CEPREMAP

CENTRE POUR LA RECHERCHE ECONOMIQUE ET SES APPLICATIONS

Document de travail (Docweb) n°2501

Health and Economic Impacts of an Early Labor Induction Policy for High-BMI Mothers

Louis FRÉGET, Maria KOCH GREGERSEN

Février 2025

Impacts sanitaires et économiques d'une politique d'induction du travail pour les mères ayant un IMC élevé [1]

Louis FRÉGET[2], Maria KOCH GREGERSEN[3]

Résumé : Nous enrichissons la littérature sur les rendements marginaux des interventions à la naissance en étudiant une intervention courante : l'induction précoce du travail pour une part croissante de grossesses, celles des femmes ayant un IMC élevé. Nous employons une méthodologie de régression en discontinuité, en exploitant les recommandations de la société d'obstétrique danoise qui préconisent une induction du travail à 7 jours après la date prévue d'accouchement, contre 10 à 13 jours pour les mères ayant un IMC pré-grossesse d'au moins 35. Nos résultats montrent que l'induction précoce améliore la santé maternelle et néonatale immédiate, réduit la consommation de soins de santé durant la première année de l'enfant et diminue le risque de dépression post-partum de la mère. Elle n'a toutefois pas d'effet sur l'offre de travail et la fécondité jusqu'à deux ans après la naissance.

Mots-clés : Interventions précoces, Induction du travail, Santé maternelle et infantile, Discontinuité de régression

Health and Economic Impacts of an Early Labor Induction Policy for High-BMI Mothers

Abstract: A large economics literature studies the marginal returns of birth interventions. Still, it is almost non-existent on a fairly common intervention: medically initiating labor to prevent the health risks of a pregnancy lasting too long. Because labor induction can also have side effects, the optimal timing of birth remains debated and can depend on the specific population of mothers under study. In this paper, we assess the effects of an early labor induction policy for a fast growing share of pregnancies: high-BMI women. We provide the first piece of causal evidence on the topic by exploiting Danish guidelines which recommend routine induction at 7 days after the expected due date instead of 10-13 days after for mothers with a pre-pregnancy BMI of at least 35. Early labor induction improves immediate maternal and neonatal

[1]Use of municipal health data in our research project was approved by the Danish Patient Safety Authority (approval 3-3013-2507/1). We gratefully acknowledge financial support from Danish Council for Independent Research (grant DFF:0218-00003B). We warmly thank Hans H. Sievertsen (U. Bristol) and Miriam Wüst (U. Copenhagen), as well as Petra Persson, Maya-Rossin Slater, Ciarian S. Phibbs and Elliot Main (Stanford U.), as well as participants of the Aarhus LPP seminar, RGS Doctoral Conference, PhD Seminar of the department of economics of Copenhagen University, DGPE 2022, WiSo HHE seminar, Gender Coffee and Labor and Public Workshop at Stanford University for their comments. All remaining errors are ours.

[2]Cepremap louis.freget@econ.ku.dk

[3]University of Aarhus mgregersen@econ.au.dk

health, reduces universal nurse visits during the first year of life of the child, as well as maternal postpartum depression risks.

Keywords : Early Health Interventions, Labor Induction, Child & Maternal Health, Regression Discontinuity

JEL Codes: I11, J13, J18

Health and Economic Impacts of an Early Labor Induction Policy for High-BMI Mothers*

Louis Freget[1] and Maria Koch Gregersen[2]

[1]Centre pour la recherche économique et ses applications (CEPREMAP)

[2]University of Aarhus

Version: February 12, 2025.

Abstract

A large economics literature studies the marginal returns of birth interventions. Still, it is almost non-existent on a fairly common intervention: medically initiating labor to prevent the health risks of a pregnancy lasting too long. Because labor induction can also have side effects, the optimal timing of birth remains debated and can depend on the specific population of mothers under study. In this paper, we assess the effects of an early labor induction policy for a fast growing share of pregnancies: high-BMI women. We provide the first piece of causal evidence on the topic by exploiting Danish guidelines which recommend routine induction at 7 days after the expected due date instead of 10-13 days after for mothers with a pre-pregnancy BMI of at least 35. Early labor induction improves immediate maternal and neonatal health, reduces universal nurse visits during the first year of life of the child, as well as maternal postpartum depression risks.

