Low-risk isn't no-risk: Perinatal treatments and the health of low-income 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

January 2019

Abstract

We investigate the effects of perinatal medical treatments on low-income newborns who are classified as low-risk. A policy rule in The Netherlands states that low-risk deliveries before week 37 should be supervised by physicians and later deliveries only by midwives with no physician present. This creates large discontinuities in the probability of receiving medical interventions only physicians are allowed to perform. Using a regression discontinuity design, we find that babies born slightly before the week-37 cutoff are significantly less likely to die than babies born slightly later. Our data suggest that physician supervision of birth reduces the likelihood of adverse events such as fetal distress or emergency C-section. Our results indicate that low-income women benefit from receiving a higher level of medical care even if no explicit risk factors have been recognized, pointing to challenges in identifying all high-risk pregnancies. "Back-of-the-envelope" calculations suggest this additional care is highly cost-effective.

Keywords: Perinatal care, prematurity, mortality, midwives, birth, medical treatments, medical interventions

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 Simeonova, Kosali Simon, Arthur van Soest, and participants at Aarhus University, Cornell University, Impaq International, Lund University, Stockholm School of Economics, Uppsala University, University of Copenhagen, University of Hamburg, University of Mainz, University of Munich, University of Southern Denmark, Universite de Sherbrooke, University of York, Tilburg University, and the meetings of the American Society of Health Economists, American Health Economics Workshop, European Economic Association, EuroEpi, European Workshop on Econometrics and Health Economics, iHEA-ECHE, 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

Medical expenditures increased tremendously over the last few decades throughout the entire developed world. In the United States, health care spending in 2016 reached almost 18% of gross domestic product. Most of this increase is attributed to changes in medical technologies (Newhouse, 1992). However, there is substantial heterogeneity in treatment effects across patients (Chandra and Skinner, 2012). As such, decision-makers are debating the possibility of slowing down the growth of medical expenditures by reducing the use of inefficient technologies. One specific policy focus concerns the technology of childbirth and proposes the expanded use of lower-cost physician extenders instead of higher-cost physicians for low-risk patients (Fuchs, 1998). In this paper, we investigate the impact of perinatal medical care on the short-term health of low-income low-risk newborns.

Empirical estimation of the returns to medical treatments 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 care for low-risk deliveries. 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 the medical professional supervising the delivery (OB/GYN instead of midwife) and possibly also in two other important medical inputs: 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).

We investigate the effect of the week-37 rule using a regression discontinuity (RD) design among the sample of low-income mothers, defined as women residing in postal codes in the lowest quartile of average monthly household income.[1] There are three reasons for our focus on low-income women. First, the week-37 rule may be more relevant for them because the classification as low- or high-risk may be more error-prone among low socioeconomic status (SES) women. In general, the week-37 rule is most relevant for women (and deliveries) with unrecognized risks: a premature delivery might be life-saving because it will make them end up with the care that they need because of the unrecognized risk. If higher-income women have better health education, are better able to communicate with midwives, and follow their instructions, midwives might be better able to identify higher-income mothers at risk and refer them to an OB/GYN. The week-37 rule would then be of little benefit for them. In contrast, higher rates of unrecognized risks among lower-income women might mean that the extra care provided under the week-37 rule can be beneficial in general, i.e., that low SES may in itself be regarded as a proxy for risk factors. Consistent with this, 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.

Second, it is a stylized fact that there are significant socioeconomic disparities in the prevalence of prematurity (Institute of Medicine, 2007). Given the increased burden of prematurity on families with lower socioeconomic sta-

[1] 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.

tus, studying the returns to perinatal medical interventions for this group is interesting in its own right.

Finally, the key identification assumption in the RD design requires that individuals do not have precise control over the assignment variable. We exclude induced and stimulated births from our sample, as well as planned C-sections, which are rare within our sample.[2] Hence, a manipulation of birth dates with the goal of ending up on a specific side of the cutoff is unlikely in our sample of spontaneous births. Nevertheless, as we will argue later on, physicians may induce births differentially across cutoff and thus excluding induced and stimulated births may lead to a bias toward zero, especially among higher-income women. Correspondingly, we find evidence that the key identification assumption is potentially violated among higher-income births. While the identification assumptions in this RD design are more likely to hold among low-income mothers, it is important to emphasize that they are ultimately not directly testable. Therefore, we devote a significant portion of the results section to discussing their plausibility.

To preview our results, we find that the week-37 rule generates substantial variation in medical care parameters among low-income women with pregnancies classified as low-risk. For example, the probability that a birth is supervised by an obstetrician increases by 28.4 percentage points below the 37-week threshold. Similarly, newborns slightly below the week-37 cutoff are 16.4 percentage points more likely be delivered in a hospital and 4.5 percentage points more likely to be admitted to a neonatal intensive care unit (NICU). These estimates are economically large and correspond to increases of 23–62% when compared to the mean above the cutoff. Our results also indicate economically large health gains to preterm newborns resulting from the week-37 rule: babies born slightly before 37 completed gestational weeks are significantly less likely to die when compared to newborns who are slightly above the week-37 cutoff.

[2] The rate of planned C-sections is generally very low in The Netherlands. 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.

In order to shed some light on the mechanisms behind these health gains, we further examine the presence of discontinuities among a number of health outcomes that are related to the skill of the birth attendant. Our results point to economically large reductions in the likelihood of presence of meconium in amniotic fluid, emergency C-section or birth trauma/perinatal asphyxia (oxygen deprivation). Finally, our "back-of-the-envelope" calculations suggest that the week-37 rule is a highly cost-effective measure. This is especially important in light of the growing emphasis on cost reduction through increased use of physician extenders (Institute of Medicine, 2011).

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; Daysal et al., 2018). Research on the returns to medical care 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.[3] 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 authors 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 makes several contributions. First, we exploit a new source of

[3]Some 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.

variation in perinatal medical treatments that affects children on the borderline of prematurity (37 weeks of gestation). Prematurity is the most important cause of death within the first month of life and it ranks as the second cause of all deaths for children under 5 (March of Dimes et al., 2012). Focusing on the understudied group of premature children on the borderline of prematurity 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 additional medical care 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.

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 are 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.[4]

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.

[4]During 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.

3 Empirical Strategy

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

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

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

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

where the unit of observation is infant i born in year t at gestational age a, Y_{iat} is a measure of infant health or of medical treatments, $W37_a$ is an indicator for prematurity (gestational age strictly below 37 completed weeks, or 259 days), and $f(\cdot)$ is a first-degree polynomial in normalized gestational age that is allowed to differ 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 start by calculating a rule-of-thumb bandwidth following 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, which are generally between 10–20 days. We choose to be more conservative in our baseline regressions and use a bandwidth of 7 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. It is constructed by linking individual birth records submitted by midwives (LVR-1), obstetricians/gynecologists (LVR-2) and pediatricians

(LNR).[5]

The data include detailed information on the birth process. For each delivery, we observe the date and time of birth, type of birth attendant (midwife or OB/GYN), delivery location (home or hospital), method of delivery (vaginal, planned C-section, emergency C-section), use of interventions during vaginal delivery (labor augmentation, induction, use of forceps or vacuum), as well as the presence of complications during pregnancy or delivery. In the case of complications, we can observe the date and the reason for referral from midwife to an obstetrician. The data also provide rich background information on newborns (gender, gestational age in days, birth weight, parity, plurality) and basic demographic characteristics of mothers (age, ethnicity, 4-digit residential postal code). We complement the individual-level Perined data with a secondary postal code-level data set from Statistics Netherlands (Kerncijfers postcodegebieden 2004). These data provide a snapshot of average characteristics in the postal code of residence of the mother as of January 1, 2004, such as average monthly household income, average area density, and the share of residents 0-15 years old.[6]

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 and 28-day mortality.[7]

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

[5]Perined data does not include information on births supervised by general practitioners, a very small share of all primary care deliveries.

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

[7]We do not have information on longer term mortality rates.

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.[8]

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.[9] The third group (infant characteristics) includes birth weight and indicators for gender, congenital anomalies and birth position.[10] 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.[11]

Our analysis sample includes live deliveries by low-risk women with gestational age between 252 and 265 days. 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; Kooy et al., 2011). We focus on low-risk women because the week-37 rule does not apply to high-risk women.[12] This has the added benefit that women in this category are homogenous in terms of their prenatal care. As a result,

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

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

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

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

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

we are able to identify the effects of perinatal medical care abstracting from the effects of prenatal care. In addition, we exclude multiple births, which are automatically referred to obstetricians.

Focusing on spontaneous births by low-risk women could lead to a violation of the RD assumptions if certain types of women are more likely to be referred to an OB/GYN (i.e., be classified as high-risk) before the cut-off. Referrals under the week-37 rule can be made because of premature onset of labor, but also because of the "threat of prematurity," which midwives can potentially assess relatively shortly 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 take women's health and preferences into account, then referral patterns on each side of the threshold may be different and a comparison of low-risk (i.e., non-referred) births right above and right below the threshold would be misleading. We address this potential bias by defining low-risk women as women who were not referred to an obstetrician by gestational age of 238 days, 14 days before the lower bound of our target interval.[13] It is currently impossible to predict with any degree of reliability so much in advance whether a spontaneous delivery will occur just on the left, or just on the right of the cutoff (Institute of Medicine, 2007). Therefore, midwives are unable to selectively refer certain types of women to an obstetrician based on whether they are expected to deliver before or after the week-37 cutoff.

