

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

Johannes Gutenberg University Mainz

This Version: November 2016

Abstract

We investigate heterogeneity in the impact of early-life medical interventions on low-risk newborns' 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. We provide suggestive evidence that potential difficulties in risk screening among low-income women cause this heterogeneity. "Back-of-the-envelope" calculations suggest that providing additional treatments to all low-risk births as per the current policy rule is barely efficient, but a targeted policy focusing on low-income areas would be highly cost-effective.

Keywords: Medical interventions, birth, heterogeneity, mortality

JEL Classifications: I11, I12, I18, J13

*Daysal: Department of Business and Economics, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: meltem.daysal@sam.sdu.dk); Trandafir: Department of Business and Economics, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: mircea.trandafir@sam.sdu.dk); Van Ewijk: Faculty of Law and Economics, Johannes Gutenberg University Mainz, Jacob-Welder-Weg 4, 55128 Mainz, Germany (e-mail: vanewijk@uni-mainz.de). Douglas Almond, Abby Alpert, Marianne Bertrand, Kitt Carpenter, John Cawley, Gordon Dahl, Hendrik Jürges, Amanda Kowalski, Amalia Miller, Hessel Oosterbeek, Martin Salm, Diane Whitmore Schanzenbach, Emilia Simonova, 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, Health and Gender Workshop at the University of Essex 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 findings, interpretations, and conclusions in the paper are entirely those of the authors. They do not necessarily represent the views of the Perined or its affiliated organizations.

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. One particular dimension along which there may be heterogeneity in the health benefits of treatments is socioeconomic status because of potential differences in (unobserved) health characteristics, health behaviors, and quality of care provided. This may lead to heterogeneity in the optimal treatment approach as well as to heterogeneity in diagnostic accuracy. In this paper, we investigate the heterogeneity by socioeconomic status 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 interventions 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 treatments 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

¹It 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.

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 37 percentage points below the 37-week threshold. Similarly, newborns slightly below the week-37 cutoff are 24 percentage points more likely to be delivered in a hospital and 10 percentage points more likely to be admitted to a neonatal intensive care unit (NICU). These estimates are economically large and correspond to increases of 36–88% 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 interventions. 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 interventions 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 interventions 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.³

There are several channels that may explain the heterogeneity in the returns to childbirth technologies. It is possible that the week-37 rule changes the quality of providers differentially across the income distribution because

²Postal 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.

³Alternative measures of socio-economic disadvantage, such as being first- or second-generation immigrant, generally yield similar conclusions but with less precision due to smaller sample sizes.

of differences in the quality of care provided by either midwives or hospitals. Unfortunately, our data do not allow us to explore this channel. Second, 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. Finally, treatment heterogeneity may arise if midwives are less able to identify higher-risk mothers among lower-income individuals (e.g., due to difficulties in communications about health-related issues or due to unobserved health behaviors). If this is the case, some high-risk low-income women may be classified as low-risk and their infants may benefit more from the additional medical treatments provided due to the week-37 rule. We provide suggestive evidence on this channel by showing that average predicted newborn health improves monotonically with income.

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 for 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. While Almond and Doyle (2011) show that longer hospital stays do not affect infant health outcomes after uncomplicated deliveries, Miller (2006) finds that midwifery-promoting public policies were associated with lower neonatal mortality.⁴ Particularly relevant to our study is Daysal et al. (2015), who also use data from the Netherlands and 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. The au-

⁴Some medical research finds higher rates of adverse events such as a low Apgar score or asphyxia in midwife-supervised as opposed to physician-supervised deliveries Wernham et al. (2016), but no increases in infant mortality Wiegerinck et al. (2015). However, these studies likely suffer from an omitted variable bias.

thors provide suggestive evidence that proximity to other medical technologies, such as neonatal intensive care units (NICU), may be an important channel contributing to the health gains from a hospital birth.

Our paper is most closely related to an emerging literature, both in economics and medical sciences, studying the heterogeneity in returns to medical interventions (Kravitz et al., 2004; Chandra and Skinner, 2012). To the best of our knowledge, the only paper to explicitly examine heterogeneity in the 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.

Our paper makes several contributions. First, we exploit a new source of variation in early-life medical interventions. To this end, we utilize a policy rule in The Netherlands that affects children on the borderline of prematurity (37 weeks of gestation). Focusing on this relatively understudied group of children yields widely relevant findings because “[o]n a global level, given their relatively larger numbers, babies born at 34 to 36 weeks are likely to have the greatest public health impact and to be of the most importance in the planning of services.” (March of Dimes et al., 2012, p. 30)

Second, our results suggest that ensuring access to medical interventions 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 past medical history, the current health status of the mother and the fetus, and their development throughout pregnancy. 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, who are predicted to have babies in poorer health, suggesting that risk selection is more difficult and less

precise for women with lower socioeconomic status.

Finally, our results imply that crafting cost-effective health policies requires a solid understanding of the potential sources of heterogeneity in treatment effects. In our context, “back-of-the-envelope” calculations suggest that the week-37 rule is barely efficient as currently applied to all low-risk births. However, a targeted policy focusing on low-income areas would turn it into a highly cost-effective measure. Failing to account for this heterogeneity would result either in waste in health care or even in the elimination of a program that saves lives among the poor, potentially contributing to existing health disparities. 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. The safety of this system is currently hotly debated in The Netherlands among researchers, the general public, as well as policy makers ([Evers et al., 2010](#); [Wiegerinck et al., 2015](#)).

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.⁵

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 interventions 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 arbi-

⁵During obstetrician-supervised deliveries, there is also a midwife present, but this is a different midwife from the one who supervised the prenatal care. He or she is employed by the hospital (rather than by a midwifery practice) and is specialized in dealing with higher-risk deliveries.

trary 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:

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

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. The coefficient of interest β captures the intention-to-treat effect of the week-37 rule: 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 triangular kernel which places higher weights on observations closer to the cutoff and we cluster the standard errors in all regressions at the

gestational day level.

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 (3.438 for the triangular 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 the week-37 cutoff.

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 midwives (LVR-1), obstetricians/gynecologists (LVR-2) and paediatricians (LNR).⁶

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

⁶Perined 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).

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.⁷

Our outcomes include a number of variables pertaining to medical interventions 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 treatments during and after delivery: obstetrician supervision of birth, delivery in a hospital, use of forceps or vacuum, and admission to a 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.⁸

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.⁹

⁷Average 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.

⁸We 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.

⁹Alternatively, we can define gestational age at the onset of labor as gestational age at

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.¹⁰ The third group (infant characteristics) includes birth weight and indicators for gender, congenital anomalies and birth position.¹¹ The final group (postal code characteristics) includes the average characteristics of the postal code of residence of the mother: monthly household income, area density and the fraction of residents 0–15 years old.¹²

Our analysis sample includes live deliveries by low-risk women with gestational age between 245 and 272 days. We focus on low-risk women because the week-37 rule does not apply to high-risk women.¹³ 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

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.

¹⁰We 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).

¹¹Specifically, 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).

¹²Some 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.

¹³The 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.

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, 7 days before the lower bound of our target interval.¹⁴ We also restrict our sample to first 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.¹⁵ 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 85,246 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.¹⁶

¹⁴We thank Gordon Dahl for this suggestion. The date of referral is missing for about 4.5 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.

¹⁵Indeed, we find that parity is discontinuous across the week-37 cutoff among all low-risk births: newborns below the cutoff are significantly more likely to be first-born relative to those slightly above the cutoff. This holds both overall and across the income distribution (results available upon request).

¹⁶Income quartiles are defined using the entire population of births. Appendix Table A2 details how the analysis sample is constructed.

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 interventions 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 within our bandwidth. 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. 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 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.¹⁷

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 before and after the cutoff.¹⁸ The Figure shows that the distribution of the covariates is

¹⁷The estimated discontinuity at the cutoff is 139.493 (s.e. 239.298). 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: 0.914 (s.e. 56.199), 31.946 (s.e. 62.696), 53.382 (s.e. 59.831), 53.252 (s.e. 68.971).

¹⁸For 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.

smooth around the discontinuity. In order to examine this issue more formally, we also estimate in Table 1 local-linear regressions using the covariates as the dependent variables. The results 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, and average characteristics in the postal code of the residence of the mother. It is worth noting that even in the few cases where we find statistically significant differences, the difference in the magnitudes is very small with no clear pattern. For example, infants born before day 259 are on average 31 grams lighter than those born after the cutoff.¹⁹ On the other hand, mothers of preterm infants reside in postal codes where the average monthly household income is higher by €15.

When we split the sample by household income, we find some evidence that preterm babies in the highest income quartile have worse observable characteristics than at-term babies from the same quartile (a higher likelihood of severe congenital anomalies and perhaps a higher likelihood of breech presentation). However, we conduct a large number of tests so a few significant results may be expected purely based on chance. The patterns otherwise generally follow those in the full sample and indicate smooth distributions across the cutoff for the majority of covariates. This lends support to the claim that the variation in medical interventions near the week-37 cutoff is as good as random.

5.2 The Discontinuity in Medical Treatments

If the Dutch institutional rule governing the supervision of premature births is binding, then we should observe a discontinuity in receipt of medical interventions at 37 completed gestational weeks. To examine this, in Figure 3 we plot several medical treatments around the cutoff. Visually, there is a substan-

¹⁹The 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. Our estimated jump may be an artifact, resulting from a slight nonlinearity in the relation between gestational age and birth weight.

tial 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).²⁰

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). Each cell reports the coefficient of $W37$ from a different regression. The results suggest that preterm newborns receive significantly more medical interventions: they are, on average, 37 percentage points more likely to be supervised by an obstetrician; 24 percentage points more likely to be delivered in a hospital; 10 percentage points more likely to be admitted to a NICU within the first seven days of life; and 3 percentage points more likely to be delivered by use of forceps and vacuum. These estimates are economically large and represent increases of 25–88% 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, 8.43 percentage points more likely to be admitted to a NICU, which represents a 75% increase at the mean above the cutoff. Preterm newborns in the highest income quartile, on the other hand, are 11.62 percentage points more likely to be admitted to a NICU, a 119% increase at the mean above the cutoff. Similarly, the week-37 rule increases the probability of a hospital birth by 27% for preterm newborns

²⁰There 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 by OB/GYNs.

in the lowest quartile and by 39% 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.

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. 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 the cutoff.

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 interventions.²¹ 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 0.19 fewer infant deaths per 1,000 births for 28-day mortality and a 0.03 percentage point reduction in the probability of low Apgar score. In stark contrast to these findings,

²¹Appendix Figures A10–A13 provide the corresponding visual evidence.

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.

5.4 Robustness Checks

We next investigate the robustness of our results to model specification. If the key assumption in our RD design is satisfied (i.e., the variation in receipt of medical interventions 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. The results, shown in Panel B of Table 4, support our main conclusions: the week-37 rule leads to health gains among the poorest quartile, but not among the other income quartiles. Panel C focuses on the choice of kernel and reports results based on a rectangular kernel which places equal weights on observations. Our results again point to health gains for babies slightly below the week-37 cutoff in the lowest income quartile and no benefits for preterm babies in the higher income quartiles. In Panels D and E, 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.