Keywords: Early Health Interventions, Labor Induction, Child & Maternal Health, Regression Discontinuity
JEL Codes: I11, J13, J18

*Use of municipal health data in our research project was approved by the Danish Patient Safety Authority (approval 3-3013-2507/1). We gratefully acknowledge financial support from Danish Council for Independent Research (grant DFF:0218-00003B). We warmly thank Hans H. Sievertsen (U. Bristol) and Miriam Wüst (U. Copenhagen), as well as Petra Persson, Maya-Rossin Slater, Ciarian S. Phibbs and Elliot Main (Stanford U.), as well as participants of the Aarhus LPP seminar, RGS Doctoral Conference, PhD Seminar of the department of economics of Copenhagen University, DGPE 2022, WiSo HHE seminar, Gender Coffee and Labor and Public Workshop at Stanford University for their comments. All remaining errors are ours.

Contact information. Freget: `louis.freget@cepremap.org`, Gregersen: `mgregersen@econ.au.dk`

1 Introduction

Following Almond et al. (2010), a dense health economics literature documents the returns of birth interventions and their longer-run consequences. Still, it is almost nonexistent on a fairly common [1] intervention: medically initiating labor to prevent the health risks of a pregnancy lasting too long (such as increased risks of stillbirth). Yet, concerns have been raised regarding policies that move routine induction earlier than 14 days after the expected due date (EDD) since labor induction can have side effects as well (Maimburg, 2020; Spillane, 2020; Rydahl et al., 2019).[2] Moreover, the existing literature concentrates almost exclusively on the effect of induction on immediate health without considering other policy-relevant outcomes such as health beyond birth, maternal mental health, or parental labor supply and fertility. Finally, the existing causal evidence is based on samples of (mainly) low-risk pregnancies. This makes it unclear whether the optimal time for routine induction varies for different sub-groups of pregnancies.

This paper provides the first piece of causal evidence on the consequences of early routine induction of labor for high-BMI women on a range of medical and economic outcomes. This is an important policy question. First, high-BMI mothers face increased morbidity and mortality risks. This indicates they may benefit from having their labor induced earlier than the general population of pregnancies. However, a recent review points to the absence of causal evidence on the net benefits on high-BMI as an indication for labor induction (Coates et al., 2020). Second, the number of overweight and obese pregnancies has been increasing in high- and middle income countries (Chen et al., 2018).[3]

We exploit a discontinuity in the recommended time for routine induction based on mothers' pre-pregnancy BMI to identify causal effects. In Denmark, women with a pre-pregnancy BMI $\geq$ 35 are recommended routine induction of labor at 7 days after the EDD instead of at

[1] one in every fourth birth at term was induced in developed countries in 2011 (WHO, 2011).

[2] See Gregersen (2024) for a review on the discussion on early induction in general.

[3] Globally, Chen et al. (2018) estimate that in 2014, 38.9 million pregnancies involved overweight mothers and 14.6 million by obese mothers

10 to 13 days after (DSOG, 2011a). Since the mother's BMI is recorded in the very *beginning of the pregnancy*, we are less concerned about women manipulating their BMI in response to early induction guidelines (which take effect and are discussed *in the end of the pregnancy*). We exploit this institutional setting and implement a regression discontinuity (RD) design to causally evaluate the intention-to-treat (ITT) effect and the local-average-treatment-effect (LATE) of early routine induction for women who have a pre-pregnancy BMI close to 35.

Leveraging rich Danish administrative data, we consider all pregnancies going at least 7 days past their EDD in Denmark from 2011 to 2017. As pregnancies with known complications are generally recommended to give birth no later than their EDD, our design allows us to isolate the effect of offering early induction for mothers with no other severe risk factors than having a high pre-pregnancy BMI. This helps determine if high-BMI in itself provides a sufficient indication for early induction, a question that has been raised in the medical literature but for which no causal evidence exists (Coates et al., 2020).

Our first stage is strong: having a BMI just above the threshold yields around 16 percentage-points (p.p) increase in the risk of being induced early (7-9 days after the EDD) - corresponding to a 163% increase relative to our control group[4]. The ITT-reduction in the gestational length (length of pregnancy) corresponds to 0.6 days, but scaling it with the increase in the early induction rate corresponds to around 4 days. This estimate is close to those found in RCTs assessing the effect of labor induction for a general population of low-risk pregnancies.