Including women who are referred to an OB/GYN between gestational age of 238 days and the onset of labor implies that there may be cases in which gestational age can be manipulated through planned C-sections, induced and stimulated births.[14] We exclude these observations from our sample. However, selection could still occur if obstetricians induce births differentially across the cutoff among this group of women (recall that only doctors can induce births,

[13]We thank Gordon Dahl for this suggestion. The date of referral is missing for less than 5 percent of our analysis sample. We exclude these observations from the main analyses.

[14]Note that these interventions can only be performed by physicians to women already in their care, meaning that they cannot be used strategically so that a woman ends up with the perinatal care associated with a particular side of the cutoff. Nevertheless, we exclude inductions, stimulated births and planned C-sections from our sample.

midwives cannot). Medical guidelines state that induction should be considered when there is a clear clinical justification. For uncomplicated pregnancies, the main reason for induced births is to avoid risks associated with continuing the pregnancy. More relevant to our setup, induction can also be medically justified when other risk factors occur, such as when the rupture of membranes is not followed by the onset of labor within 24–48 hours (National Institute for Health and Clinical Excellence, 2008). At the same time, medical guidelines are generally more restrictive for inductions before 37 completed weeks than for at term births (National Institute for Health and Care Excellence, 2008; Nederlandse Vereniging voor Obstetrie en Gynaecologie, 2006).[15] Given these considerations, we would expect the risk composition of births to vary across the cutoff (and potentially across income groups) as medical professionals should be less likely to induce births to the left of the cutoff. We indeed find that a birth just before the week-37 cutoff is less likely to be induced than a slightly later birth, especially among higher-income women (see Appendix Table A2). This type of sample selection would remove from our sample some births that are more likely to have adverse effects, more so to the right of the cutoff than to the left, leading to a bias toward zero in our results. To address this, we restrict our sample to low-income women, among whom this pattern is considerably less strong. Low-income women are defined as women residing in postal codes in the lowest quartile of average monthly household income.[16]

Finally, we only consider 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.[17] This results in an analysis sample of 7,618 low-income

[15] For example, dinoprostone is the most important medicine for inducing labor in women with an unripe cervix (Nederlandse Vereniging voor Obstetrie en Gynaecologie, 2006). Instructions for the use of dinoprostone explicitly state that its use is indicated for (close to) full-term pregnancies. Physicians would need to deviate from these instructions when using it to induce a premature birth, which they would do only after a careful consideration of pros and cons.

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

[17] Indeed, we find that parity is discontinuous across the week-37 cutoff among all low-risk

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 252 and 265.

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. Given that the moment of a spontaneous delivery cannot be reliably influenced, the women in our sample cannot precisely delay or hasten their birth to ensure a certain level of care. 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 whether this happens, 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 does not seem to be a significant jump in the number of births between day 258, when the week-37 rule applies, and day 259, when it does not. We also conduct a more formal check for the statistical significance of the difference in the number of births at the cutoff. Given that our running variable is discrete, we apply the method in Frandsen (2017). We indeed do not find evidence of a statistically significant discontinuity in the number of births for a wide range of plausible values for the k parameter.[18]

Next, we check whether there are differences in observable characteristics across the week-37 cutoff. In the absence of manipulation of birth date or

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).

[18] The p-value is greater than 0.10 for $k \geq 0.02$, which is at the lower end of the range of values suggested by Frandsen (2017).

differential selection into our sample across the week-37 cutoff, the observable characteristics should be locally balanced on both sides of the cutoff. Figure 2 presents the means of selected covariates by gestational age before and after the cutoff (for visual clarity, here and in the rest of the paper, we group the data in 4-day bins starting from the cutoff). The Figure shows that the distribution of the covariates is smooth around the discontinuity.

In order to examine this issue more formally, we estimate in Table 1 local-linear regressions using the covariates as the dependent variables. The results suggest that observations just below the week-37 cutoff are similar to those just above in terms of a range of important maternal and newborn characteristics. In cases where we find statistically significant differences, the direction of potential selection does not have a clear pattern. For example, infants born before day 259 are on average 45.6 grams lighter than those born after the cutoff.[19] On the other hand, mothers of preterm infants reside in less densely populated postal codes with slightly higher average monthly household income. In order to check whether these differences matter in the specific context of our outcomes, we check if predicted outcomes based on observable characteristics are smooth across the cutoff. In particular, we predict 7-day and 28-day mortality using a probit model including the full set of control variables. As the last panel of the Table shows, there is no significant discontinuity in predicted mortality across the cutoff.[20]

[19]The small statistically significant jump in birth weight is not surprising because birth weight and gestational age are particularly related to each other. Almond et al. (2010) exploit the variation in medical inputs across the very low birth weight threshold to estimate the marginal returns to medical care and also find a statistically significant jump in gestational age at the very low birth weight cutoff. Our estimated jump may be an artifact, resulting from a slight nonlinearity in the relation between gestational age and birth weight.

[20]We also examine the pattern of predicted mortality around the cutoff for higher-income women. Consistent with the discussion in Section 4, the results in Appendix Table A3 show that the RD assumptions are more likely to be violated among higher-income women. In particular, babies in the highest income quartile born slightly before the week-37 cutoff have higher predicted mortality rates than those born slightly after the cutoff. This pattern may be related to the selection of induced and stimulated births out of our sample, especially given that the jump in the probability of an induction is almost twice as large among the highest income group than among the lowest income quartile.

5.2 Medical Treatments and Newborn Health

If the Dutch institutional rule governing the supervision of premature births is binding, then we should observe a discontinuity in the fraction of births supervised by an OB/GYN at 37 completed gestational weeks. To examine this, in Figure 3(a) we plot the fraction of births supervised by an obstetrician as a function of gestational age. The Figure confirms a substantial drop in obstetrician supervision at the week-37 cutoff. There are two reasons why the probability of obstetrician supervision does not drop 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 (e.g., if the birth progresses too fast), 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 after 37 completed gestational weeks are also supervised by OB/GYNs.

Recall that obstetricians supervise births only in a hospital while midwife-led births can take place either at home or in a hospital. We may then observe a change in the location of birth across the cutoff if the week-37 rule shifts some of the births that would have taken place at home to the hospital. In addition, we may also see discontinuities in medical interventions at birth, such as the use of forceps or vacuum, that can be used by OB/GYNs but not by midwives. Finally, if the guideline leads to more intensive treatments in general, we may also see discontinuities in treatments performed after birth such as admission to a neonatal intensive care unit (NICU). Figures 3(b)–3(d) show large discontinuities in all these treatments with the exception of the use of forceps or vacuum during birth.

In Table 2 we provide regression estimates corresponding to these figures. Each cell in the table reports the coefficient of $W37$ from a different regression. The results confirm that babies born slightly before the week-37 cutoff are more likely to receive medical treatments. They are, on average, 28.4 percentage points more likely to be supervised by an obstetrician; 16.4 percentage points more likely to be delivered in a hospital; and 4.5 percentage points more likely to be admitted to a NICU within the first seven days of life. These estimates are economically large, representing increases of 23–62% when compared to the corresponding means for newborns with gestational age above 37 weeks.

Next, we investigate the effects of the week-37 rule on short-term newborn health as measured by 7-day and 28-day mortality, i.e., the number of deaths within the first 7 and 28 days of life per 1,000 live births. Figure 4 presents visual evidence by plotting the evolution of newborn mortality as a function of gestational age. The Figure points to large mortality gains for newborns slightly below the week-37 cutoff. The regression estimates in Table 3 support the visual evidence: babies born slightly before 37 completed gestational weeks are significantly less likely to die when compared to newborns who are slightly above the week-37 cutoff, especially within the first week of life.

A few aspects should be taken into account when thinking about the magnitudes of the effects. First, it is worth emphasizing that these estimates have relatively wide confidence intervals due to sample size, which is too small to allow us to estimate the effects precisely for rare outcomes such as mortality. The confidence intervals include much smaller but still economically important returns: for example, the lower bound of a 95-percent confidence interval indicates 1.4 fewer infant deaths per 1,000 births during the first week of life. Second, as we will show in the next section, some of the specification checks we conduct in the following section do imply slightly lower and more plausible reductions in 7-day mortality that are still statistically significant and economically large.

5.3 Robustness Checks

In this section, we investigate the robustness of the health gains induced by the week-37 rule. We start by examining sensitivity 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 Column 1 of Table 4 we present estimates from a specification that includes the full set of controls described in section 4. We again find significant health benefits for babies born slightly below the week-37 cutoff of similar magnitude 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 near the cutoff. Our findings are robust to excluding births with a gestational age of 258 or 259 days (Column 2). Column 3 focuses on the choice of kernel and reports results based on a rectangular kernel which places equal weights on all observations. Our results again point to health gains for babies slightly below the week-37 cutoff. In Figure 5, we investigate the robustness of our estimates to the choice of bandwidth. The Figure plots the estimated coefficient of $W37$ along with its 95% confidence interval. We present results for all bandwidths between 5-28 gestational days in 1-day steps. The Figure shows that the magnitudes of the estimates are quite stable across different bandwidths, even though they are less precisely estimated for 28-day mortality when using smaller bandwidths.