In Table 5, we examine the role of our sample selection criteria. Recall that our analysis sample includes women under the supervision of a midwife at least until gestational day 238. In Panels A and B, 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. In Panel C, we check the sensitivity of our findings to including observations with missing referral dates and show that the results are very similar to our baseline estimates. Finally Panel D checks the robustness of our results to including planned C-sections. These deliveries are excluded from our analysis sample to avoid a potential bias from manipulation of the running variable. However, we may have a selection bias if some births ending in an unplanned C-section are coded as planned and if this practice changes across the cutoff. The results in Panel D suggest that this is unlikely to be the case.²²

Table 6 investigates whether the health gains observed in the lowest income quartile are driven by our specific characterization of the income distribution. Using income terciles and quintiles, we confirm that only newborns from the lowest income areas experience mortality reductions from the week-37 rule. In addition, the magnitudes of the estimates generally increase as the income bins become smaller. This supports the claim that the health benefits due to the week-37 rule may be a decreasing function of income.

Finally, we check whether we observe similar reductions in adverse newborn outcomes at other points in the distribution of gestational age. If the observed gains in health in the lowest income quartile are indeed driven by the week-37 rule, then we should not observe systematic discontinuities in newborn health outcomes at other potential cutoffs. We examine cutoffs from 35 completed gestational weeks (245 days) to 41 completed gestational week (287 days), keeping the bandwidth fixed at 14 days on either side of the cutoff. Figure 5 plots the estimated coefficients and the 95% confidence interval. While the estimates are noisier at lower gestational ages due to small sample sizes, we find

²²We have also checked the robustness of our results to excluding all referrals and focusing only on women under the care of a midwife at the onset of labor and we confirmed higher health gains for the lowest income quartile with generally no benefits for the other three quartiles (results available upon request).

that there is no other cutoff where all three measures of newborn health exhibit statistically significant gains of a magnitude comparable to those observed at the week-37 cutoff.

5.5 Potential Mechanisms

Our results consistently show that the overall impact of the week-37 rule hides substantial heterogeneity in the effects by income, with significant health benefits only to low-income mothers. There are several mechanisms that may explain this heterogeneity in the returns to early-life medical interventions. To begin with, the week-37 rule may change the quality of providers differentially across the income distribution. This may be because low-income mothers have access to poorer quality midwife care or to better quality hospitals. Unfortunately, data limitations prevent us from investigating this specific channel.

Second, 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. The average distances to the closest hospital listed in the last row of Panel C in Table 1 show no consistent pattern for such residential sorting by income. In addition, the point estimates in Column 2 indicate that, if anything, low-income mothers with preterm babies actually reside farther away from a hospital. As such, this scenario is unlikely to drive our results.

Finally, 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. Alternatively, some low-income mothers may engage in unhealthy behaviors that are not included as reasons for referral to an obstetrician in the List of Obstetric Indications. In this case, the pool of low-income mothers classified as low-risk may have on average worse unobserved underlying health than the higher-income mothers classified as low-risk (in other words, 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.

There is some anecdotal evidence suggesting that this channel may be important. A survey by the Royal Dutch Organisation of Midwives reports that midwives needed on average 23 percent extra time when caring for low-income women (Buisman and Gerats, 2008). 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. In order to shed some light on this scenario, we calculate average predicted newborn health by income quartile. We find that all our measures of predicted newborn health improve monotonically as income increases, suggesting that risk selection may indeed be an important factor explaining the heterogeneity in returns to medical interventions by income.²³

5.6 Cost of a Life Saved

In order to gauge the economic significance of our findings, we conduct a “back-of-the-envelope” calculation of the cost of a life saved due to the week-37 rule using deliveries within our bandwidth. Table A3 details the calculations. Our results indicate that the week-37 rule increases the number of obstetrician-supervised hospital deliveries by 345–390 per 1,000 births. Of these, roughly 191–265 represent transfers from midwife-supervised home births to OB/GYN-supervised hospital births, while the rest are transfers from midwife-supervised hospital births. In 2016, the cost of a midwife-supervised home birth was €519.60, whereas a midwife-supervised hospital birth had an additional cost of €589.80 (NZA, 2015). On the other hand, the average cost of an uncomplicated hospital delivery under the care of an obstetrician was €2,250 (NZA, 2016). Hence, our results imply a cost increase of €561,914 per 1,000 births

²³Predictions are based on a regression model including all the observable characteristics. Average predicted 7-day mortality rates, from the first to the fourth income quartile, are: 2.285, 2.198, 2.091, 2.058. The corresponding numbers for predicted 28-day mortality are: 2.503, 2.393, 2.280, 2.235. Finally, the predicted fractions of newborns with low Apgar score are: 1.272, 1.186, 1.123, 1.094.

due to higher use of OB/GYN-supervised hospital births because of the week-37 rule.

Our estimates also suggest that prematurity leads to 13–43 more deliveries aided by use of forceps/vacuum. According to 2016 prices, use of forceps and vacuums increases the mean cost of a hospital delivery by €465 (NZA, 2016), adding €14,751 per 1,000 births. Turning to NICU admissions, we find that there are 84–119 more NICU admissions within the first week of life due to the week-37 rule. Our best estimate for the cost of a NICU stay, based on the same price listed by several hospitals and insurers, is €9,151.79.²⁴ This implies that the week-37 rule is associated with an additional cost of €918,403 per 1,000 births.

Overall, the estimated additional cost of the week-37 rule per 1,000 deliveries is roughly €1.5 million. Compared to an average reduction of 0.721–0.812 deaths per 1,000 births, the implied cost per life saved is €1.8–2.1 million.²⁵ Previous studies calculate the value of a statistical life in the Netherlands to be €2.6 million in 2009 prices (SWOV, 2012), which amounts to €2.83 million in 2015 prices. Taken together, these results suggest that the current implementation of the week-37 rule is borderline cost-effective. However, if the rule applied only to the lowest-income quartile, the estimated cost of a life saved would lie between €0.39–0.44 million. In this case, the treatments omitted from our calculation would need cost at least five times as much as those included in order to make the policy inefficient.

²⁴This estimate is based on the average length of stay in the NICU and the average length of stay in post-IC high care for babies born between weeks 34–40 (Perined, 2015).

²⁵On the one hand, these numbers represent an underestimate of the true cost of the week-37 rule because we are unable to include the costs of all treatments that increase at the cutoff. The NICU costs are underestimated because they do not include additional NICU-related costs such as transportation, certain treatments such as extracorporeal life support, or post-NICU follow-up care that does not require a hospital admission. Furthermore, some costs (e.g., additional checks after birth complications) could not be included. On the other hand, these calculations do not incorporate all benefits since the week-37 rule may not only reduce mortality, but also reduce the prevalence of conditions such as perinatal asphyxia (oxygen deprivation) which can lead to permanent brain damage and consequent additional treatments and loss of human capital.

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 interventions, 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 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 interventions 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 could be an important factor.

Our results are relevant to the ongoing policy debates on effective health policy. The fact that medical interventions improve the health outcomes of some newborns even among low-risk women living in a developed country with a long history of risk selection suggests that even relatively sophisticated models of risk selection may fail to identify all high-risk individuals. This cautions against designing “one-size-fits-all” policies and indicates that a good understanding of the causes of treatment heterogeneity should go hand in hand

with crafting policies about child birth technologies.

References

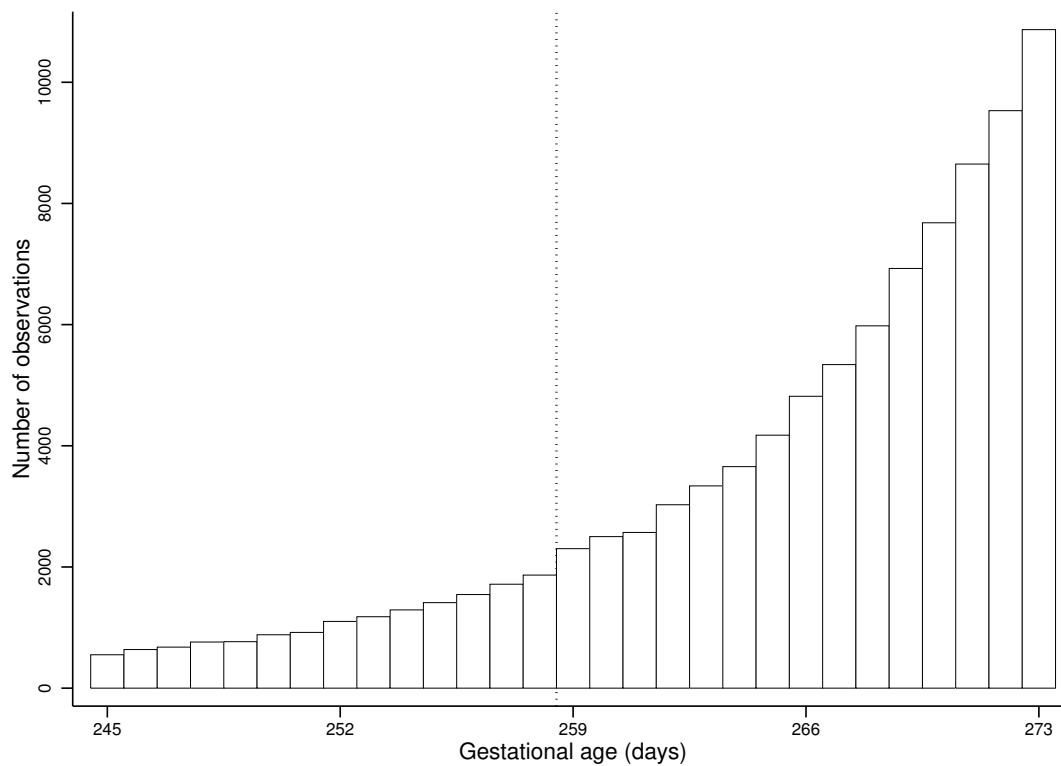
- Almond, Douglas, and Joseph Doyle (2011), “After midnight: A regression discontinuity design in length of postpartum hospital stays,” *American Economic Journal: Economic Policy* 3(3), 1–34.
- Almond, Douglas, Joseph Doyle, Amanda Kowalski, and Heidi Williams (2010), “Estimating marginal returns to 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.
- Buisman, Peter, and Geert Gerats (2008), “Verloskundigen in tijd erop vooruit in achterstandswijken,” *Tijdschrift voor Verloskundigen* September, 13–16.

- 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.
- Lee, David, and Thomas Lemieux (2010), “Regression discontinuity designs in economics,” *Journal of Economic Literature* 48(2), 281–355.
- March of Dimes, PMNCH, Save the Children, and WHO (2012), *Born Too Soon: The Global Action Report on Preterm Birth*, Geneva: World Health Organization, edited by CP Howson, MV Kinney, and JE Lawn.

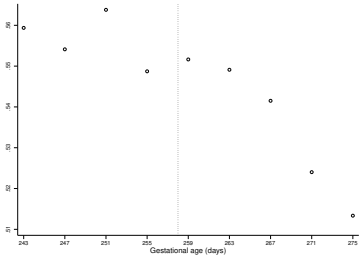
- McClellan, Mark, and Joseph Newhouse (1997), “The marginal cost-effectiveness of medical technology: A panel instrumental-variables approach,” *Journal of Econometrics* 77(1), 39–64.
- McCrary, Justin (2008), “Manipulation of the running variable in the regression discontinuity design: A density test,” *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 (2015), “Tariefbeschikking TB/CU-7130-01,” available online at https://www.nza.nl/regelgeving/tarieven-en-prestaties/TB_CU_7130_01_Maximumtarieven_verloskundige_zorg, last accessed on Oct. 28, 2016.
- (2016), “DBC-zorgproducten,” available online at <http://www.opendisdata.nl/msz/zorgproduct>, last accessed on Oct. 28, 2016.
- Perined (2015), *Perinatale Zorg in Nederland 2014*, Utrecht.
- 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.
- SWOV (2012), “Factsheet: The valuation of human costs of road deaths,” available online at https://www.swov.nl/rapport/Factsheets/UK/FS_Immaterial_costs.pdf, last accessed on Oct. 28, 2016.
- van der Kooy, Jacoba, Jashvant Poeran, Johanna de Graaf, Erwin Birnie, Semiha Denктаş, Eric Steegers, and Gouke Bonsel (2011), “Planned home compared with planned hospital births in the Netherlands,” *Obstetrics & Gynecology* 118, 1037–1046.
- Wernham, Ellie, Jason Gurney, James Stanley, Lis Ellison-Loschmann, and Diana Sarfati (2016), “A comparison of midwife-led and medical-led models of care and their relationship to adverse fetal and neonatal outcomes: A retrospective cohort study in New Zealand,” *PLoS Med* 13(9), e1002134.
- Wiegerinck, MMJ, BY van der Goes, ACJ Ravelli, JAM van der Post, J Klinkert, J Brandenburg, FCD Buist, MGAJ Wouters, P Tamminga, A de Jonge, et al. (2015), “Intrapartum and neonatal mortality in primary midwife-led and secondary obstetrician-led care in the Amsterdam region of the Netherlands: A retrospective cohort study,” *Midwifery* 31(12), 1168–1176.