Our findings are threefold. First, we find evidence that offering early induction because of high-BMI improves certain dimensions of immediate health, but increases epidural use. There are two potential channels through which early induction could improve birth outcomes. On the one hand, early induction may help ensure birth occurs before the fetus grows excessively large. Indeed, we find that early induction significantly reduces the likelihood of delivering a newborn weighing 4,5 kg or more. While a high birth weight is generally considered a bene-

[4]Our control group is defined as births on the left-side of the bandwidth of our main specification (Maternal BMI 28-35)

ficial outcome in the health economics literature (see e.g. Behrman and Rosenzweig (2004); Black et al. (2007); Figlio et al. (2014); Kreiner and Sievertsen (2020)), an excessively large birth weight can make the delivery more difficult and lead to increased risks of complications. Consistently, we find a decrease in complications potentially affected by having a large newborn (birth weight $\geq 4,5$ kg): lacerations (broad definition), hemorrhage (broad definition) and shoulder dystocia - when the infant's shoulders get stuck during delivery. Effects on these outcomes are rarely found in existing studies on moving routine induction early for the general population of low-risk pregnancies (see Appendix A for a detailed review). On the other hand, early induction may avert rare but serious complications, such as stillbirth, that can arise from overly long pregnancy lengths. Consistently, the intervention reduces a severe complication index (a joint measure of stillbirth, infant death and having an Apgar score<5 at 5 minutes). Still, the result on the complication index does not hold in all robustness checks. It is hard to draw statistically sound conclusions on complications so rare [5].

Second, we find evidence that the intervention reduces certain dimensions of healthcare usage after birth. In particular, the number of universal nurse visits during the first year is lower for women who are just above the BMI threshold. While the number of GP visits is statistically unaffected, the coefficient suggests a reduction as well.

Third, we study the consequences of the policy on maternal mental health during the child's first year of life. Leveraging unique psychometric data from a universal nurse-visiting program after birth, our ITT estimates indicate that early induction significantly lowers the score used to screen for postpartum depression risk by around 24% of a standard deviation indicating improved maternal mental health. To our knowledge, this is the only causal evidence of a mental health effect of early labor induction.

Maternal mental health could affect many dimensions of family life including maternal decision-making power power within the family (Baranov et al., 2020), the cognitive and non-cognitive skills of the children in early childhood (Baranov et al., 2020; Frank and Meara,

[5] For instance, in 2011, in the beginning of our sample period, there were only 245 stillbirths, out of 58,999 live births, according to public data from statistics Denmark.

2009), maternal educational investments (Von Hinke Kessler Scholder et al., 2019), or maternal labor supply decisions (Bryan et al., 2019). Besides improved maternal mental health, birth interventions themselves could also affect labor market decisions through fertility. For instance, Halla et al. (2020) show that avoidable C-sections may decrease fertility and thereby increase labor market supply in the short term. However, we find no effect on labor supply in our setting 2 years after birth, nor on the probability of having another child within 2 years after birth. Similarly, Gregersen (2024) finds no effect of labor induction for the broader population of low-risk pregnancies up to three years after birth, but she also did not find evidence of any immediate health effects.

We probe the robustness of these findings through multiple checks, including balance of observables, model specifications checks, multiple hypothesis testing, and two placebo analyses: one restricting on births between 2004–2009 (before the issuance of the pre-pregnancy BMI-based early induction guidelines) and another one in which we restrict the sample to women who gave birth before seven days past their EDD, who thus were unaffected by the BMI-based early induction policy but still exposed to other BMI-based guidelines.

We contribute to the literature by offering causal evidence on the effects of early routine induction for high-BMI women. We also expand the broader literature on labor induction by examining health outcomes beyond immediate health. Furthermore, our findings also speak to a large body of literature in health economics studying the marginal returns of medical care, following the seminal work of Almond et al. (2010). A strand of this literature has focused on birth interventions including C-sections (Card et al., 2018; Costa-Ramón et al., 2022), interventions at very low birth weight cutoffs (Almond et al., 2010), home births (Daysal et al., 2015), being discharged from the hospital (Sievertsen and Wüst, 2017; Card et al., 2023), and hospital crowding (Hoe, 2022).

However, to our knowledge, the only economic paper that analyzes a labor induction policy is Gregersen (2024). Despite a large increase in the induction risk, she finds no severe consequences of introducing additional surveillance and moving routine induction forward to

10 to 13 days after the EDD instead of at 14 days after for low-risk first-time pregnancies. By contrast, we study a population with higher health risks by focusing on high-BMI women, but note we also consider a different treatment: offering routine induction at 7 days after the EDD instead of at 10-13 days after. Our findings indicate beneficial health effects for high-BMI women which shows that optimal induction policies can differ across sub-populations.