In order to check whether the linear approximation is appropriate, in Appendix Figure A1 we superimpose a Lowess curve and our local-linear approximation on the raw 7-day and 28-day mortality data in a 21-day window on each side of the cutoff. The figures suggest that a linear approximation is well-suited within the baseline bandwidth of 7 days, especially for 28-day mortality. To investigate whether our results for 7-day mortality are driven by our functional form assumption, we consider a specification similar to equation (1) but with a second-degree polynomial in gestational age. Similar to Figure 5, in Appendix Figure A2 we plot the estimated coefficients from this specification for every bandwidth between 5 and 28 days. Although this specification tends to overfit the data for small bandwidths, our results still indicate economically large but slightly smaller effect sizes that are generally statistically significant and stable for bandwidths starting from 7 days.

In Columns 4–6 of Table 4, 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. Columns 4–5 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 gain substantially from the medical treatments induced by the week-37 rule. In Column 6, we check the robustness of our results to including planned C-sections and induced births. These deliveries are excluded from our analysis sample to avoid a potential bias from manipulation of the running variable. However, physicians are less likely to induce births to the left of the cutoff (see Appendix Table A2), which could lead to a bias toward zero. Moreover, selection bias could occur if some births ending in an unplanned C-section are coded as planned and if this practice changes across the cutoff. Adding planned C-sections and induced births leads to slightly larger point estimates but does not change the results in a meaningful way.[21]

The last two columns investigate whether the observed health gains are driven by our specific classification of the income distribution. Using income terciles and quintiles, we confirm that newborns from the lowest income areas experience mortality reductions from the week-37 rule.

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 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 7 days on either side of the cutoff. Figure 6 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 that there is no other cutoff where 7-day mortality exhibits statistically significant gains of a magnitude comparable to those observed at the week-37 cutoff. Consistent with our baseline estimates, we do not find any statistically significant discontinuities

[21] We 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 found qualitatively similar health gains (results available upon request).

in 28-day mortality across the distribution of gestational age.

It is important to emphasize that the key identification assumptions of an RD design are ultimately not testable. In section 5.1 and in this section we examine several scenarios that could lead to violations of these assumptions, but there may still be situations (e.g., due the risk classification employed by midwives in the Netherlands) under which they may not hold. In that spirit, none of the tests presented are sufficient on their own to claim the validity of our RD design. However, together they consistently suggest that the RD assumptions hold and that the week-37 rule improves short-term newborn health. The reduction in 7-day mortality is statistically significant and economically large across all specification checks. While health gains within the first 28 days of life are less precisely estimated when using smaller bandwidths, the estimated effect is economically large and mostly stable across our robustness checks.

5.4 Potential Mechanisms and Cost of a Life Saved

In this section, we first investigate potential explanations for the observed mortality gains. Ideally, we would have liked to pinpoint the causes of death that drive our results. Our data does not include the cause of death but has a number of variables that can hint at the importance of the medical professional attending the birth. In particular, we examine whether the week-37 rule affects the likelihood of presence of meconium in amniotic fluid, emergency C-section or birth trauma/perinatal asphyxia (oxygen deprivation). Presence of meconium in amniotic fluid is generally considered the most direct measure of fetal distress during labor and can lead to severe complications if breathed in by the baby (meconium aspiration). Among the most common causes of emergency C-sections are fetal and maternal distress and maternal hemorrhage. Finally, birth trauma/perinatal asphyxia may be due to poor handling of the newborn by the birth attendant. The results, presented in Table 5, suggest that the week-37 rule is successful in reducing these adverse outcomes. Newborns slightly below the cutoff are 1.3 percentage points less likely to be exposed to meconium in the amniotic fluid, 1.7 percentage points less likely to be delivered via an emergency C-section and 1.2 percentage points less likely to experience birth trauma or perinatal asphyxia. All of these estimates are economically large, representing approximately 25% reductions when compared to the mean of the outcomes above the cutoff, even though only the first two are statistically significant.

We next try to gauge the economic significance of our findings by conducting 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 6 details the calculations. Our results indicate that the week-37 rule increases the number of obstetrician-supervised hospital deliveries by 284 per 1,000 births. Of these, roughly 164 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 €421,075 per 1,000 births due to higher use of OB/GYN-supervised hospital births because of the week-37 rule.

Our estimates also suggest that prematurity leads to 3.5 fewer 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), subtracting €1,628 from the costs per 1,000 births. Turning to NICU admissions, we find that there are 45 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.[22] This implies that the week-37 rule is associated with an additional cost of €411,190 per 1,000 births.

Overall, the estimated additional cost of the week-37 rule per 1,000 deliveries is €830,637. Compared to an average reduction in 7-day (28-day) mortality

[22] This 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).

of 3.6 (1.5) deaths per 1,000 births, the implied cost per life saved is €230,669 (€549,000).[23] 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 2016 prices. Taken together, these results suggest that the current implementation of the week-37 rule is highly cost-effective.[24]

6 Conclusions

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

Using data for the period 2000–2008, we find that the week-37 rule leads to statistically and economically significant increases in the receipt of perinatal medical treatments. Our results also indicate that low-income preterm newborns are significantly less likely to die when compared to newborns who are

[23]On 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 does not only reduce mortality, but also reduces the use of emergency C-sections, or the occurrence of birth trauma/perinatal asphyxiation which can lead to permanent brain damage and consequent additional treatments and loss of human capital.

[24]We reach the same conclusion even if we use the lower bound of the confidence interval for the reduction in 7-day mortality, i.e., 1.47 deaths per 1,000 births. In that case, the cost per life saved is €564,123, still much lower than the value of a statistical life.

slightly above the week-37 cutoff. While we are not able to investigate causes of death directly, we provide suggestive evidence on the importance of the role of the birth attendant: the week-37 rule leads to large reductions in the presence of meconium in the amniotic fluid, emergency C-sections and birth trauma or perinatal asphyxiation.

Our results are relevant to the ongoing policy debates on effective health policy. Our results caution against designing "one-size-fits-all" policies when crafting policies about child birth technologies. The fact that medical interventions improve the health outcomes of some low-income 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. Particularly among lower socioeconomic status women, some risks may remain unrecognized until the moment of birth, meaning that poorer women can benefit from receiving a higher level of medical care even if no explicit risk factors have been recognized.

References

Almond, Douglas and Joseph Doyle (2011) "After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays," *American Economic Journal: Economic Policy*, Vol. 3, No. 3, pp. 1–34, DOI: 10.1257/pol.3.3.1.

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*, Vol. 125, No. 2, pp. 591–634, DOI: http://www.mitpressjournals.org/loi/qjec.

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*, Vol. 55, No. 3, pp. 216–225, URL: http://www.sciencedirect.com.ezproxy.usherbrooke.ca/science/article/B6W6R-4YYH3WF-9/2/081d82d6acec27a06fd26f683ccf633b, DOI: 10.1016/j.jmwh.2010.01.001.

Bais, Joke and Maria Pel (2006) "The basis of the Dutch obstetric system: risk selection," *European Clinics in Obstetrics and Gynaecology*, Vol. 2, No. 4, pp. 209–212, URL: http://www.springerlink.com.ezproxy.usherbrooke.ca/content/p211740450151478/abstract/, DOI: 10.1007/s11296-006-0053-6.

Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell (2016) "Heaping-Induced Bias in Regression-Discontinuity Designs," *Economic Inquiry*, Vol. 54, No. 1, pp. 268–293, URL: http://onlinelibrary.wiley.com.proxy1-bib.sdu.dk:2048/doi/10.1111/ecin.12225/abstract, DOI: 10.1111/ecin.12225.

Bharadwaj, Prashant, Katrine Løken, and Christopher Neilson (2013) "Early Life Health Interventions and Academic Achievement," *American Economic Review*, Vol. 103, No. 5, pp. 1862–1891.

Buisman, Peter and Geert Gerats (2008) "Verloskundigen in tijd erop vooruit in achterstandswijken," *Tijdschrift voor Verloskundigen*, Vol. September, pp. 13–16.

Chandra, Amitabh and Jonathan Skinner (2012) "Technology Growth and Expenditure Growth in Health Care," *Journal of Economic Literature*, Vol. 50, No. 3, pp. 645–680.

Cutler, David, Mark McClellan, Joseph Newhouse, and Dahlia Remler (1998) "Are Medical Prices Declining? Evidence from Heart Attack Treatments," *Quarterly Journal of Economics*, Vol. 113, No. 4, pp. 991–1024, URL: http://www.jstor.org/stable/2586972.

Cutler, David and Ellen Meara (2000) "The Technology of Birth: Is It Worth It?" *NBER/Frontiers in Health Policy Research*, Vol. 3, No. 1, pp. 33–67, DOI: 10.1162/109623100300091069.

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

Daysal, N. Meltem, Marianne Simonsen, Mircea Trandafir, and Sanni Breining (2018) "Spillover effects of early-life medical interventions," mimeo.