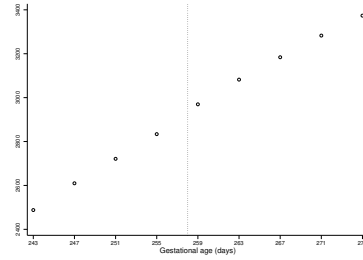


]

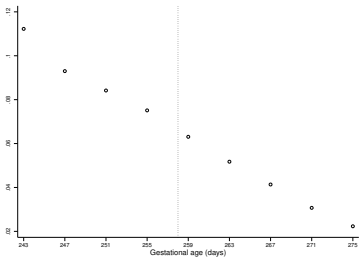
Figure 1: Frequency of births around the week-37 cutoff



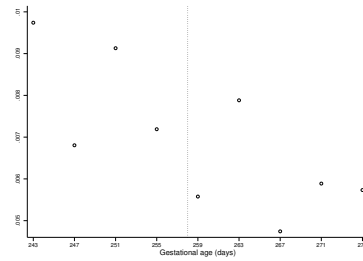
(a) Gender: male



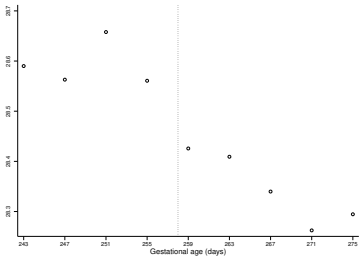
(b) Birth weight



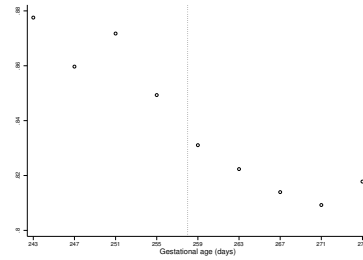
(c) Breech birth



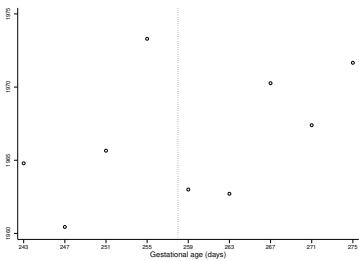
(d) Mild congenital anomaly



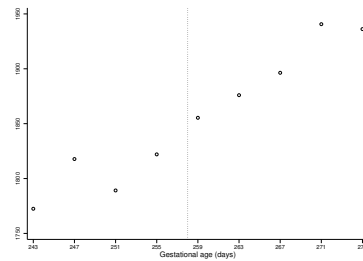
(e) Mother's age



(f) Mother's ethnicity: Dutch

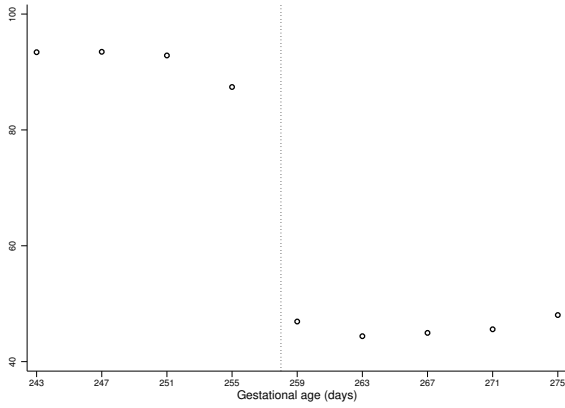


(g) Average household income

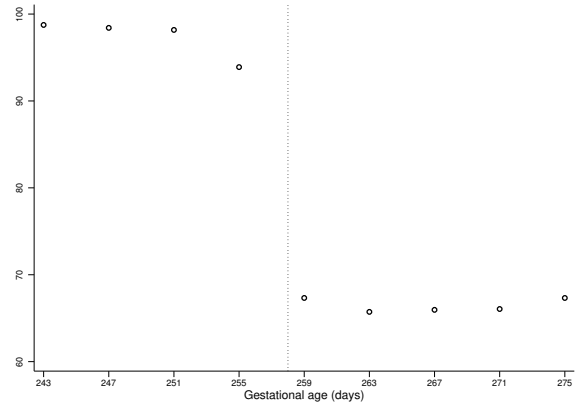


(h) Average density

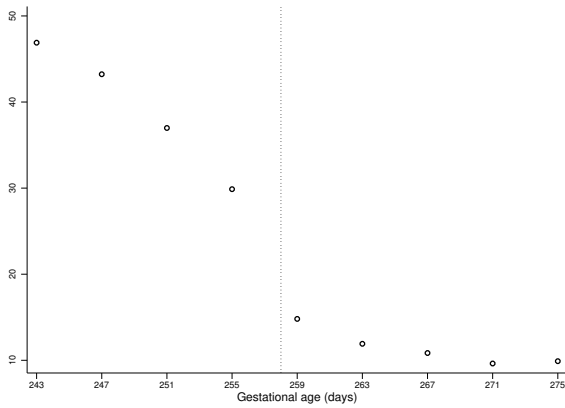
Figure 2: Distribution of selected covariates around the week-37 cutoff



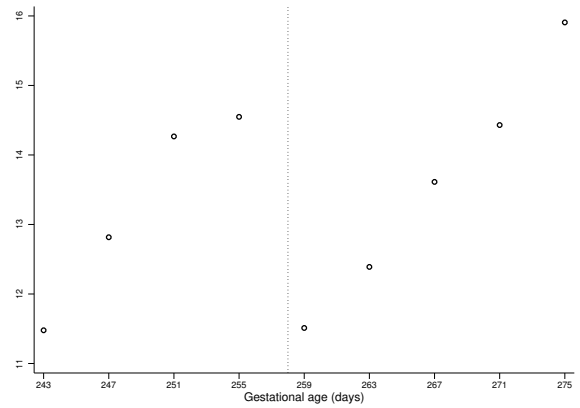
(a) Obstetrician supervision



(b) Hospital birth

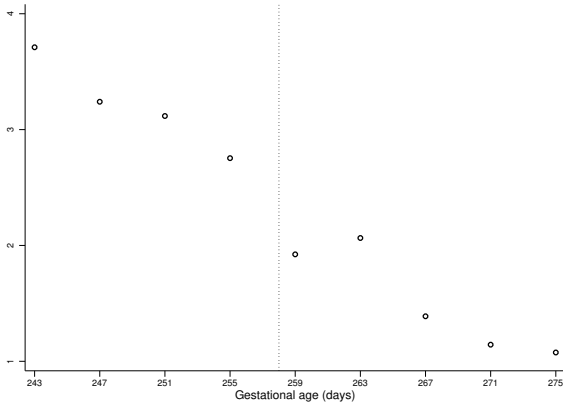


(c) NICU admission

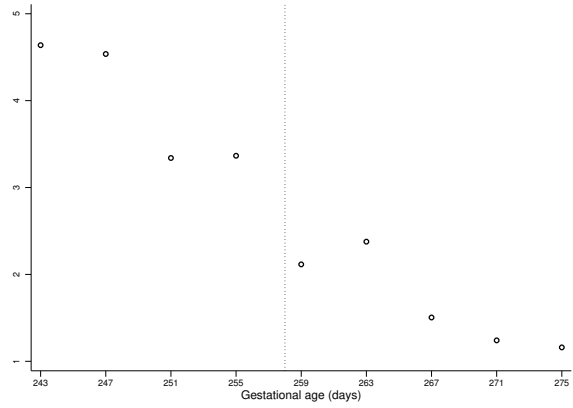


(d) Use of forceps/vacuum

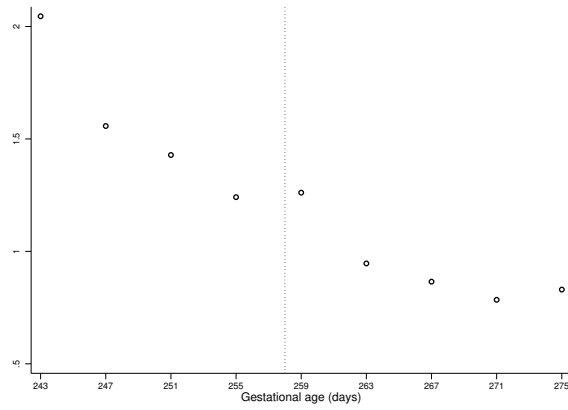
Figure 3: Medical treatments around the week-37 cutoff



(a) 7-day mortality

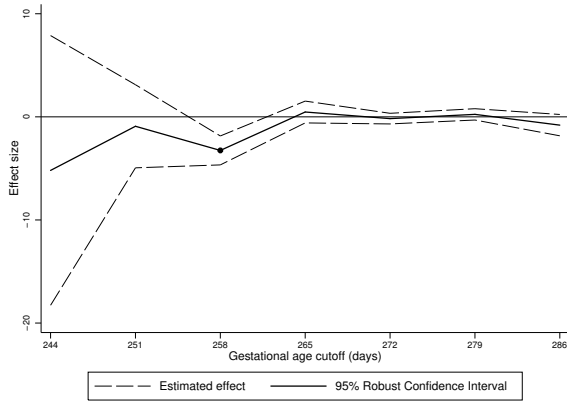


(b) 28-day mortality

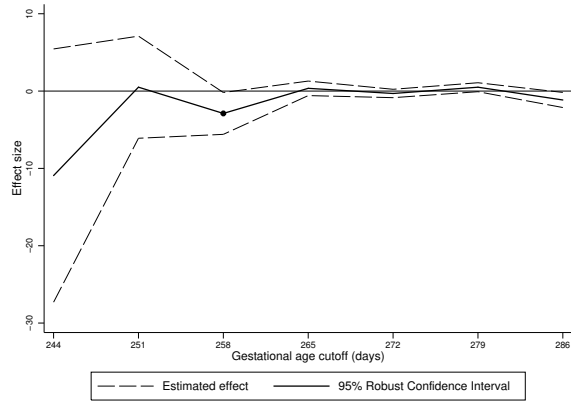


(c) Low Apgar score

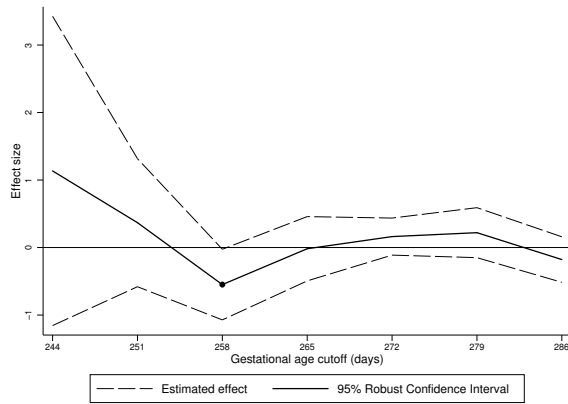
Figure 4: Newborn health around the week-37 cutoff