We also enrich the scarce economic literature on the effect of birth interventions on postpartum depression. While Tonei (2019) finds increased risks of maternal postpartum depression when having an unplanned C-section, our findings suggest that other (milder) types of birth interventions may also impact maternal postpartum depression risk. Thus, our findings highlight the importance of evaluating potential mental health effects of birth interventions.

2 Institutional setting

Since 2011, Danish women with a pre-pregnancy BMI$\geq$35 have been recommended induction at 7 days after the EDD[6] even in the absence of any other pregnancy risks (DSOG, 2011b). Similar women with a pre-pregnancy BMI<35 were generally recommended induction at 10-13 days after the EDD.

In Figure 1, we empirically show that the rate of births with induced labor rises from 7 days after the EDD for women above the BMI$\geq$ 35 cutoff relative to women with a BMI<35. At 10 days after the EDD, the induction rate for women with a BMI<35 begins to rise too, and at around 12 days after the EDD the induction rates for both groups have converged.

[6] Gestational length is rather precisely estimated for most women in Denmark during our sample period. It is typically determined using crown-rump length measurement during the first-trimester ultrasound scan taking place when assessing the risk for Downs Syndrome (Zizzo et al., 2017; DSOG, 2015).

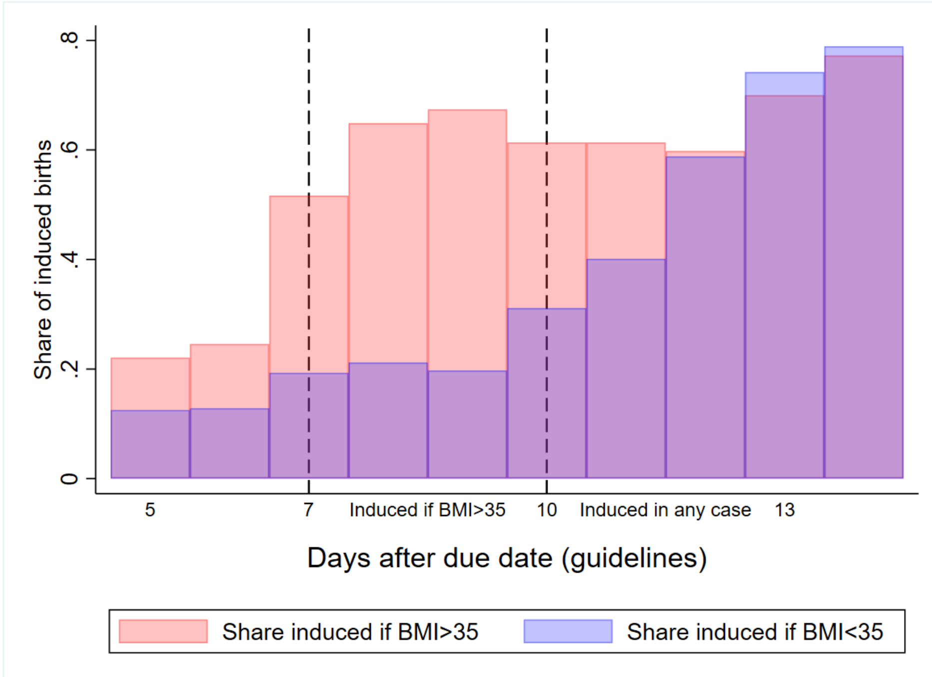


Fig. 1 Average induction rates for ongoing births by days after the due date and BMI category

Notes: The figure shows the share of births with induced labor by day after the estimated due date (EDD). The sample includes all births in Denmark between 2011 and 2017 with a gestational length of more than four days past the EDD and with information on BMI.

Since we show the rate of induced labor by the time of birth (and not the time of induction), it is to be expected that the relative difference continues to rise during 7-9 days after the EDD as depicted in Figure 1. First, a woman may not necessarily go into labor on the day of induction.[7] Second, hospitals tend to postpone induction of labor on busy days (Maibom et al., 2021). Third, even when labor is successfully initiated at the day of induction, the duration of labor may in many cases last more than one calendar day.

