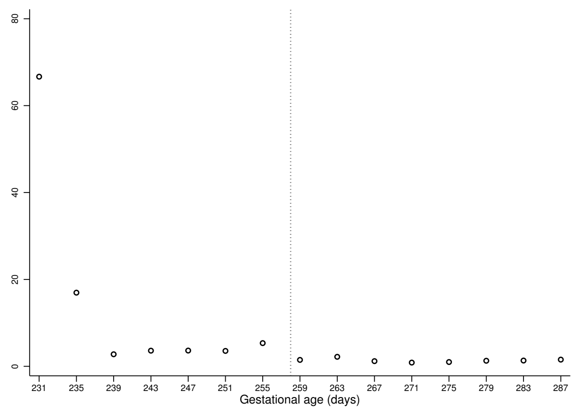


(b) 28-day mortality

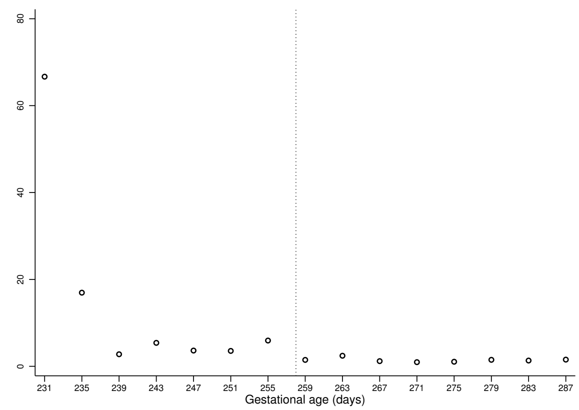


(c) Low Apgar score

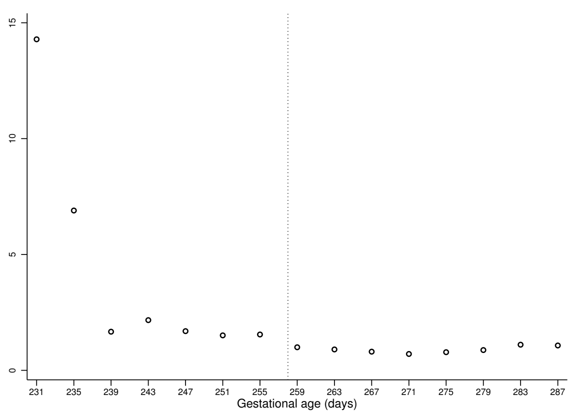
Figure A10: Newborn health around the week-37 cutoff, first income quartile