(a) 7-day mortality



(b) 28-day mortality



(c) Low Apgar score

Figure 5: Newborn health across the gestational age distribution, first income quartile

Table 1: Covariates around the week-37 cutoff

	All	Quartile of average household income in postal code			
	(1)	First (2)	Second (3)	Third (4)	Fourth (5)
A. Maternal characteristics					
Age	0.116* (0.061)	0.060 (0.123)	0.208 (0.137)	0.033 (0.105)	-0.001 (0.092)
Mean outcome	28.354	26.843	28.135	28.701	29.787
Observations	85,246	21,259	22,050	21,137	20,800
Ethnicity: Dutch	0.007 (0.006)	0.009 (0.006)	-0.000 (0.020)	0.005 (0.007)	0.003 (0.011)
Mean outcome	0.817	0.697	0.832	0.874	0.865
Observations	385,246	21,259	22,050	21,137	20,800
Ethnicity: Mediterranean	0.004 (0.004)	0.010 (0.006)	0.007 (0.012)	0.005 (0.004)	-0.000 (0.003)
Mean outcome	0.063	0.129	0.057	0.036	0.028
Observations	85,246	21,259	22,050	21,137	20,800
B. Newborn characteristics					
Male	-0.009 (0.007)	-0.018 (0.018)	-0.003 (0.014)	-0.035** (0.015)	0.022* (0.012)
Mean outcome	0.542	0.537	0.541	0.548	0.543
Observations	85,246	21,259	22,050	21,137	20,800
Birth weight	-31.133*** (5.613)	-32.556*** (7.987)	-18.524* (9.170)	-38.960*** (8.668)	-38.719*** (13.315)
Mean outcome	3148.143	3113.914	3142.691	3161.847	3175.258
Observations	85,246	21,259	22,050	21,137	20,800
Mild congenital anomaly	0.001 (0.001)	-0.001 (0.002)	0.004 (0.004)	0.000 (0.002)	-0.001 (0.003)
Mean outcome	0.006	0.006	0.005	0.006	0.006
Observations	85,246	21,259	22,050	21,137	20,800
Severe congenital anomaly	0.002 (0.003)	-0.003 (0.004)	-0.002 (0.002)	0.003 (0.004)	0.011*** (0.004)
Mean outcome	0.009	0.010	0.010	0.008	0.008
Observations	85,246	21,259	22,050	21,137	20,800
Breech birth	0.006 (0.006)	-0.006 (0.008)	0.008 (0.011)	0.008 (0.007)	0.011* (0.007)
Mean outcome	0.045	0.039	0.046	0.046	0.048
Observations	85,246	21,259	22,050	21,137	20,800

Table 1: Covariates around the week-37 cutoff (cont'd)

	All	Quartile of average household income in postal code			
	(1)	First (2)	Second (3)	Third (4)	Fourth (5)
C. Residential characteristics					
Average household income	15.291*** (2.789)	9.324*** (2.483)	-0.967 (1.182)	2.541** (1.073)	10.240 (6.233)
Mean outcome	1,967.265	1,637.355	1,849.366	2,021.428	2,376.055
Observations	85,246	21,259	22,050	21,137	20,800
Average density	-14.339 (24.758)	-101.004* (49.305)	79.416 (81.366)	-53.342 (62.854)	49.458 (58.361)
Mean outcome	1,894.099	2,300.358	1,771.135	1,564.385	1,938.739
Observations	85,246	21,259	22,050	21,137	20,800
% 0-15 year-old	-0.032 (0.073)	0.053 (0.076)	-0.148 (0.104)	0.171* (0.093)	-0.364 (0.215)
Mean outcome	18.824	17.718	18.099	19.039	20.509
Observations	85,246	21,259	22,050	21,137	20,800
Distance to nearest hospital	0.055 (0.061)	0.426** (0.188)	-0.216 (0.225)	-0.104 (0.094)	0.105* (0.057)
Mean outcome	4.836	4.135	5.740	5.312	4.128
Observations	84,856	21,259	22,050	20,747	20,800

Notes: Each cell reports the estimated coefficient of $W37$ (an indicator for prematurity) from a different regression. The dependent variable is listed in the row heading and the sample in the column heading. All specifications are local linear regressions using a triangular kernel and include a dummy indicating a preterm delivery and a first-degree polynomial in normalized gestational age, allowed to vary on each side of the cutoff. Samples restricted to observations with gestational age within a 14-day bandwidth around day 259. 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 2: Medical treatments around the week-37 cutoff

	All	Quartile of average household income in postal code			
	(1)	First (2)	Second (3)	Third (4)	Fourth (5)
Obstetrician supervision	36.908*** (3.093)	34.452*** (2.986)	36.949*** (3.129)	39.086*** (3.526)	37.036*** (3.352)
Mean outcome	45.053	45.788	44.857	44.645	44.916
Observations	85,246	21,259	22,050	21,137	20,800
Hospital birth	23.912*** (2.157)	19.101*** (1.456)	24.664*** (2.763)	26.571*** (3.232)	25.509*** (2.110)
Mean outcome	65.975	71.395	64.679	62.831	64.944
Observations	85,246	21,259	22,050	21,137	20,800
NICU admission	10.048*** (0.984)	8.425*** (2.080)	8.355*** (1.254)	11.851*** (0.843)	11.617*** (1.743)
Mean outcome	11.371	11.201	12.501	11.928	9.791
Observations	85,246	21,259	22,050	21,137	20,800
Use of forceps or vacuum	3.245*** (1.056)	1.313 (1.024)	3.666*** (0.740)	3.455 (2.383)	4.262** (1.755)
Mean outcome	13.104	12.152	12.693	13.380	14.238
Observations	85,131	21,230	22,022	21,104	20,775

Notes: Each cell reports the estimated coefficient of $W37$ (an indicator for prematurity) from a different regression. The dependent variable is listed in the row heading and the sample in the column heading. All treatment variables measure the percentage of births receiving that treatment. All specifications are local linear regressions using a triangular kernel and include a first-degree polynomial in normalized gestational age, allowed to vary on each side of the cutoff. Samples restricted to observations with gestational age within a 14-day bandwidth around day 259. 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			
	(1)	First (2)	Second (3)	Third (4)	Fourth (5)
7-day mortality	0.530 (0.359)	-3.257*** (0.716)	3.929 (3.013)	0.849 (1.410)	0.330 (1.501)
Mean outcome	1.518	1.696	1.315	1.328	1.740
Observations	85,246	21,259	22,050	21,137	20,800
28-day mortality	0.900 (0.787)	-2.891** (1.380)	4.638 (3.686)	0.742 (1.383)	0.823 (0.815)
Mean outcome	1.702	1.866	1.425	1.501	2.030
Observations	85,246	21,259	22,050	21,137	20,800
Low Apgar score	-0.246* (0.134)	-0.549* (0.267)	0.110 (0.427)	0.202 (0.282)	-0.770** (0.309)
Mean outcome	0.919	1.161	0.824	0.827	0.866
Observations	85,102	21,224	22,008	21,101	20,769

Notes: See notes in Table 2.

Table 4: Sensitivity to model specification

	All	Quartile of average household income in postal code			
	(1)	First (2)	Second (3)	Third (4)	Fourth (5)
A. Including controls					
7-day mortality	0.432 (0.293)	-3.322*** (0.767)	3.969 (2.859)	0.851 (1.362)	-0.159 (1.562)
Mean outcome	1.518	1.696	1.315	1.328	1.740
Observations	85,246	21,259	22,050	21,137	20,800
28-day mortality	0.786 (0.672)	-2.929* (1.474)	4.702 (3.532)	0.733 (1.330)	0.296 (0.936)
Mean outcome	1.702	1.866	1.425	1.501	2.030
Observations	85,246	21,259	22,050	21,137	20,800
Low Apgar score	-0.242 (0.146)	-0.542* (0.264)	0.157 (0.449)	0.179 (0.284)	-0.886** (0.328)
Mean outcome	0.919	1.161	0.824	0.827	0.866
Observations	85,102	21,224	22,008	21,101	20,769
B. Donut regressions					
7-day mortality	0.517 (0.396)	-3.571*** (0.693)	4.463 (3.020)	0.301 (1.399)	0.576 (1.600)
Mean outcome	1.496	1.696	1.245	1.373	1.680
Observations	82,944	20,677	21,470	20,568	20,229
28-day mortality	0.789 (0.811)	-3.271** (1.346)	5.141 (3.697)	0.149 (1.343)	0.803 (0.935)
Mean outcome	1.686	1.871	1.358	1.552	1.980
Observations	82,944	20,677	21,470	20,568	20,229
Low Apgar score	-0.209 (0.160)	-0.665** (0.266)	0.218 (0.427)	0.091 (0.296)	-0.500* (0.292)
Mean outcome	0.903	1.159	0.805	0.831	0.817
Observations	82,801	20,642	21,429	20,532	20,198
C. Rectangular kernel					
7-day mortality	0.491 (0.561)	-3.940** (1.792)	3.443 (2.463)	0.790 (1.214)	1.449 (1.385)
Mean outcome	1.487	1.614	1.240	1.346	1.760
Observations	96,666	24,150	24,919	24,007	23,590
28-day mortality	0.570 (0.848)	-4.241* (2.376)	3.436 (2.964)	0.918 (1.167)	1.947* (0.983)
Mean outcome	1.659	1.760	1.336	1.495	2.061
Observations	96,666	24,150	24,919	24,007	23,590
Low Apgar score	-0.182 (0.132)	-0.872** (0.364)	-0.047 (0.397)	0.447 (0.321)	-0.259 (0.317)
Mean outcome	0.890	1.073	0.798	0.849	0.841
Observations	96,504	24,107	24,872	23,970	23,555

Table 4: Sensitivity to model specification (cont'd)

	All	Quartile of average household income in postal code			
	(1)	First (2)	Second (3)	Third (4)	Fourth (5)
D. 7-day bandwidth					
7-day mortality	0.893** (0.393)	-3.601*** (1.086)	3.133 (4.022)	3.543*** (1.100)	0.155 (1.938)
Mean outcome	2.086	2.194	2.149	1.319	2.688
Observations	30,576	7,618	7,883	7,644	7,431
28-day mortality	1.937* (0.910)	-1.513 (1.584)	4.095 (4.958)	3.189** (1.265)	1.671** (0.751)
Mean outcome	2.365	2.377	2.328	1.507	3.264
Observations	30,576	7,618	7,883	7,644	7,431
Low Apgar score	-0.588*** (0.108)	-0.293 (0.294)	-0.545 (0.437)	-0.198 (0.338)	-1.313*** (0.332)
Mean outcome	1.114	1.317	0.969	0.944	1.230
Observations	30,530	7,610	7,868	7,630	7,422
E. 21-day bandwidth					
7-day mortality	0.290 (0.415)	-3.304*** (0.952)	3.656 (2.411)	0.179 (1.350)	0.429 (1.311)
Mean outcome	1.250	1.369	1.096	1.186	1.356
Observations	188081	46,675	48,343	46,719	46,344
28-day mortality	0.528 (0.740)	-3.322** (1.423)	3.994 (2.933)	0.216 (1.294)	1.034 (0.800)
Mean outcome	1.374	1.534	1.165	1.281	1.522
Observations	188081	46,675	48,343	46,719	46,344
Low Apgar score	-0.121 (0.129)	-0.738** (0.274)	0.264 (0.377)	0.281 (0.264)	-0.307 (0.271)
Mean outcome	0.878	1.036	0.803	0.855	0.819
Observations	187777	46,585	48,271	46,650	46,271

Notes: See notes in Table 2.