Thus, a woman with a pre-pregnancy BMI just above the cutoff and an otherwise un-

[7]Depending on how ready her body is (how ripe the woman's cervix is), different means of inducing labor may be implemented, and in many cases, induced labor will result in a combination of different means. When the woman's cervix is not ripe, the standard induction process typically involves taking a drug for up to 3 days (until labor kicks in or until the body is ripe enough to directly ripe the amniotic sack). If the drug is unsuccessful, the woman will typically have another appointment at the hospital to determine the best action forward (postponing induction, trying another method of induction or planning an C-section). Sometimes it is possible to rupture the amniotic sack right away. Further down the process, contraction-stimulating medicine may be necessary if the contractions do not kick in or are not strong enough.

complicated pregnancy will potentially be affected by the guidelines in two ways relative to a similar woman just below the cutoff. First, she is more likely to give birth earlier (reduced gestational length). Second, she may have a different birth mode. Had she not been induced at 7 days after the EDD, she might go into spontaneous labor before the general time of routine induction of labor at 10-13 days after the EDD or she would be recommended labor induction at that point in time.[8] When complying by the guidelines, mothers are thus exposed to *early* induction instead of the alternative which may involve later induction of labor or later spontaneous onset of labor.

As for the target population, it is important to note that this BMI cutoff for early routine induction of labor applies only to relatively uncomplicated high-BMI pregnancies that do not have other complications warranting an even earlier induction or a planned C-section. Indeed, pregnancies with obstetrical complications are typically recommended induction or elective C-section no later than their EDD (DSOG, 2011b; Hedegaard et al., 2014). Those types of pregnancies include cases of insulin-treated (gestational) diabetes, intrauterine growth restriction, preeclampsia, a history of previous intrauterine death, or a multiple birth. Only mild cases of gestational diabetes (GDM) with no need for insulin treatment[9] and with no suspicion of a large baby[10] are allowed passing their EDD together with some type of low-risk pregnancies with a previous C-section. Moreover, pregnancies may be induced even earlier if any problems are detected during surveillance or any type of complication occurs. Thus, we estimate the effect of early induction for high-BMI mothers without any other major complications.

[8] In principle, the mother could also develop a complication which would have led her to be induced earlier than 10-13 days after the EDD too.

[9] These mild cases of GDM which are recommended induction at 7 days after the EDD. Similarly women of at least 40 of age were recommended induction at 7 days after the EDD. We drop both groups of pregnancies since their recommended induction status does not differ across the BMI cutoff. Guidelines for women with a prior C-section are less uniform. They are recommended to give birth earlier than 10 days after the EDD. If allowed to be induced (instead of receiving a planned C-section) their induction procedure differs because of their elevated risk of uterus rupture (DSOG, 2011b). We control for having had a prior C-section in our main specification.

[10] Notice however that it is difficult to correctly predict the size of a large baby, especially for overweight and obese women (DSOG, 2007b, 2018). See Appendix A.2 for details on antenatal care particularly relevant for high-BMI mothers.

This policy setting allows us to address a central question in the medical and birth policy literature: although early induction is widely recommended when having a high BMI leads to conditions like preeclampsia, the net benefit of early induction because of high BMI is currently uncertain.

3 Data

This paper combines data from several administrative sources. Most of our health outcomes and control variables are extracted from the Danish Medical Birth Register (MFR), the National Patient Register (LPR), Danish National Health Service Register for Primary Care (SYSI), and the Death Records. Additional sources of administrative data from Statistics Denmark allows us to get family background information. Nurse records from Danish municipalities provides us with mental health data and nurse visits data. Our unit of observation is the birth for maternal outcomes, but when evaluating children's outcome the number is slightly larger because of a few number of multiple births. For details on our control variables see Appendix B).

3.1 Pre-pregnancy BMI

The MFR includes a measure of mothers' pre-pregnancy BMI. It is calculated based on reported weight and height at the mother's first pregnancy consultation taking place at her family GP around pregnancy week 6-10. The available measure disregards extreme cases where the mother's weight is either lower than 30 kg or higher than 249 kg or her height is either less than 50 cm or taller than 210 cm. We additionally put BMI values>44 to missing (0.05%). We can observe the mother's pre-pregnancy BMI for 92% of all pregnancies recorded in the MFR during the years 2011-2017 (our main sample period) and which passed the EDD with at least 7 days. Since the available measure of pre-pregnancy BMI is 3-digit, we treat it as a continuous measure.