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*, Vol. 7, No. 3, pp. 28–50.

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*, Vol. 11, No. 2, URL: http://www.degruyter.com/view/j/fhep.2008.11.2/fhep.2008.11.2.1102/fhep.2008.11.2.1102.xml?format=INT, DOI: 10.2202/1558-9544.1102.

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.)*, Vol. 341, p. c5639, URL: http://www.bmj.com/cgi/doi/10.1136/bmj.c5639, DOI: 10.1136/bmj.c5639.

Frandsen, Brigham R. (2017) "Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable Is Discrete," in Cattaneo, Matias D. and Juan Carlos Escanciano eds. *Advances in Econometrics*, Vol. 38: Emerald Publishing Limited, pp. 281–315, DOI: 10.1108/S0731-905320170000038012.

Fuchs, Victor (1998) *Who Shall Live?: Health, Economics, and Social Choice*: World Scientific.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw (2001) "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, Vol. 69, No. 1, pp. 201–209, URL: http://onlinelibrary.wiley.com.ezproxy.usherbrooke.ca/doi/10.1111/1468-0262.00183/abstract, DOI: 10.1111/1468-0262.00183.

Imbens, Guido and Thomas Lemieux (2008) "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, Vol. 142, No. 2, pp. 615–635, URL: http://www.sciencedirect.com/science/article/pii/S0304407607001091, DOI: 10.1016/j.jeconom.2007.05.001.

Institute of Medicine (2007) *Preterm Birth: Causes, Consequences, and Prevention*, Washington, D.C.: The National Academies Press.

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

van der Kooy, Jacoba, Jashvant Poeran, Johanna de Graaf, Erwin Birnie, Semiha Denktaş, Eric Steegers, and Gouke Bonsel (2011) "Planned

Home Compared With Planned Hospital Births in The Netherlands," *Obstetrics & Gynecology*, Vol. 118, pp. 1037–1046, URL: http://ovidsp.tx.ovid.com.ezproxy.usherbrooke.ca/sp-3.4.2a/ovidweb.cgi?&S=CENPFPBELNDDKFLCNCBLGDFBKIPAAA00&Link+Set=S.sh.15.16.18%7c11%7csl_10, DOI: 10.1097/AOG.0b013e3182319737.

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*, Vol. 206, No. 2, pp. 108–112, URL: http://www.sciencedirect.com/science/article/pii/S0002937811021697, DOI: 10.1016/j.ajog.2011.10.864.

Lee, David and Thomas Lemieux (2010) "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, Vol. 48, No. 2, pp. 281–355, DOI: http://www.aeaweb.org/jel/index.php.

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.

McCrary, Justin (2008) "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, Vol. 142, No. 2, pp. 698–714, URL: http://www.sciencedirect.com/science/article/pii/S0304407607001133, DOI: 10.1016/j.jeconom.2007.05.005.

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*, Vol. 6, No. 1, pp. 1–34, DOI: http://www.bepress.com/bejeap/advances/.

National Institute for Health and Care Excellence (2008) "Induction of labour," available online at https://www.nice.org.uk/guidance/cg70, last accessed on Jan 15, 2019.

Nederlandse Vereniging voor Obstetrie en Gynaecologie (2006) "Inductie van de baring," available online at https://www.nvog.nl/wp-content/uploads/2017/12/Inductie-van-de-baring-1.0-20-09-2006.pdf, last accessed on Jan 15, 2019.

Newhouse, Joseph (1992) "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives*, Vol. 6, No. 3, pp. 3–21, DOI: http://www.aeaweb.org/jep/.

Norwitz, Errol R and Aaron B Caughey (2011) "Progesterone Supplementation and the Prevention of Preterm Birth," *Reviews in Obstetrics and Gynecology*, Vol. 4, No. 2, pp. 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, URL: https://www.perined.nl/producten/publicaties/jaarboeken.

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*, Vol. 78, pp. 3–6.

Skinner, Jonathan, Douglas Staiger, and Elliott Fisher (2006) "Is Technological Change In Medicine Always Worth It? The Case Of Acute Myocardial Infarction," *Health Affairs*, Vol. 25, No. 2, pp. w34–w47, URL: http://content.healthaffairs.org/content/25/2/w34.abstract, DOI: 10.1377/hlthaff.25.w34.

SWOV (2012) "Factsheet: The valuation of human costs of road deaths," URL: https://www.swov.nl/rapport/Factsheets/UK/FS_Immaterial_costs.pdf, available online at https://www.swov.nl/rapport/Factsheets/UK/FS_Immaterial_costs.pdf, last accessed on Oct. 28, 2016.

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*, Vol. 13, No. 9, p. e1002134.

Wiegerinck, MMJ, BY van der Goes, ACJ Ravelli, JAM van der Post, J Klinkert, J Brandenbarg, 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*, Vol. 31, No. 12, pp. 1168–1176.

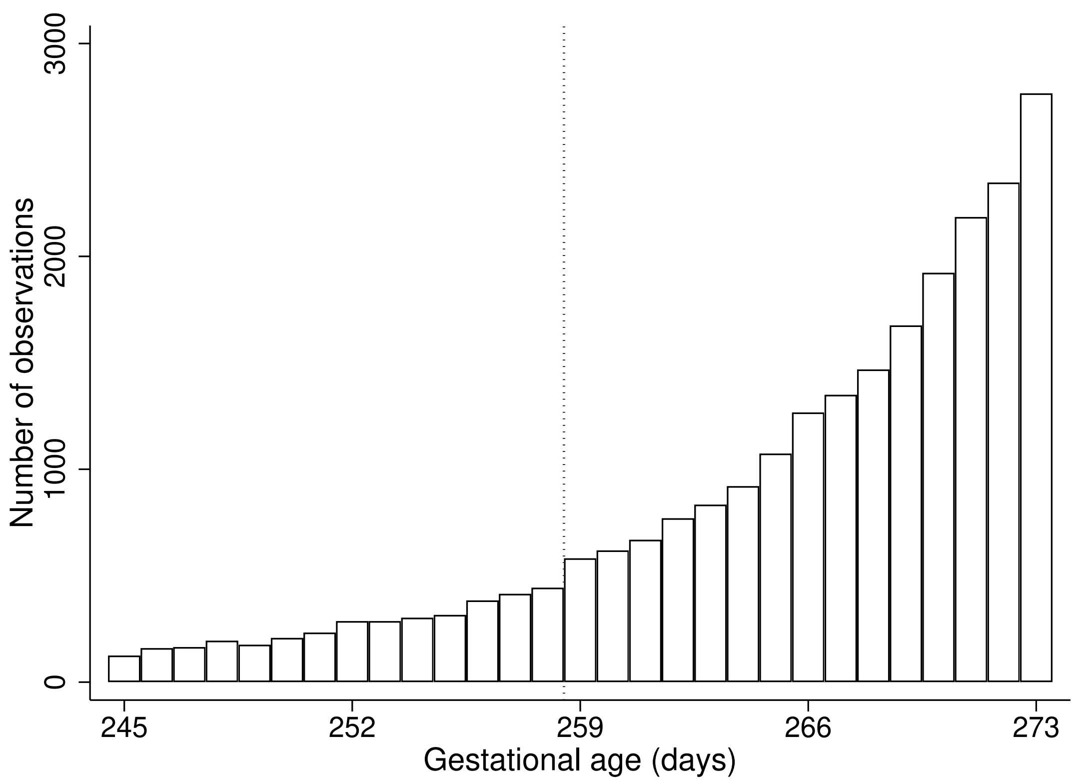
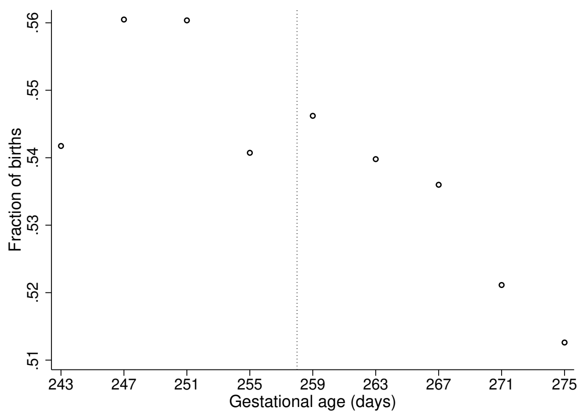
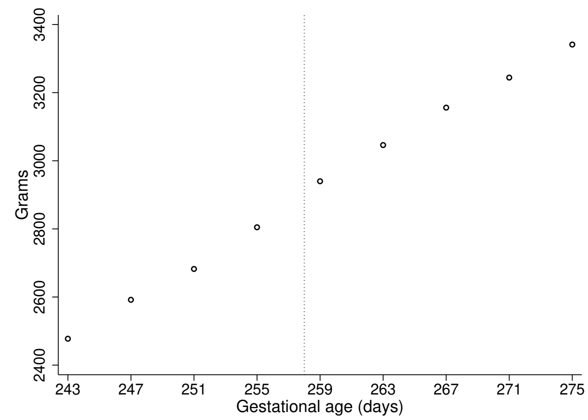


Figure 1: Frequency of births around the week-37 cutoff

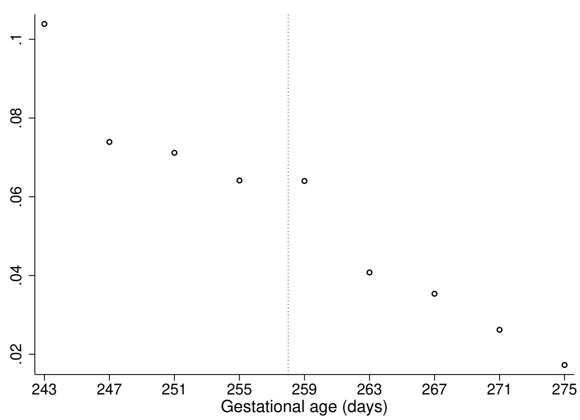
Notes: Each bar represents the number of births with gestational age indicated on the horizontal axis. The sample includes first births to low-risk mothers living in postal codes in the first quartile of average household income.