Table 5: Robustness to the definition of analysis sample

	All	Quartile of average household income in postal code			
	(1)	First (2)	Second (3)	Third (4)	Fourth (5)
A. Under the supervision of a midwife at least until gestational day 245					
7-day mortality	0.723** (0.341)	-3.126*** (0.635)	4.093 (3.072)	1.100 (1.405)	0.571 (1.528)
Mean outcome	1.505	1.717	1.273	1.341	1.697
Observations	83,902	20,886	21,721	20,782	20,513
28-day mortality	1.095 (0.781)	-2.762** (1.323)	4.823 (3.764)	0.979 (1.393)	1.072 (0.779)
Mean outcome	1.691	1.889	1.384	1.516	1.989
Observations	83,902	20,886	21,721	20,782	20,513
Low Apgar score	-0.191 (0.141)	-0.490* (0.252)	0.214 (0.453)	0.216 (0.284)	-0.724** (0.304)
Mean outcome	0.917	1.169	0.821	0.812	0.867
Observations	83,760	20,851	21,681	20,746	20,482
B. Under the supervision of a midwife at least until gestational day 231					
7-day mortality	0.614* (0.338)	-3.211*** (0.705)	3.908 (2.972)	1.192 (1.345)	0.311 (1.478)
Mean outcome	1.522	1.685	1.307	1.319	1.787
Observations	85,991	21,455	22,255	21,317	20,964
28-day mortality	0.981 (0.752)	-2.851** (1.355)	4.608 (3.635)	1.090 (1.322)	0.803 (0.811)
Mean outcome	1.705	1.854	1.416	1.491	2.076
Observations	85,991	21,455	22,255	21,317	20,964
Low Apgar score	-0.254 (0.149)	-0.624** (0.293)	0.020 (0.458)	0.217 (0.301)	-0.645* (0.333)
Mean outcome	0.922	1.165	0.835	0.822	0.866
Observations	85,847	21,420	22,213	21,281	20,933

Table 5: Robustness to the definition of analysis sample (cont'd)

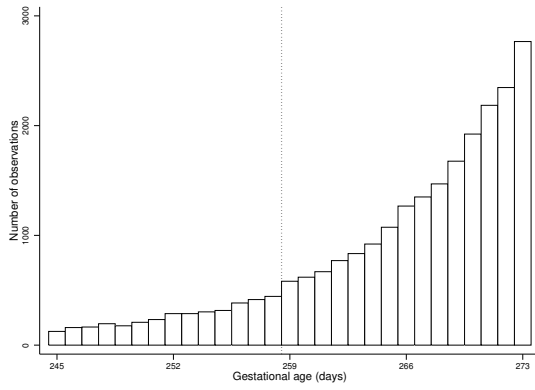
	All	Quartile of average household income in postal code			
	(1)	First (2)	Second (3)	Third (4)	Fourth (5)
C. Including spontaneous natural births with unknown referral date					
7-day mortality	0.739*	-2.746**	4.635*	0.536	0.282
	(0.410)	(1.236)	(2.378)	(1.440)	(1.461)
Mean outcome	1.548	1.672	1.421	1.387	1.714
Observations	89,065	22,289	22,954	21,982	21,840
28-day mortality	0.900	-2.400	4.868	0.130	0.739
	(0.712)	(2.040)	(3.102)	(1.573)	(0.772)
Mean outcome	1.751	1.834	1.579	1.609	1.990
Observations	89,065	22,289	22,954	21,982	21,840
Low Apgar score	-0.284**	-0.681**	0.230	0.128	-0.837**
	(0.135)	(0.288)	(0.343)	(0.292)	(0.308)
Mean outcome	0.936	1.151	0.860	0.844	0.886
Observations	88,916	22,251	22,911	21,946	21,808
D. Including C-sections					
7-day mortality	-0.025	-4.073***	4.340	-0.633	-0.040
	(0.480)	(0.643)	(3.245)	(1.124)	(1.390)
Mean outcome	1.608	1.800	1.505	1.446	1.688
Observations	106,576	26,384	27,669	26,494	26,029
28-day mortality	0.248	-4.044***	4.866	-0.693	0.547
	(0.772)	(1.428)	(3.774)	(1.113)	(0.875)
Mean outcome	1.798	2.070	1.591	1.626	1.916
Observations	106,576	26,384	27,669	26,494	26,029
Low apgar score	-0.264**	-0.313	0.019	-0.016	-0.768**
	(0.125)	(0.299)	(0.537)	(0.318)	(0.358)
Mean outcome	0.997	1.176	0.914	0.910	0.991
Observations	106,576	26,341	27,618	26,449	25,995

Notes: See notes in Table 2.

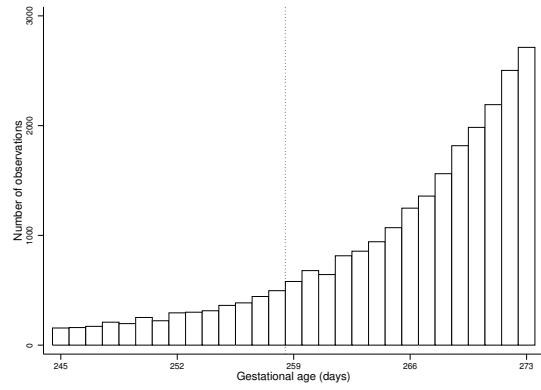
Table 6: Robustness to alternative income bins

	Tercile of average household income in postal code			Quintile of average household income in postal code				
	First (1)	Second (2)	Third (3)	First (4)	Second (5)	Third (6)	Fourth (7)	Fifth (8)
7-day mortality	-2.165** (0.884)	3.570 (2.509)	0.139 (1.572)	-3.406*** (0.739)	0.843 (1.787)	4.241 (3.867)	1.429 (2.955)	-0.708 (1.234)
Mean outcome	1.559	1.218	1.787	1.675	1.462	1.258	1.661	1.538
Observations	28,573	28,826	27,847	17,185	17,356	17,334	16,951	16,420
28-day mortality	-1.908* (0.957)	4.009 (3.035)	0.520 (0.831)	-2.930* (1.597)	0.738 (1.793)	5.050 (4.734)	1.003 (3.004)	0.402 (1.491)
Mean outcome	1.685	1.386	2.048	1.885	1.601	1.328	2.022	1.684
Observations	28,573	28,826	27,847	17,185	17,356	17,334	16,951	16,420
Low apgar score	-0.550*** (0.156)	0.107 (0.315)	-0.306 (0.296)	-0.446 (0.277)	-0.230 (0.537)	0.172 (0.420)	0.159 (0.489)	-0.915** (0.360)
Mean outcome	1.055	0.820	0.882	1.189	0.927	0.771	0.825	0.880
Observations	28,520	28,778	27,804	17,155	17,326	17,302	16,921	16,398

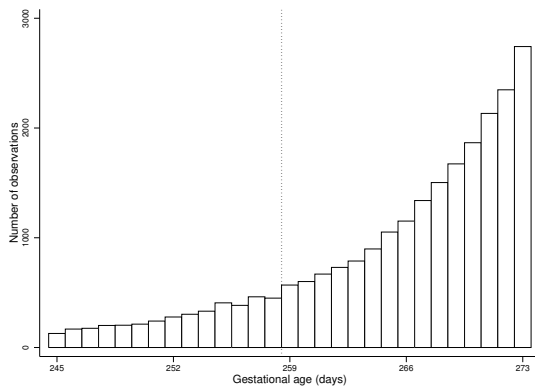
Notes: See notes in Table 2.



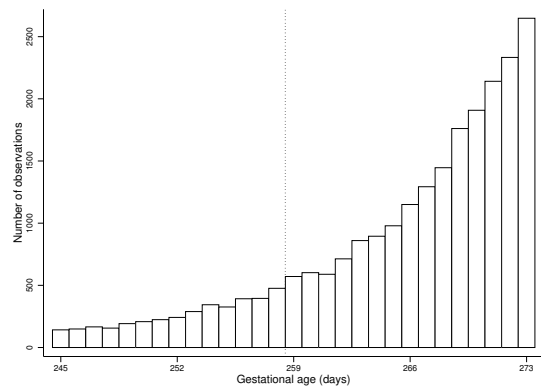
(a) First quartile



(b) Second quartile



(c) Third quartile



(d) Fourth quartile

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

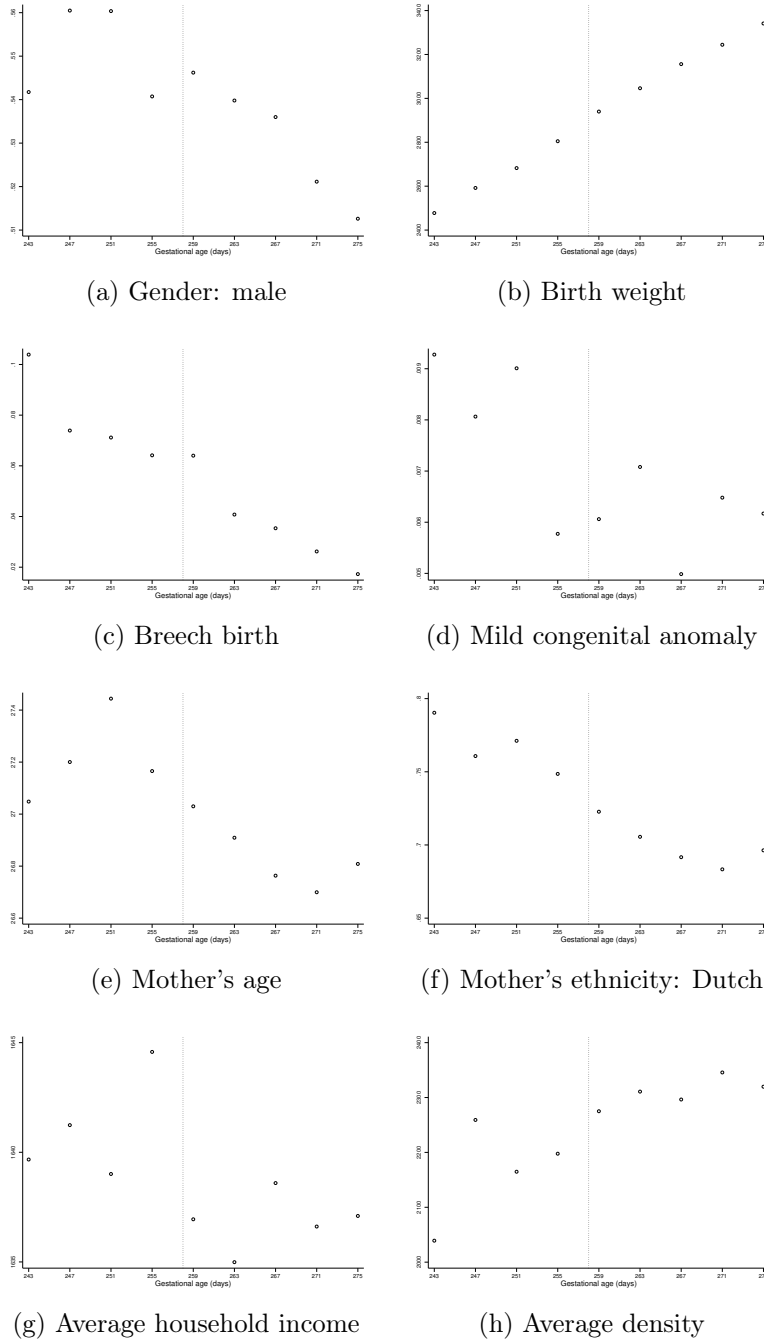
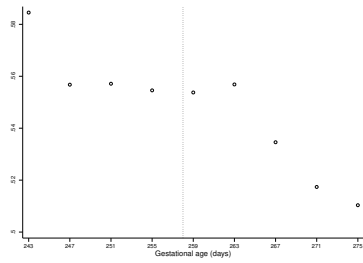
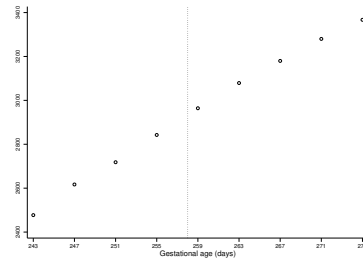