3.2 Instrument for early induction

To obtain a measure of the intended treatment when complying with the guidelines, we construct an measure of *early* induction. The medical birth register contains information on time of birth and whether labor was induced, but not the *time* of labor induction. Instead we use information from LPR and follow Hedegaard et al. (2015)'s Danish medical study in defining (medically) induced labor as any of the following procedure codes: BKHD2, BKHD20, BKHD20A, and BKHD21.[11] Our indicator of early induction is equal to one when any of these procedure codes is registered between 7-9 days after the EDD. We allow for the fact that inductions tend to be postponed on busy days (Maibom et al., 2021) but require that the registered inductions take place prior to the point in time where mothers below the BMI threshold potentially face recommendations of routine induction too.

3.3 Immediate health & birth outcomes

We assess the impact of the intervention on available outcomes commonly found in the medical literature on induction (see Appendix A for a review). From the Danish Medical Birth Register, we extract information on pregnancy length, whether labor was induced, the use of epidural, vacuum extracttion, birth weight, birth weight $\geq 4,5$ kg (macrosomia), and general measures of the child's health (Apgar score<7 at 5 minutes, admission to the Neonatal Intensive Care Unit (NICU), and uterus rupture). Note that while a higher birth weight may be linked to beneficial longer-term outcomes (Behrman and Rosenzweig, 2004; Black et al., 2007; Figlio et al., 2014; Kreiner and Sievertsen, 2020), a birth weight≥4,5 kg includes risks for both the mother and child and is thus not considered beneficial.[12]

We supplement with additional information from LPR. First, we use information on

[11] While it is standard to be medically induced in many circumstances, we do not capture pure mechanical inductions. Specifically, we will not include inductions of women with a precious C-section or other major uterus surgery as they are advised only to be mechanically induced because of their increased risk of uterus rupture (DSOG, 2014). In 2014, most hospital use the drug misoprostol for labor induction (Hedegaard et al., 2014). Yet, if the woman's body is already ripe, labor may be induced by rupturing the amniotic sack.

[12] See Appendix 3.1 for a detailed description of high-BMI pregnancies in Denmark and the associated risks - including increased risk of macrosomia (birth weight≥4,5 kg).

emergency C-section (ICD-10 code: O821). Second, we add information on shoulder dystocia (ICD-10 code: O660), lacerations (ICD-10 codes:070), severe lacerations (3rd, fourth as well as high obstetric lacerations (ICD-10 codes: O702, O703 and O714)[13], and postpartum hemorrhage (O72). The latter code records any kind of hemorrhage since 2012 (Rydahl et al., 2019). We interpret it as a proxy of relatively severe bleeding assuming that it is more likely to be recorded the larger the bleeding.

To study rare but severe outcomes, we pool Apgar score at 5 minutes less than 5 with child mortality (encompassing stillbirth and infant death which is defined as death up to a year of age).[14] Still, the index is positive in fewer than 10 observations on each side of the cutoff. Caution is warranted when analyzing effects on such a rare outcome.

3.4 Short-term health outcomes

From the Danish National Health Service Register for Primary Care (SYSI), we extract the number of GP visits for the child during the first year of life. To obtain the number of nurse-visits during the first year of life as well as a detailed measure of mother's mental health, we leverage data from the Danish universal nurse home visiting program (NHV). The goal of those Nurse visits are to monitor family health including parental mental health, support infant development, and refer families with needs to other specialists. The program proposes up to five universal nurse visits within the first year after childbirth. Additional visits may be scheduled by the nurse when needed. In the spirit of Hirani et al. (2020), we use three outcomes to measure nurse visits: universal visits (out of five), additional nurse visits, and a total measure summing both universal and additional nurse visits.

During these visits, mothers are offered a mental health screening by a standard home visit by the nurse around 2 months after birth. The most frequently used instrument for screening is the Edinburgh Postnatal Depression Scale (EPDS) (Cox and Holden, 2003; Smith-Nielsen

[13] results are robust to excluding obstetric high vaginal laceration (O714)

[14] The Apgar score<4 at 5 minutes outcome is more common in the medical literature, but to avoid having too few observations to comply with the rules and regulations of Statistics Denmark, we use the cutoff below 5. Using the below 4 cutoff does not change our result.

et al., 2018) ranging from 0-30 with 30 being the most adverse outcome. We create an indicator for the score being strictly above 11 (12 or more) which is the cutoff for detecting moderate to severe postpartum depression (in the sense of ICD-10) in the Danish version of the scale (Smith-Nielsen et al., 2018). We additionally consider the total EPDS score. We are only able to obtain data on nurse visits from 62 out of 98 Danish municipalities. For the EPDS scores, we further need to restrict the sample to the 58 municipalities that systematically register their screening for maternal mental health problems.