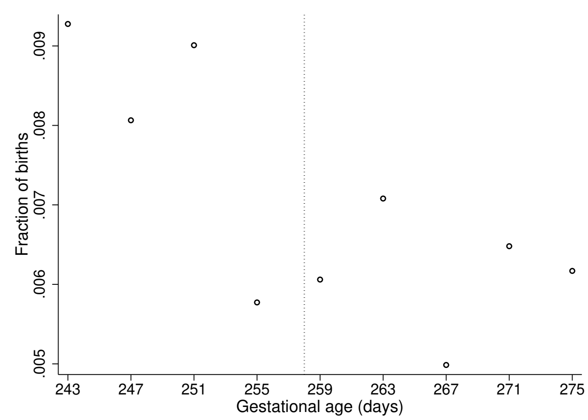
(a) Share male



(b) Birth weight (grams)

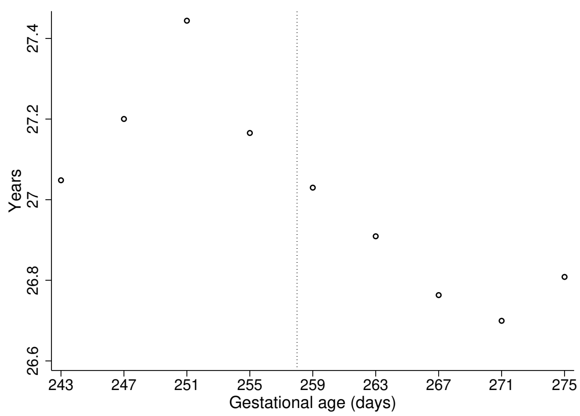


(c) Share breech birth

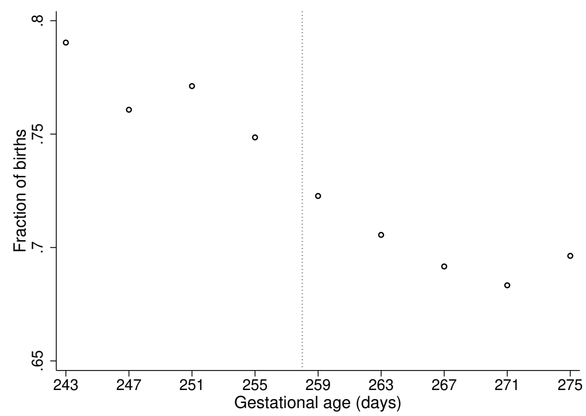


(d) Share mild congenital anomaly

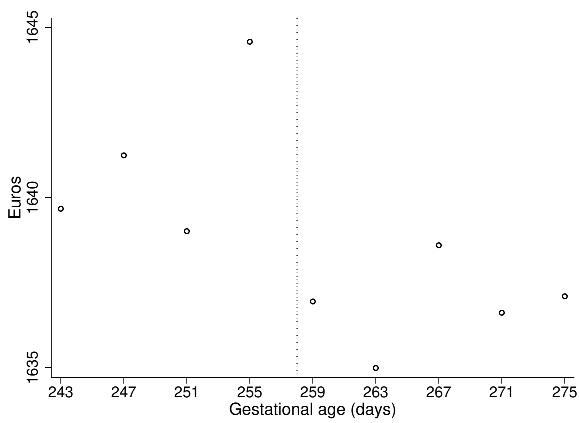
Figure 2: Distribution of selected covariates around the week-37 cutoff



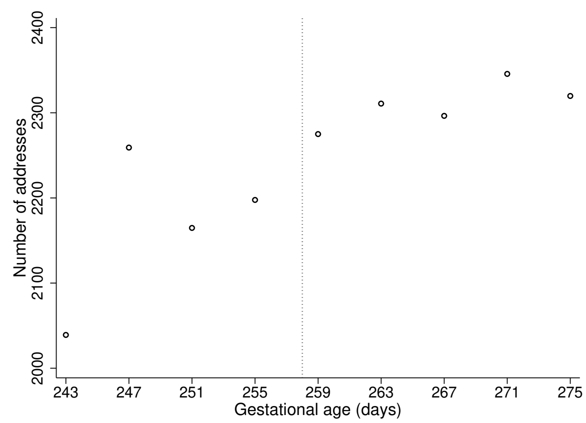
(e) Mother's age (years)



(f) Share ethnic Dutch mother



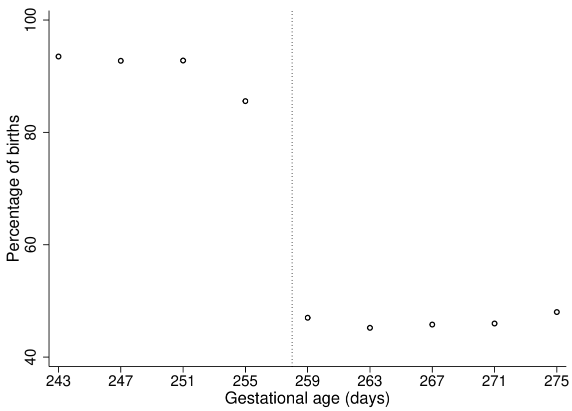
(g) Average postal code household income (euros/month)



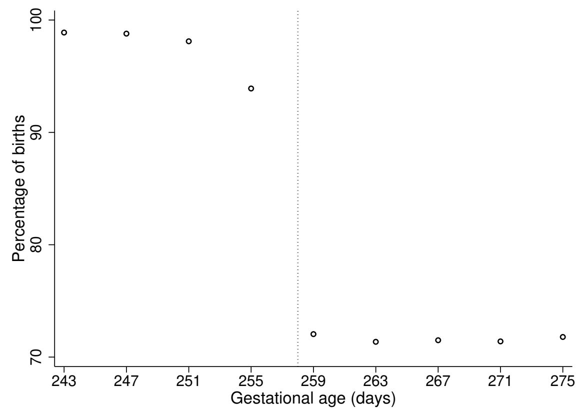
(h) Average postal code density

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

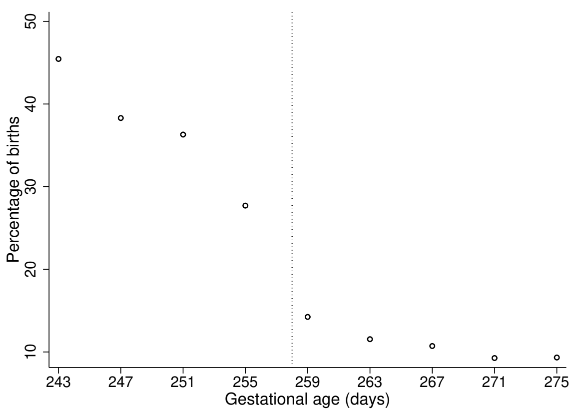
Notes: Each point represents the average value of the corresponding covariate for births with gestational age in a 4-day interval with lower limit indicated on the horizontal axis. Average postal code density is the average number of addresses per square kilometer in a circle with a radius of 1 km around each address in the postal code. The sample includes first births to low-risk mothers living in postal codes in the first quartile of average household income.