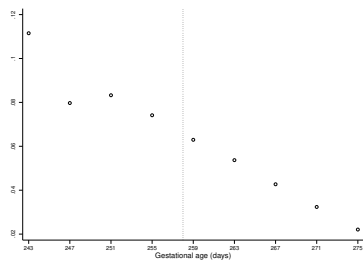
Figure A2: Distribution of selected covariates around the week-37 cutoff, first income quartile



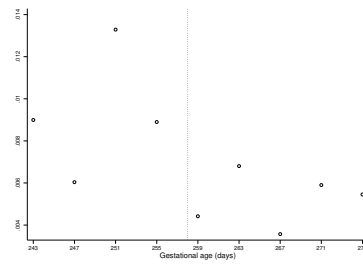
(a) Gender: male



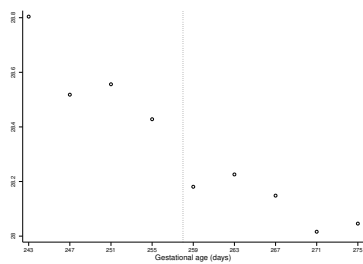
(b) Birth weight



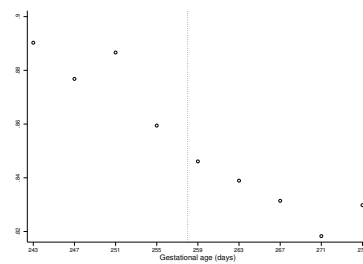
(c) Breech birth



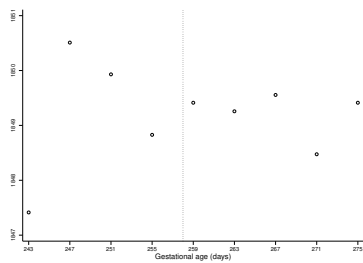
(d) Mild congenital anomaly



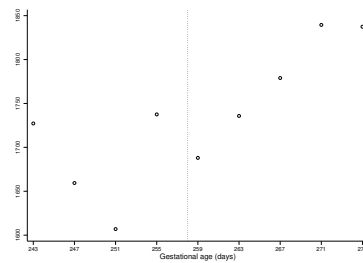
(e) Mother's age



(f) Mother's ethnicity: Dutch

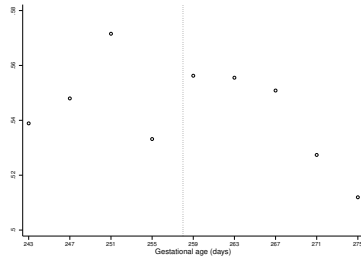


(g) Average household income

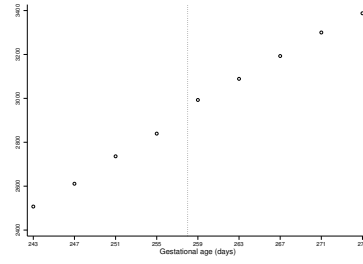


(h) Average density

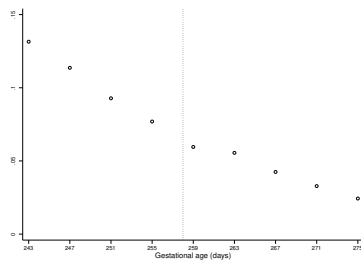
Figure A3: Distribution of selected covariates around the week-37 cutoff, second income quartile



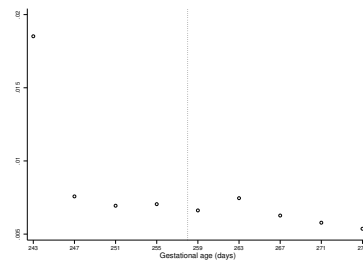
(a) Gender: male



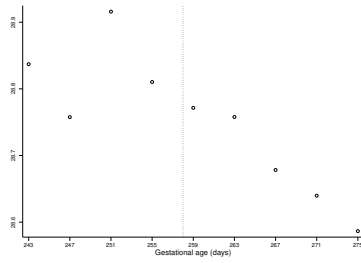
(b) Birth weight



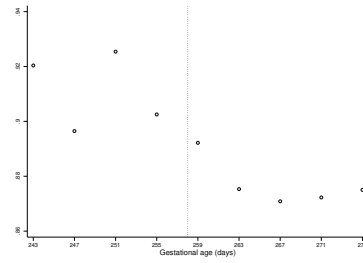
(c) Breech birth



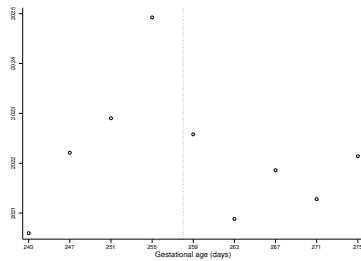
(d) Mild congenital anomaly



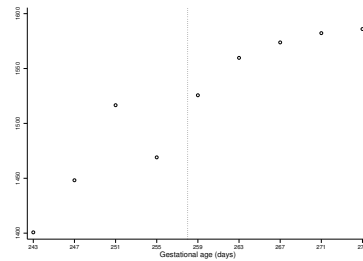
(e) Mother's age



(f) Mother's ethnicity: Dutch

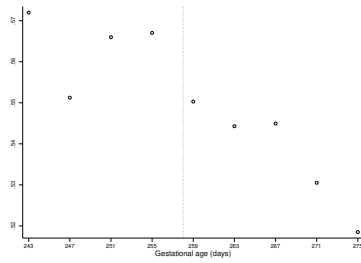


(g) Average household income

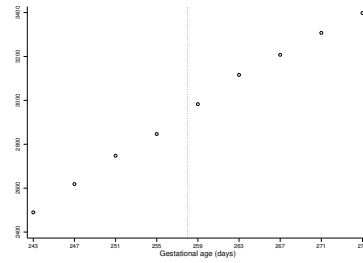


(h) Average density

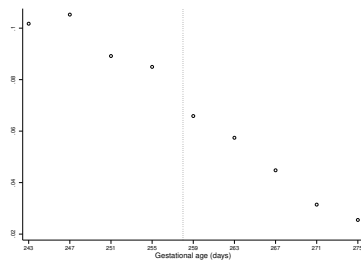
Figure A4: Distribution of selected covariates around the week-37 cutoff, third income quartile



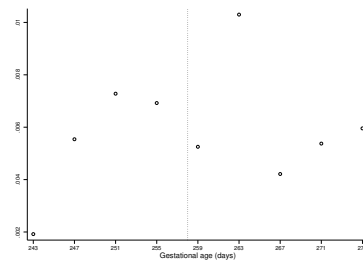
(a) Gender: male



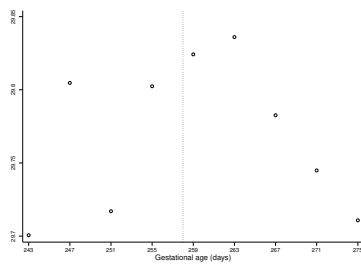
(b) Birth weight



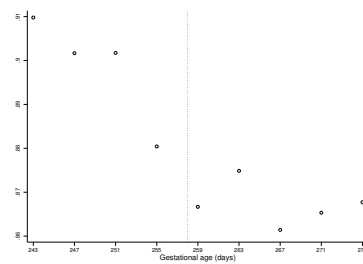
(c) Breech birth



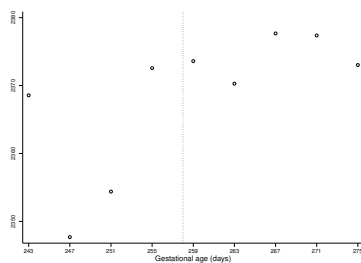
(d) Mild congenital anomaly



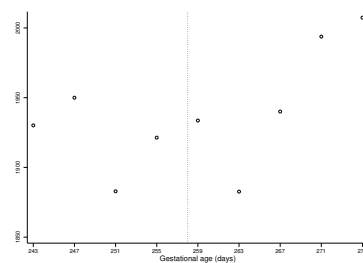
(e) Mother's age



(f) Mother's ethnicity: Dutch

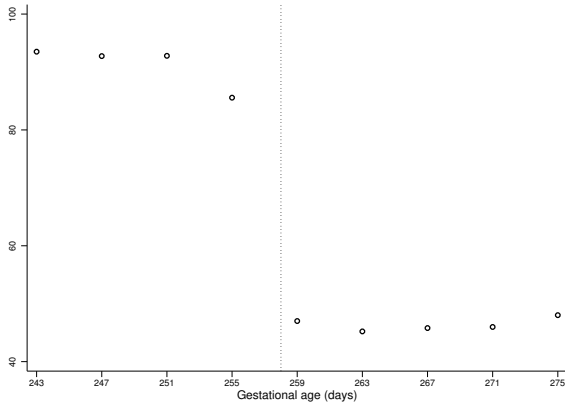


(g) Average household income

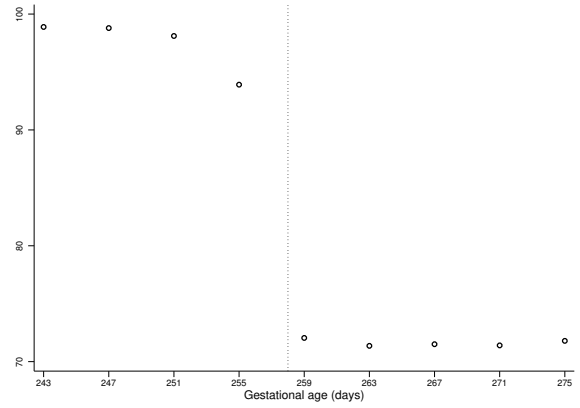


(h) Average density

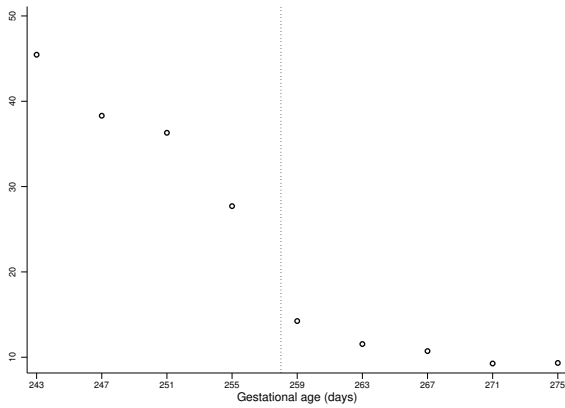
Figure A5: Distribution of selected covariates around the week-37 cutoff, fourth income quartile