(a) 7-day mortality

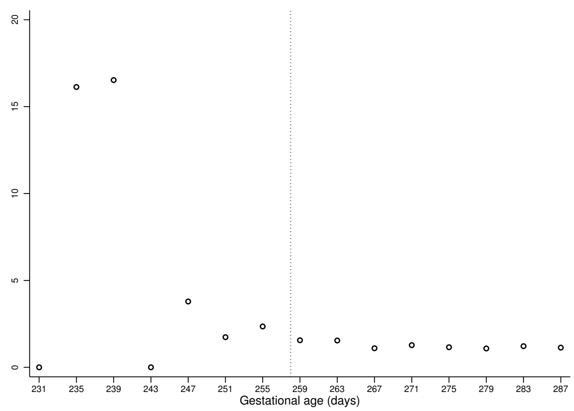


(b) 28-day mortality

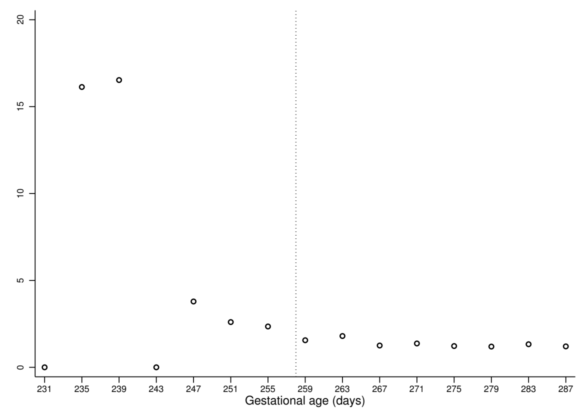


(c) Low Apgar score

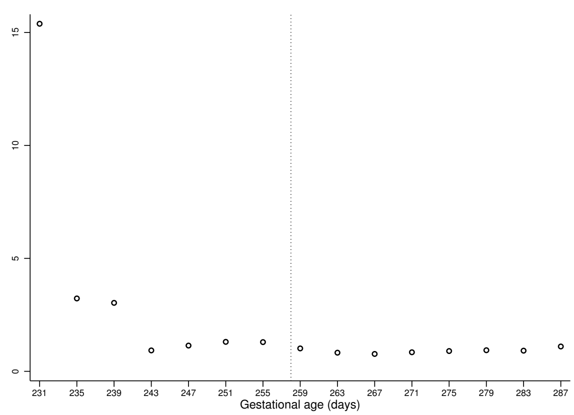
Figure A11: Newborn health around the week-37 cutoff, second income quartile



(a) 7-day mortality

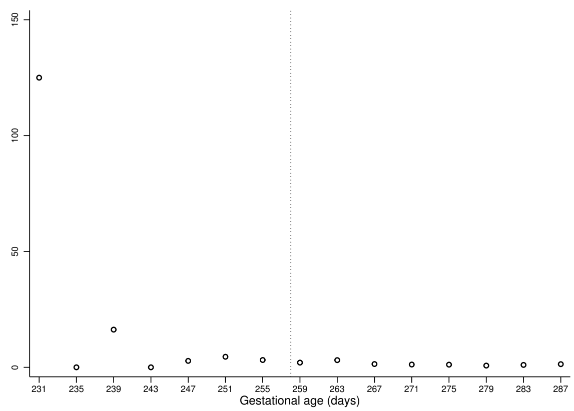


(b) 28-day mortality

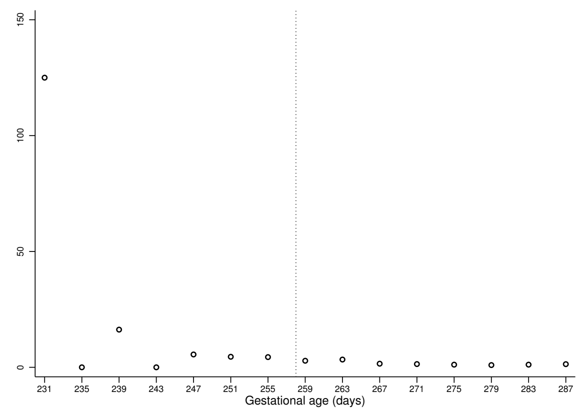


(c) Low Apgar score

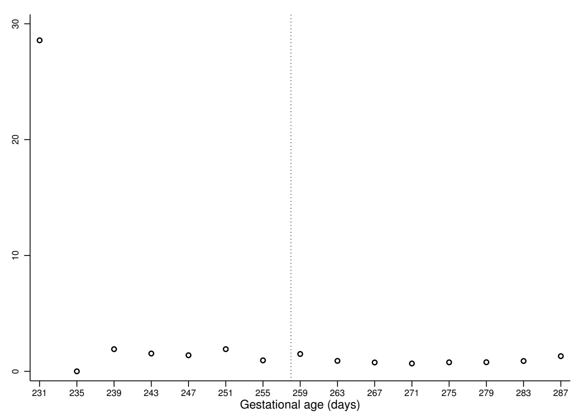
Figure A12: Newborn health around the week-37 cutoff, third income quartile