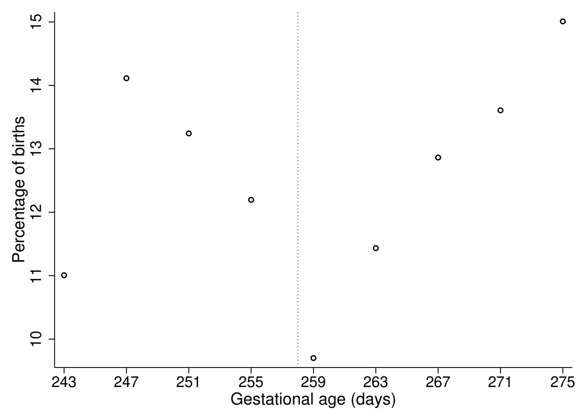
(a) Percent supervised by obstetrician



(b) Percent hospital birth



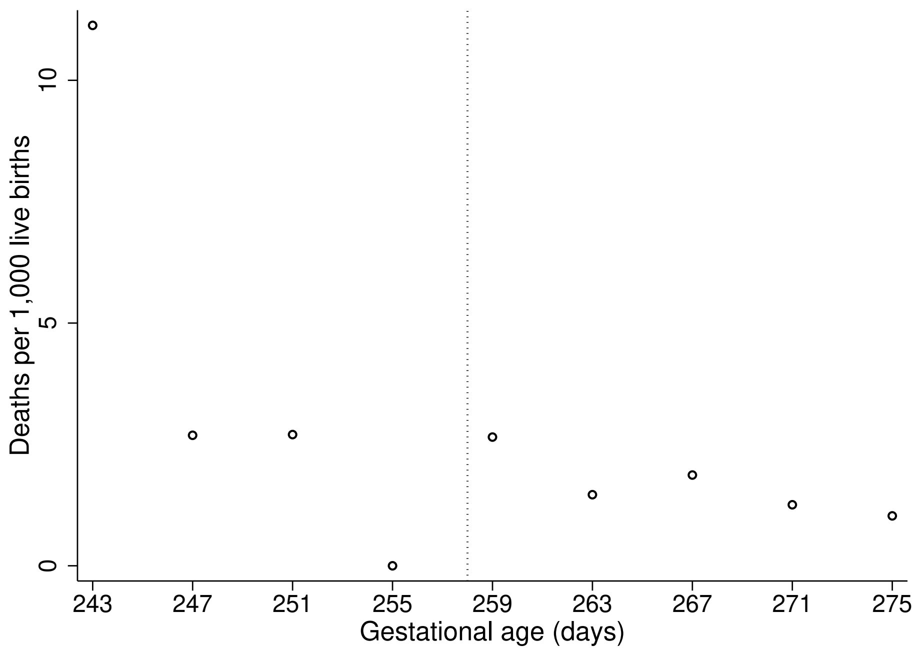
(c) Percent admitted to NICU



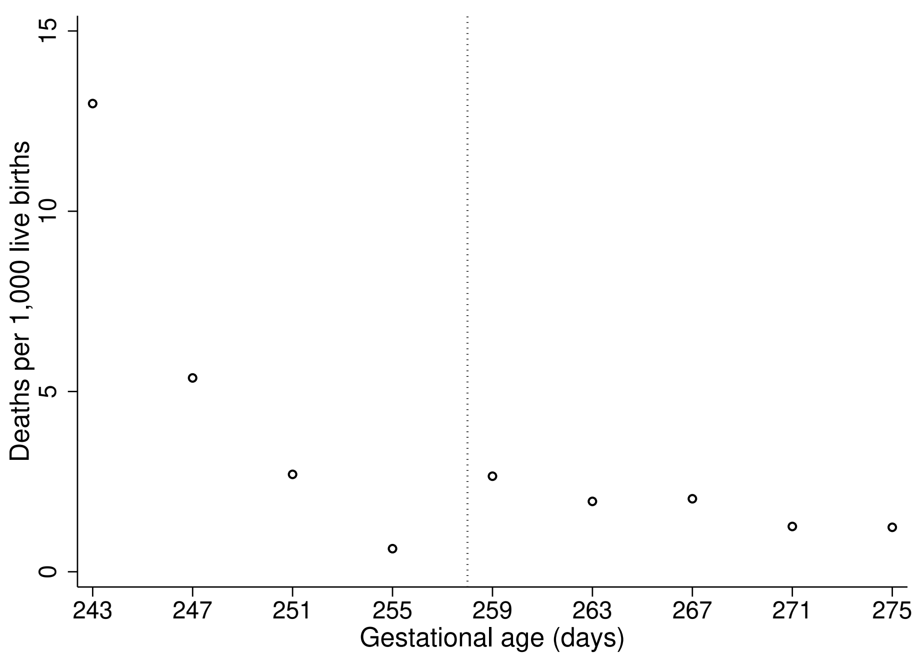
(d) Percent use of forceps/vacuum

Figure 3: Medical treatments around the week-37 cutoff

Notes: Each point represents the average value of the corresponding outcome for births with gestational age in a 4-day interval with lower limit indicated on the horizontal axis. The sample includes first births to low-risk mothers living in postal codes in the first quartile of average household income.



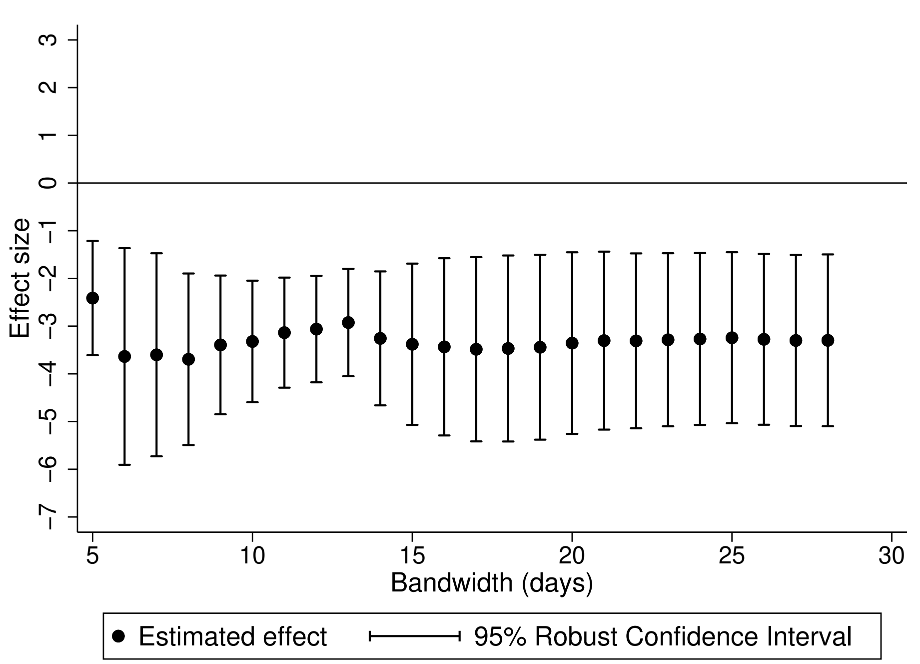
(a) 7-day mortality



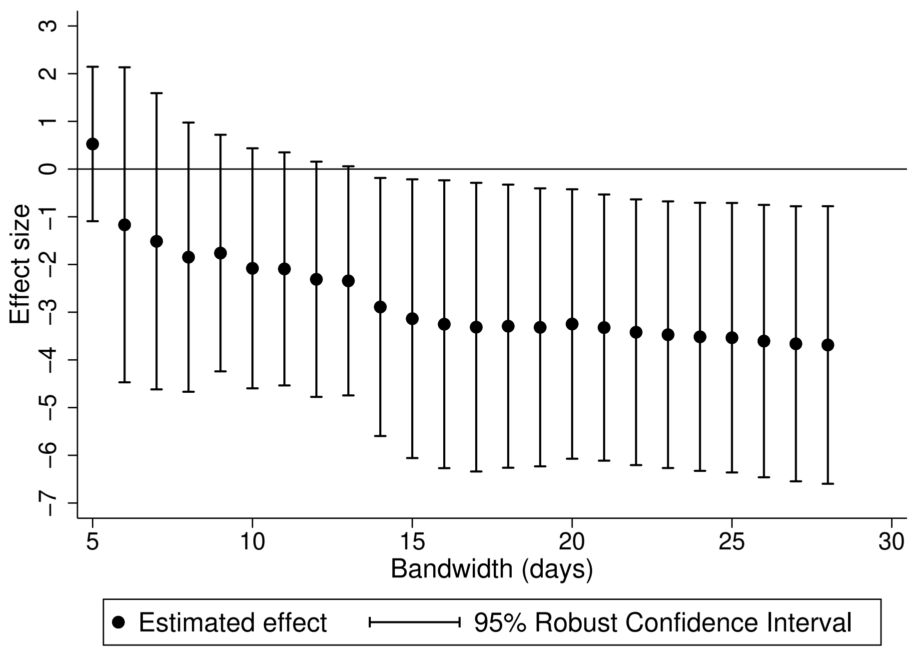
(b) 28-day mortality

Figure 4: Short-term newborn health around the week-37 cutoff

Notes: Each point represents the number of deaths within 7 and 28 days after birth, respectively, per 1,000 live births with gestational age in a 4-day interval with lower limit indicated on the horizontal axis. The sample includes first births to low-risk mothers living in postal codes in the first quartile of average household income.