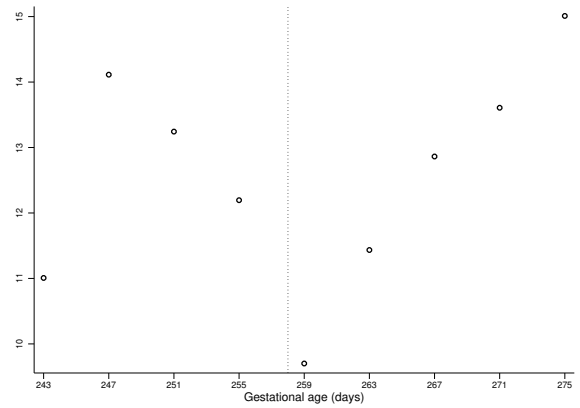
(a) Obstetrician supervision



(b) Hospital birth

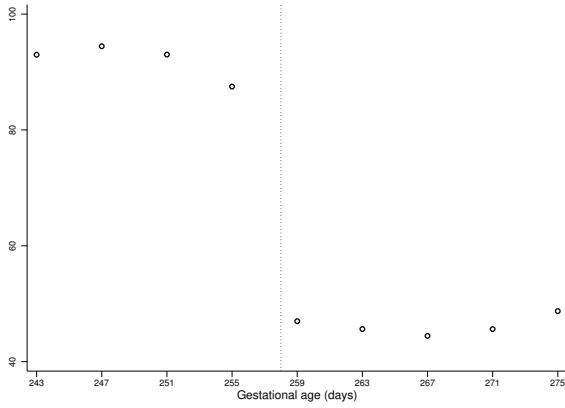


(c) NICU admission

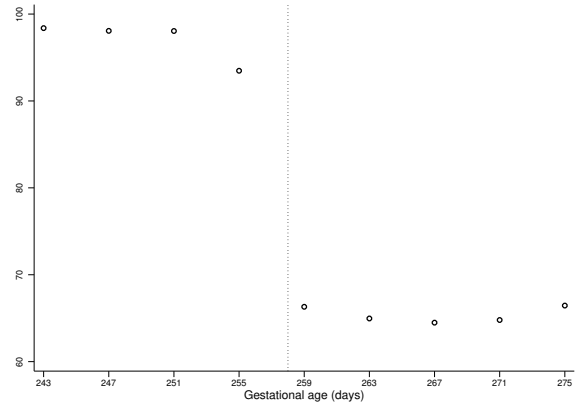


(d) Use of forceps/vacuum

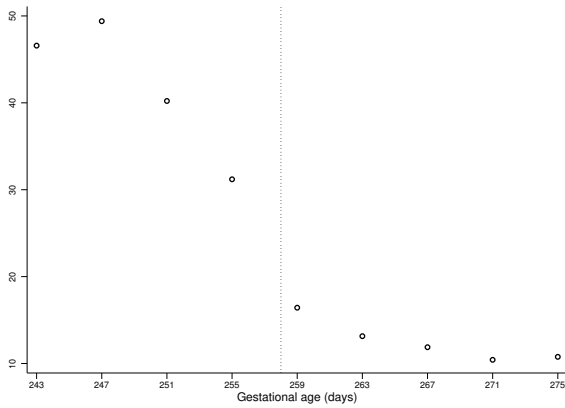
Figure A6: Medical treatments around the week-37 cutoff, first income quartile



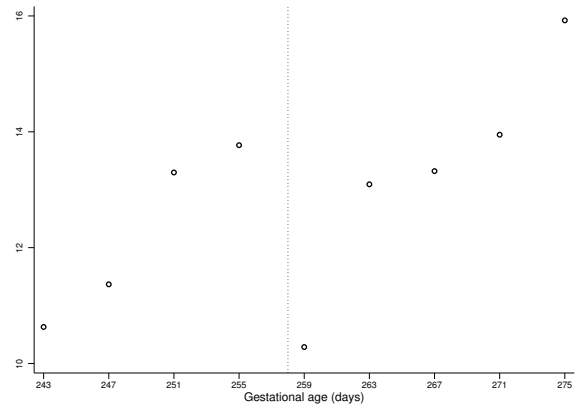
(a) Obstetrician supervision



(b) Hospital birth

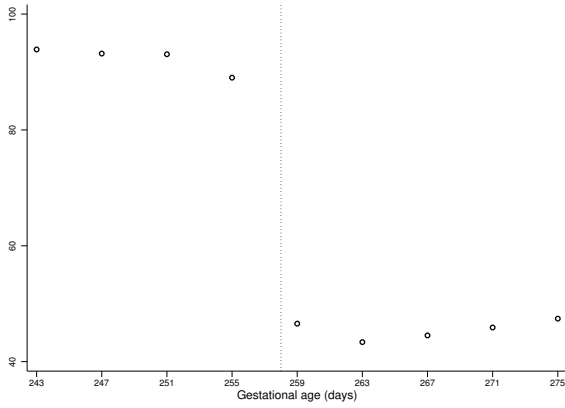


(c) NICU admission

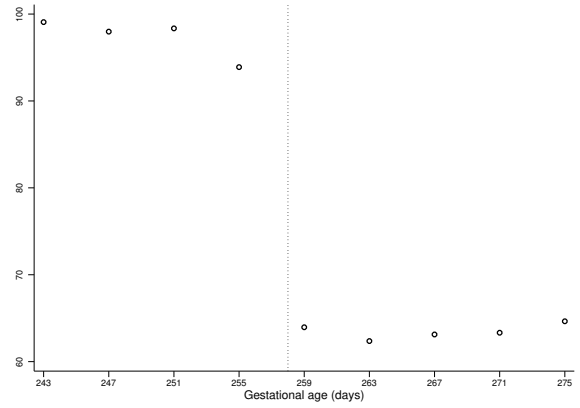


(d) Use of forceps/vacuum

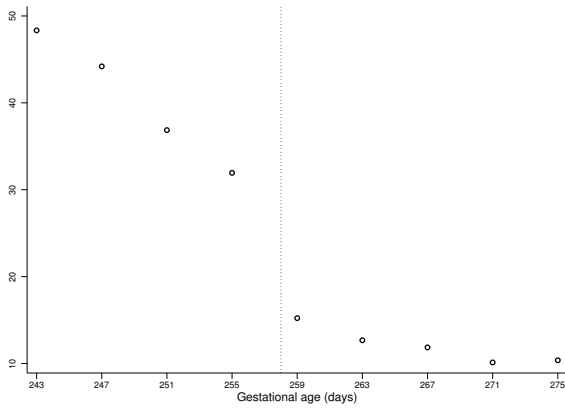
Figure A7: Medical treatments around the week-37 cutoff, second income quartile



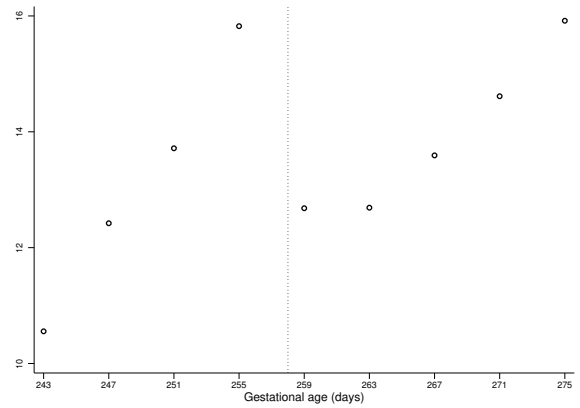
(a) Obstetrician supervision



(b) Hospital birth

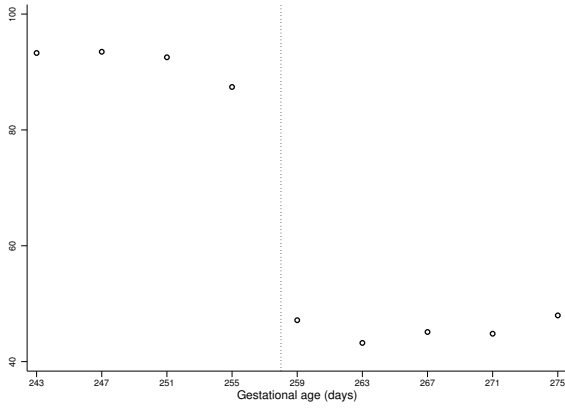


(c) NICU admission

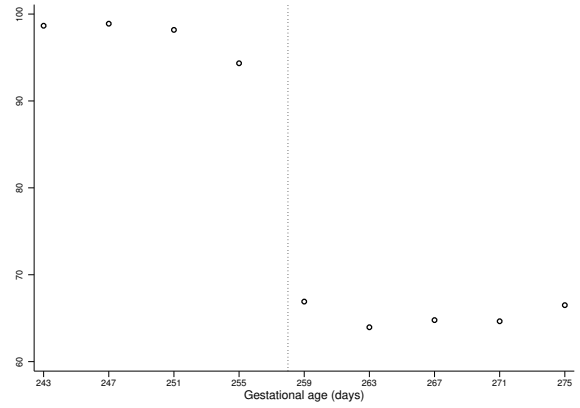


(d) Use of forceps/vacuum

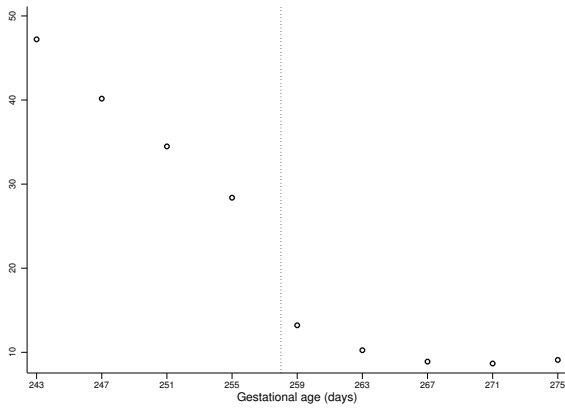
Figure A8: Medical treatments around the week-37 cutoff, third income quartile



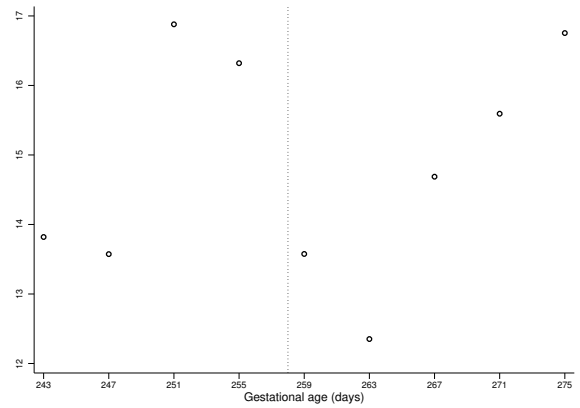
(a) Obstetrician supervision



(b) Hospital birth

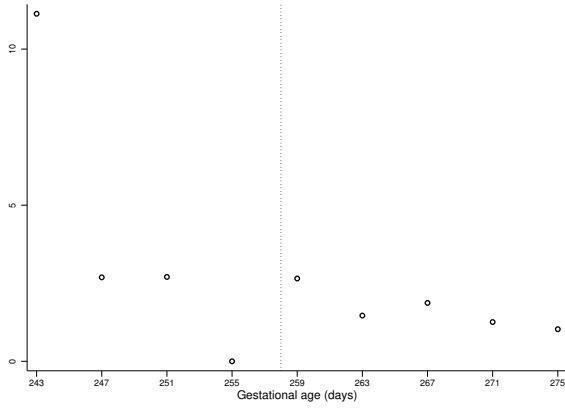


(c) NICU admission

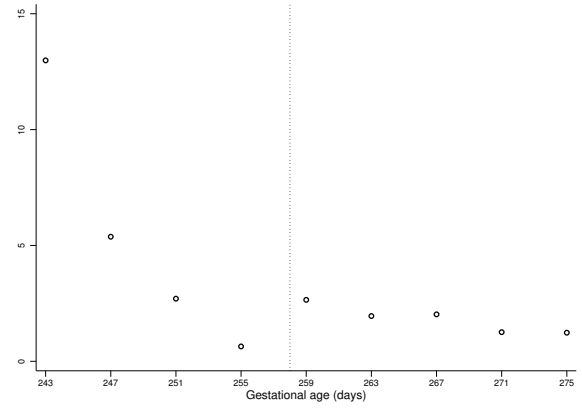


(d) Use of forceps/vacuum

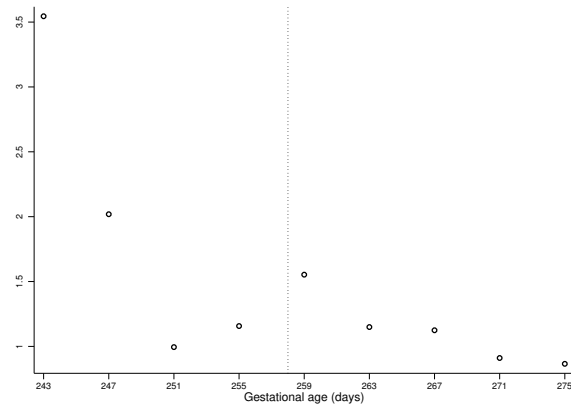
Figure A9: Medical treatments around the week-37 cutoff, fourth income quartile