(a) 7-day mortality



(b) 28-day mortality



(c) Low Apgar score

Figure A13: Newborn health around the week-37 cutoff, fourth income quartile

Table A1: Optimal bandwidth, gestational age in days

	All	Quartile of average household income in postal code			
		First	Second	Third	Fourth
A. Health outcomes					
7-day mortality	11.98	12.60	14.07	12.37	10.79
28-day mortality	11.71	14.43	14.37	10.12	11.49
Low apgar score	13.24	15.09	16.19	16.88	13.14
B. Treatments					
Obstetrician supervision	6.21	7.93	8.26	8.21	8.29
Hospital birth	7.52	10.03	9.79	9.81	9.91
NICU admission	7.75	10.31	9.26	10.73	10.67
Use of forceps or vacuum	12.40	13.79	17.56	15.18	15.74
Emergency C-section	7.64	10.30	9.97	9.64	10.21
Any delivery intervention	8.33	10.51	10.85	11.29	10.98

Notes: See section 3 for details on the calculation of optimal bandwidths.

Table A2: Comparison of characteristics around the discontinuity by income quartile

	First quartile			Second quartile			Third quartile			Fourth quartile		
	Gestational age		Clustered	Gestational age		Clustered	Gestational age		Clustered	Gestational age		Clustered
	< 37 weeks	≥ 37 weeks	p-value	< 37 weeks	≥ 37 weeks	p-value	< 37 weeks	≥ 37 weeks	p-value	< 37 weeks	≥ 37 weeks	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Maternal characteristics												
Age	27.260	27.082	0.159	28.420	28.267	0.241	28.876	28.803	0.489	29.845	29.857	0.900
20–24	0.220	0.242	0.028	0.153	0.161	0.537	0.110	0.130	0.012	0.083	0.083	0.962
25–29	0.392	0.366	0.016	0.378	0.399	0.069	0.425	0.408	0.385	0.341	0.349	0.610
30–34	0.262	0.257	0.595	0.347	0.320	0.075	0.354	0.341	0.355	0.431	0.426	0.808
35–39	0.055	0.063	0.172	0.080	0.080	0.983	0.086	0.090	0.424	0.116	0.116	0.985
40 and above	0.008	0.004	0.042	0.005	0.005	0.902	0.005	0.006	0.317	0.010	0.011	0.883
Ethnicity												
Dutch	0.745	0.732	0.044	0.859	0.854	0.782	0.906	0.893	0.122	0.881	0.876	0.648
Mediterranean	0.116	0.117	0.936	0.052	0.047	0.641	0.032	0.029	0.601	0.021	0.019	0.561
B. Newborn characteristics												
Male	0.539	0.550	0.502	0.551	0.567	0.332	0.536	0.568	0.042	0.567	0.552	0.193
Birth weight	2,837	2,873	0.000	2,875	2,900	0.013	2,879	2,917	0.000	2,891	2,926	0.016
Low birth weight	0.156	0.136	0.114	0.125	0.129	0.601	0.113	0.105	0.527	0.104	0.101	0.853
Congenital anomaly												
Mild	0.006	0.006	0.982	0.011	0.006	0.116	0.007	0.008	0.456	0.008	0.008	0.842
Severe	0.011	0.013	0.657	0.011	0.013	0.328	0.011	0.010	0.709	0.019	0.010	0.003
Breech birth	0.058	0.066	0.245	0.072	0.069	0.704	0.072	0.068	0.601	0.082	0.075	0.301
C. Residential characteristics												
Average household income	1,643	1,635	0.001	1,849	1,850	0.631	2,024	2,022	0.039	2,370	2,369	0.911
Average density	2,194	2,291	0.043	1,708	1,658	0.423	1,494	1,521	0.647	1,919	1,862	0.300
Percent 0–15 year-old	17.719	17.683	0.599	18.040	18.161	0.164	19.182	19.022	0.129	20.414	20.702	0.179
Distance to nearest hospital (km)	4.438	4.069	0.032	5.668	5.789	0.477	5.236	5.306	0.464	4.252	4.206	0.473
Number of observations	3,698	17,686		3,959	18,247		3,944	17,321		3,701	17,241	

Notes: Each cell in the first two columns in each panel represents the mean of the corresponding variable in the row after controlling for gestational age. The last column in each panel presents the p-value for differences in means clustered at the gestational day level.