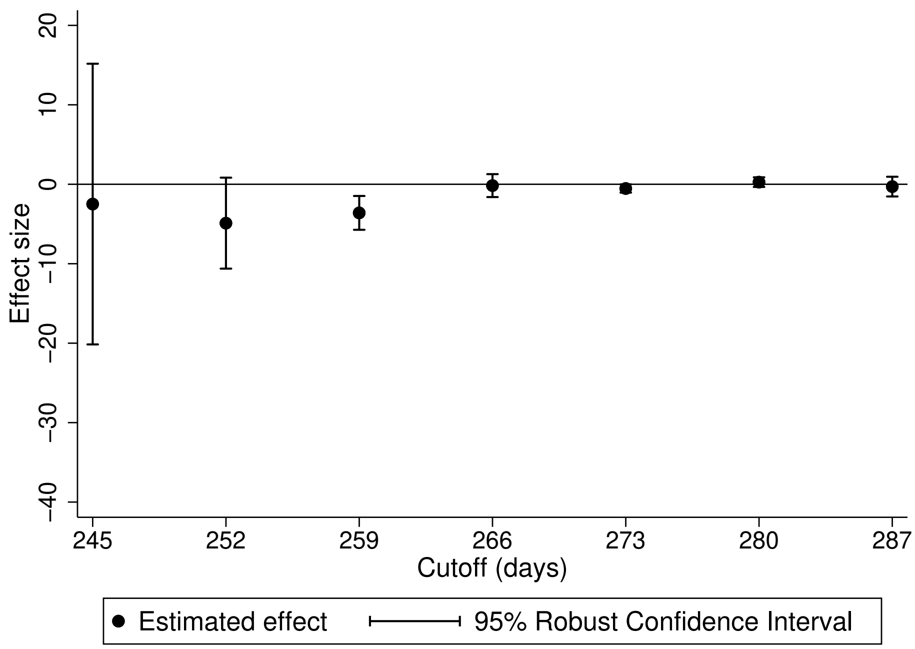
(a) 7-day mortality



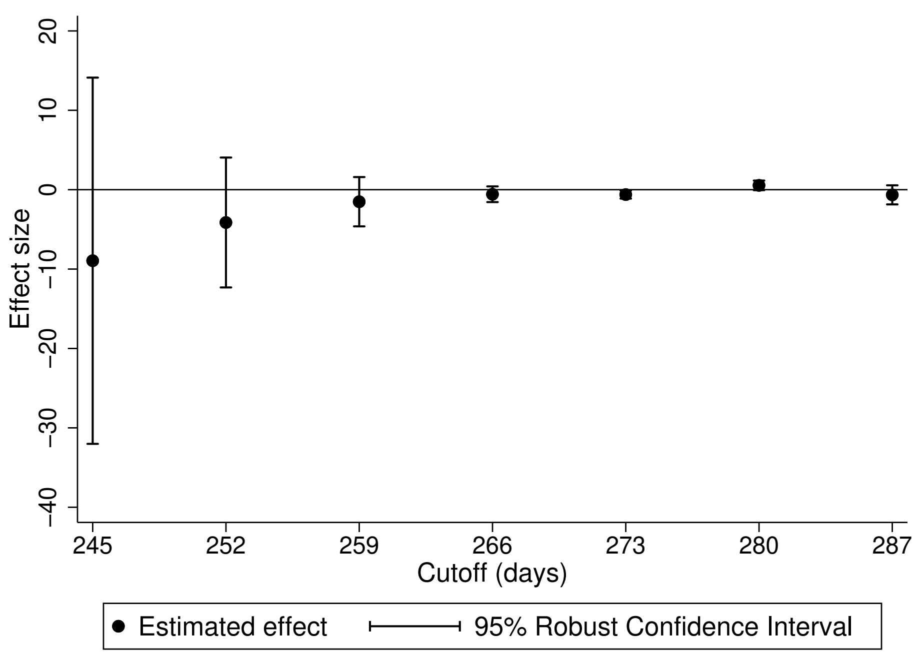
(b) 28-day mortality

Figure 5: Effects of week-37 rule on newborn mortality for bandwidths between 5–28 days

Notes: Each point and bar represent the coefficient estimate of $W37$ and the corresponding 95% confidence interval, respectively, from a local-linear regression similar to equation (1), using the bandwidth indicated on the horizontal axis and with dependent variable indicated in the figure title. The sample includes first births to low-risk mothers living in postal codes in the first quartile of average household income.



(a) 7-day mortality



(b) 28-day mortality

Figure 6: Newborn health across the gestational age distribution

Notes: Each point and bar represent the coefficient estimate and the corresponding 95% confidence interval, respectively, of an indicator that the birth had a gestational age below the cutoff indicated on the horizontal axis. The results are obtained from a local-linear regression similar to equation (1), using a bandwidth of 7 days on each side of the cutoff and with dependent variable indicated in the figure title. The sample includes first births to low-risk mothers living in postal codes in the first quartile of average household income.

Table 1: Covariates around the week-37 cutoff

	A. Maternal characteristics			B. Characteristics of postal code of residence			
	Age	Ethnicity: Dutch	Ethnicity: Mediterranean	Average monthly hh. income	Average density	Percent residents aged 0–15	Distance to closest hospital
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$W37$	−0.048	0.013*	0.022***	6.595**	−136.490**	0.220***	0.605***
	(0.179)	(0.006)	(0.007)	(2.527)	(58.835)	(0.049)	(0.148)
Mean outcome	26.956	0.718	0.124	1634.809	2283.560	17.725	4.106
Observations	7,618	7,618	7,618	7,618	7,618	7,618	7,618

	C. Newborn characteristics					D. Predicted mortality	
	Boy	Birth weight	Mild congenital anomaly	Severe congenital anomaly	Breech birth	7-day	28-day
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
$W37$	−0.022	−45.604***	−0.001	−0.006	−0.007	0.280	0.236
	(0.015)	(5.767)	(0.002)	(0.005)	(0.010)	(0.191)	(0.205)
Mean outcome	0.551	2983.012	0.007	0.012	0.053	3.020	3.305
Observations	7,618	7,618	7,618	7,618	7,618	7,617	7,617

Notes: Each cell reports the estimated coefficient of $W37$ (an indicator for prematurity) from a different regression with dependent variable listed in the column heading. All specifications are local-linear regressions using a triangular kernel and include a first-degree polynomial in normalized gestational age allowed to differ on each side of the cutoff. Mean outcome calculated using 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

	Obstetrician supervision (1)	Hospital birth (2)	Use of forceps or vacuum (3)	NICU admission (4)
$W37$	28.434***	16.405***	−0.350	4.493**
	(2.695)	(1.580)	(0.968)	(1.180)
Mean outcome	46.114	71.914	10.538	13.074
Observations	7,618	7,618	7,614	7,618

Notes: Each cell reports the estimated coefficient of $W37$ (an indicator for prematurity) from a different regression with dependent variable listed in the column heading. All specifications are local-linear regressions using a triangular kernel and include a first-degree polynomial in normalized gestational age allowed to differ on each side of the cutoff. Mean outcome calculated using 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

	7-day mortality (1)	28-day mortality (2)
$W37$	-3.601^{***}	-1.513
	(1.086)	(1.584)
Mean outcome	2.194	2.377
Observations	7,618	7,618

Notes: Each cell reports the estimated coefficient of $W37$ (an indicator for prematurity) from a different regression with dependent variable listed in the column heading. All specifications are local-linear regressions using a triangular kernel and include a first-degree polynomial in normalized gestational age allowed to differ on each side of the cutoff. Mean outcome calculated using observations to the right of the cutoff. Robust standard errors clustered at the gestational day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Robustness checks

	Including controls	Donut regression	Rectangular kernel	Earliest referral date		Including C-sections	Alternative income classifications	
				Day 245	Day 231	and induced births	Lowest tercile	Lowest quintile
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
7-day mortality	-3.683^{**}	-5.772^{**}	-3.891^{***}	-3.695^{***}	-3.541^{***}	-3.990^{***}	-3.220^{***}	-3.770^{***}
	(1.243)	(1.865)	(0.761)	(1.125)	(1.061)	(0.776)	(0.579)	(1.216)
Mean outcome	2.194	2.251	1.930	2.230	2.175	2.565	2.184	2.458
Observations	7,618	6,592	9,172	7,466	7,717	9,172	10,261	6,193
28-day mortality	-1.650	-5.712^{**}	-2.448^{*}	-1.558	-1.490	-2.144	-1.720^{*}	-1.111
	(1.761)	(1.859)	(1.322)	(1.630)	(1.551)	(1.346)	(0.921)	(1.919)
Mean outcome	2.377	2.455	2.227	2.416	2.356	2.867	2.321	2.681
Observations	7,618	6,592	9,172	7,466	7,717	9,172	10,261	6,193

Notes: Each cell reports the estimated coefficient of $W37$ (an indicator for prematurity) from a different regression with dependent variable listed in the row heading. All specifications are local-linear regressions using a triangular kernel (unless specified otherwise) and include a first-degree polynomial in normalized gestational age allowed to differ on each side of the cutoff. Mean outcome calculated using 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 5: Potential mechanisms

	Meconium in amniotic fluid (1)	Emergency C-section (2)	Birth trauma or perinatal asphyxia (3)
$W37$	-1.314^{**}	-1.684^{**}	-1.223
	(0.478)	(0.593)	(2.067)
Mean outcome	5.180	6.915	4.540
Observations	6,774	7,614	6,774

Notes: Each cell reports the estimated coefficient of $W37$ (an indicator for prematurity) from a different regression with dependent variable listed in the column heading. All specifications are local-linear regressions using a triangular kernel and include a first-degree polynomial in normalized gestational age allowed to differ on each side of the cutoff. Mean outcome calculated using 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 6: Calculation of healthcare costs related to the week-37 rule

	Estimated number of additional treatments per 1,000 births (1)	Cost per delivery (2)	Estimated additional costs per 1,000 births (3)
Transfer from hospital+midwife	120.29	1,141	137,203
Transfer from home+midwife	164.05	1,730	283,872
Use of forceps or vacuum	−3.50	465	-1,628
NICU admission	44.93	9,152	411,190
Total			830,637

Notes: All prices are in 2016 Euros.