(a) 7-day mortality

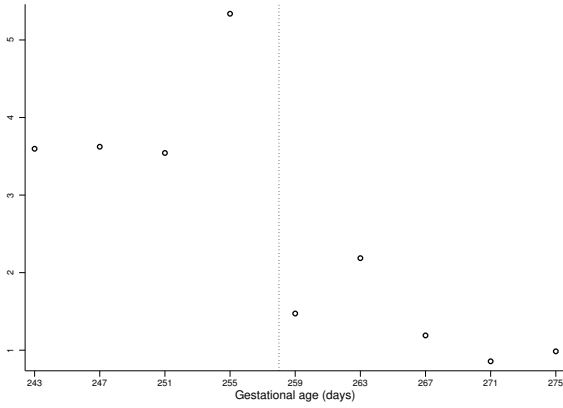


(b) 28-day mortality

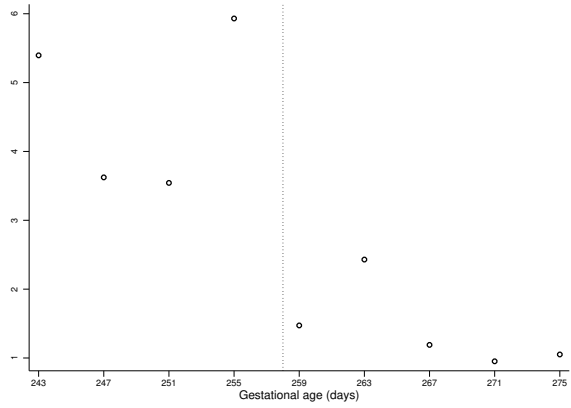


(c) Low Apgar score

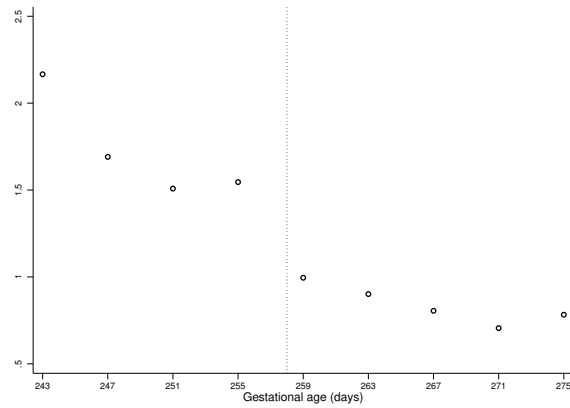
Figure A10: Newborn health around the week-37 cutoff, first income quartile



(a) 7-day mortality

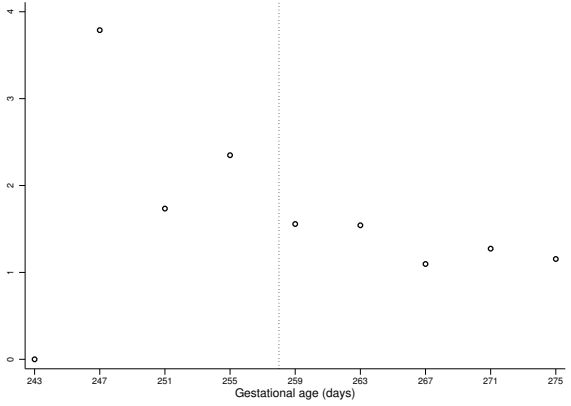


(b) 28-day mortality

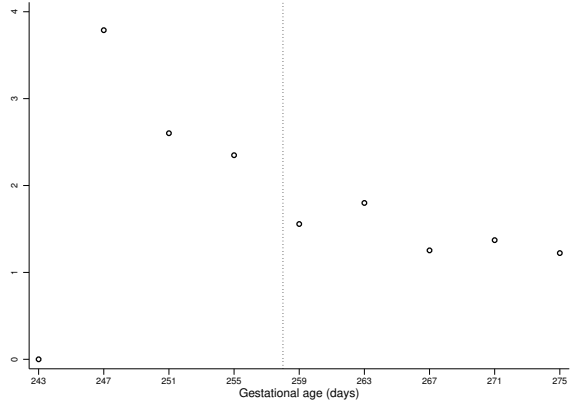


(c) Low Apgar score

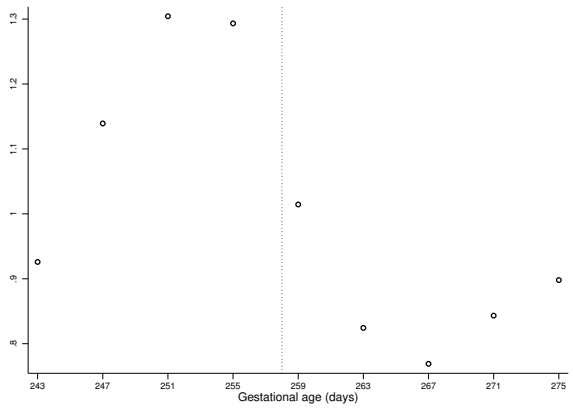
Figure A11: Newborn health around the week-37 cutoff, second income quartile



(a) 7-day mortality

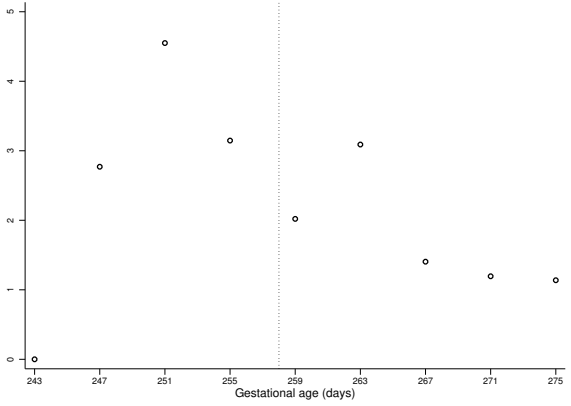


(b) 28-day mortality

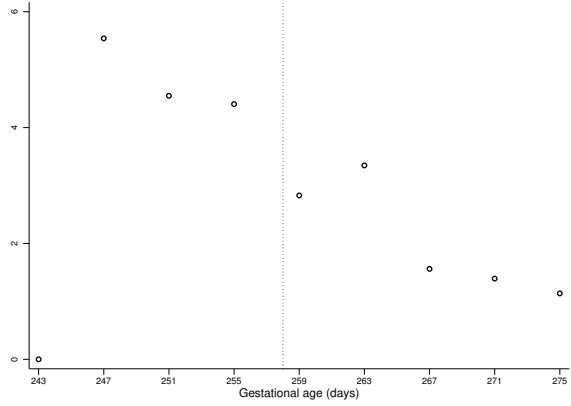


(c) Low Apgar score

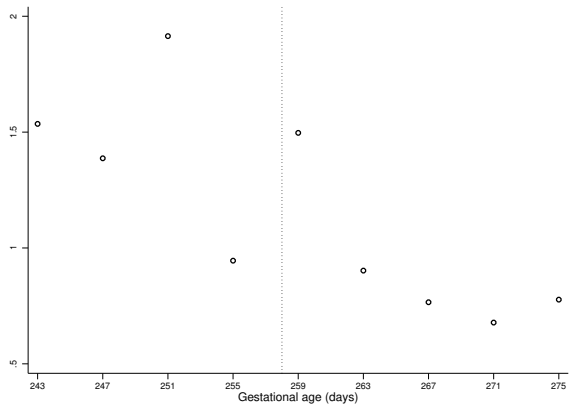
Figure A12: Newborn health around the week-37 cutoff, third income quartile



(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	15.23	16.02	17.88	15.72	13.71
28-day mortality	14.88	18.34	18.27	12.86	14.60
Low apgar score	16.83	19.18	20.58	21.45	16.70
B. Treatments					
Obstetrician supervision	7.89	10.08	10.50	10.43	10.53
Hospital birth	9.56	12.74	12.44	12.47	12.59
NICU admission	9.85	13.10	11.77	13.64	13.56
Use of forceps or vacuum	15.75	17.52	22.32	19.29	20.00

Notes: Each cell provides the calculated optimal bandwidth corresponding to the outcome in the row and the sample in the column heading. The optimal bandwidths are calculated using a rule-of-thumb approach. See section 3 for details.

Table A2: Sample construction

	Observations
Initial sample (universe of births)	1,630,062
Stillbirths	-9,263
Missing information on birth attendant	-3,377
Missing referral date	-48,189
Planned C-section	-107,162
Induced/stimulated birth	-233,400
Live spontaneous births	1,228,671
High-risk births	-282,134
Referral prior to gestational day 238	-46,937
Multiple births	-420
Birth occurred outside bandwidth (245–272 gestational days)	-737,347
Low-risk births	161,833
Higher-order birth	-76,587
Low-risk first-births (analysis sample)	85,246

Table A3: Calculation of the cost of a life saved

	Quartile of average household income in postal code				Overall
	First (1)	Second (2)	Third (3)	Fourth (4)	
	<i>Estimated number of additional treatments</i>				<i>Cost per birth</i>
Transfer from hospital+midwife	153.51	122.85	125.15	115.27	1,141
Transfer from home+midwife	191.01	246.64	265.71	255.09	1,730
NICU admission	84.25	83.55	118.51	116.17	9,152
Use of forceps or vacuum	13.13	36.66	34.55	42.62	465
Observations	21,259	22,050	21,137	20,800	
	<i>Estimated additional costs per 1,000 births</i>				<i>Cost per 1,000 births</i>
Transfer from hospital+midwife	175,094	140,123	142,746	131,477	147,385
Transfer from home+midwife	330,524	426,786	459,785	441,408	414,529
NICU admission	771,038	764,632	1,084,579	1,063,163	918,403
Use of forceps or vacuum	6,105	17,047	16,066	19,818	14,751
Total costs per 1,000 births	1,282,761				1,495,068
Number of lives saved (7-day mortality)	3,257				0.812
Number of lives saved (28-day mortality)	2,891				0.721
Cost per life saved (7-day mortality):	393,847				1,840,664
Cost per life saved (28-day mortality)	443,708				2,073,692

Notes: All prices are in 2016 Euros.