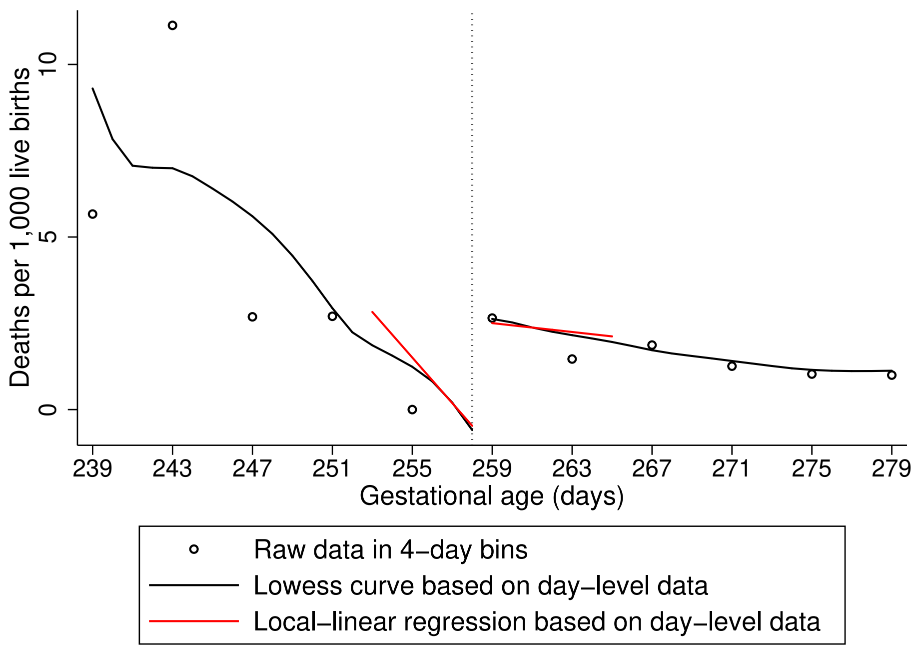
Low-risk isn't no-risk: Perinatal treatments and the health of low-income newborns

Online Appendix — Not For Publication

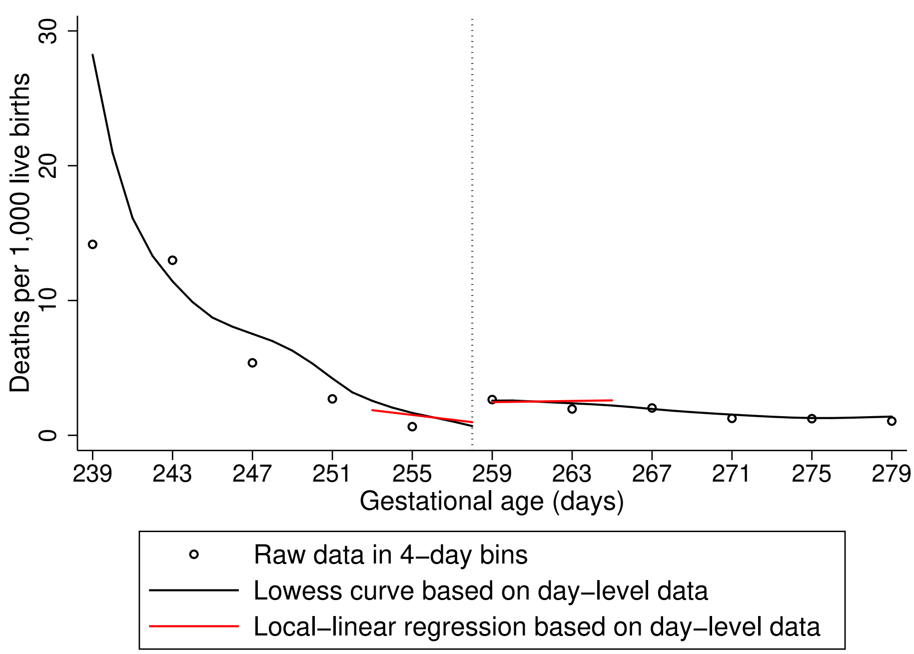
N. Meltem Daysal
University of Southern Denmark and IZA

Mircea Trandafir
University of Southern Denmark and IZA

Reyn van Ewijk
Johannes Gutenberg-University Mainz



(a) 7-day mortality



(b) 28-day mortality

Figure A1: Short-term newborn health around the week-37 cutoff

Notes: Each point represents the number of deaths within 7 and 28 days after birth, respectively, per 1,000 live births with gestational age in a 4-day interval with lower limit indicated on the horizontal axis. The Lowess curve and the (weighted) local-linear regressions are based on data at the gestational day level. The Lowess curve uses a tricube weighting function and a bandwidth of 0.8. The sample includes first births to low-risk mothers living in postal codes in the first quartile of average household income.

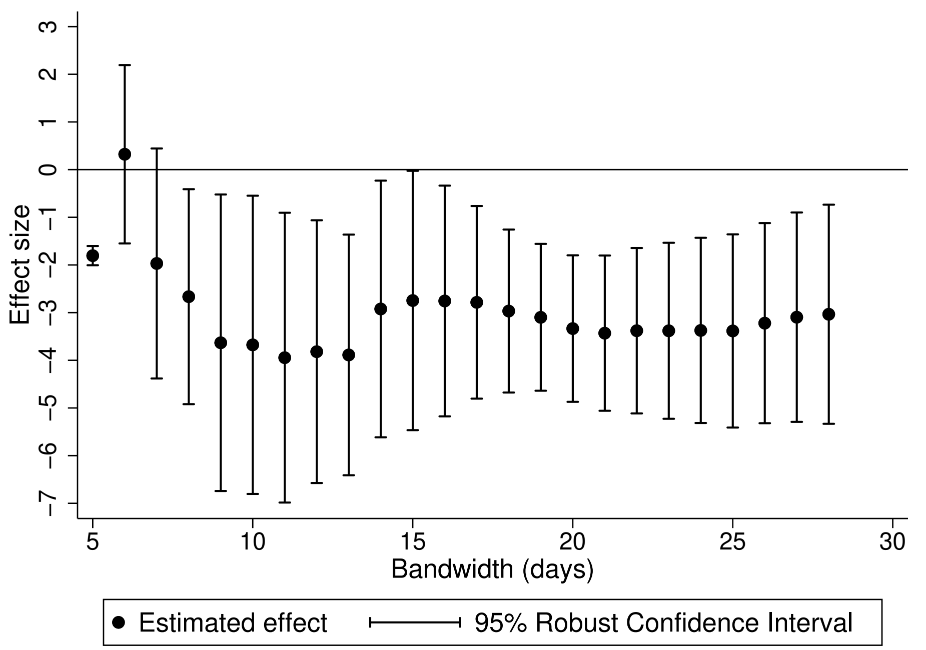


Figure A2: Effects of week-37 rule on newborn 7-day mortality for bandwidths between 5–28 days

Notes: Each point and bar represent the coefficient estimate of W37 and the corresponding 95% confidence interval, respectively, from a local regression using a second-degree polynomial in gestational age within the bandwidth indicated on the horizontal axis. The sample includes first births to low-risk mothers living in postal codes in the first quartile of average household income.

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: Probability of birth induction around the week-37 cutoff

	Quartile of average household income in postal code			
	First (1)	Second (2)	Third (3)	Fourth (4)
Birth induction	-0.026^{***}	-0.057^{***}	-0.041^{**}	-0.050^{***}
	(0.008)	(0.009)	(0.017)	(0.005)
Mean outcome	0.126	0.143	0.14	0.132
Observations	9,160	9,633	9,311	8,986

Notes: Each cell reports the estimated coefficient of $W37$ (an indicator for prematurity) from a different regression with dependent variable listed in the row heading in the sample listed in the column heading. All specifications are local-linear regressions using a triangular kernel and include a first-degree polynomial in normalized gestational age allowed to differ on each side of the cutoff. Mean outcome calculated using 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 A3: Predicted mortality around the week-37 cutoff

	Quartile of average household income in postal code			
	First (1)	Second (2)	Third (3)	Fourth (4)
Predicted 7-day mortality	0.280	−0.084	−0.233	0.954***
	(0.191)	(0.279)	(0.179)	(0.139)
Mean outcome	3.020	2.912	2.712	2.668
Observations	7,617	7,882	7,643	7,431
Predicted 28-day mortality	0.236	−0.116	−0.253	1.049***
	(0.205)	(0.324)	(0.214)	(0.145)
Mean outcome	3.305	3.176	2.957	2.892
Observations	7,617	7,882	7,643	7,431

Notes: Each cell reports the estimated coefficient of $W37$ (an indicator for prematurity) from a different regression with dependent variable listed in the row heading in the sample listed in the column heading. All specifications are local-linear regressions using a triangular kernel and include a first-degree polynomial in normalized gestational age allowed to differ on each side of the cutoff. Mean outcome calculated using observations to the right of the cutoff. Robust standard errors clustered at the gestational day level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.