3.5 Sample selection and size

Our initial sample is the universe of 472,051 births between 2011 and 2017 in Denmark. Restricting the sample to births with information on gestational length and maternal BMI yields a sample of 456,321 births. Next, we restrict to all births taking place 7 days after the EDD or later (n=96,187). Finally, we also restrict on the mother being less than 40 years old and without a GDM[15] diagnosis before the due date, as these mothers are recommended early induction regardless of their pre-pregnancy BMI. Our final sample hence contains 91,743 births.[16] Descriptives on our final sample as well as data on how risk factors vary across the BMI cutoff are presented in Appendix B. We also build two placebo samples using the same restrictions. The first sample includes births from 2004 to 2009, before high-BMI mothers were recommended for early induction.[17] The second sample includes births from 2011 to 2017 but is restricted to cases with a gestational length of less than 7 days past the EDD.

[15] GDM is identified through ICD-10 code: O24.4 in the LPR.

[16] Main results are also robust to excluding mothers with prior C-section who recommended to have given birth before reaching 10 days after the EDD.

[17] We exclude the year of 2010 to factor out any potential announcement effects. While the early induction guidelines were implemented in 2011, the first draft of the guidelines was published already in 2009 and Zizzo et al. (2017) report that a few hospitals already had an early induction policy for high-BMI women in 2010. Results are robust to including 2010 in the placebo-period in any case.

4 Methodology

4.1 Empirical methodology & interpretation

Our identification strategy relies on local randomization in those who are offered early induction because of a high BMI and those who do not. In particular, we exploit that national guidelines discontinuous change at the cutoff of having a BMI of least 35 using a RD design. Our reduced-form equation of interest models the maternal and neonatal outcomes as a flexible function of the forcing variable (BMI of mother i) and a "jump" at the guideline-induced threshold:

$$y_{ihtj} = \gamma_0 + \gamma_1(BMI_{iht} \geq 35) + g(BMI_{iht}) + \Phi' X_{iht} + \phi_h + \tau_t + \psi_m + \mu_d + \epsilon_{iht} \quad (1)$$

Where y_{iht} is the considered outcome observed in calendar year j for mother or child i, giving birth/being born at hospital h in calendar year t, month-of-birth m, and day-of-week d. X_{iht} is a vector of control variables and g() is a smooth function of the mother's pre-pregnancy BMI which is allowed to differ at each side of the cutoff. Our main sample includes all pregnancies going at least 7 days after the EDD to restrict attention to those affected by the guidelines. In the same spirit, we exclude mothers with GDM or who are older than 40. As described in Section 2, these conditions are in themselves advised to be induced at 7 days after the EDD regardless of BMI. Moreover, we control for the following pregnancy variables: having a multiple birth, previously having given birth to a large child (birth weight $\geq 4,5$ kg, previously having had a C-section, first-time mother, insufficient fetal growth, and pre-eclampsia. Most women with these conditions are recommended giving birth prior to 7 days after the EDD, so they should not affect our identifying variation but may still impact the time and method of delivery. We measure all our pregnancy control variables prior to the EDD to avoid having bad controls (some of these conditions may be discovered at the surveillance taking place around 7 days after the EDD). As specified in Appendix B, to increase precision, we additionally control for parental background characteristics: a dummy if either parent graduated from college, monthly joint parental income in 2024 US dollars, maternal age, a dummy which equals one if either parent was born in Denmark, a dummy for whether parents cohabit, as well dummies for missing information on any of the control variables. Our set of controls also include hospital fixed effects, ϕ_h, to capture local variation in practices and year fixed effects, τ_t, to capture general time trends. Moreover, we include month-of-birth fixed effects, ψ_m, to capture seasonal variation and day of the week of birth fixed-effects, μ_d to account for the fact that certain birth interventions are more likely to be scheduled at certain days of the week to add precision to our results.

Our reduced form estimate of γ_1 captures the average treatment effect (ATE) effect of *offering* routine induction at 7 days after the EDD instead at 10-13 days later for mothers with a BMI just above 35. Notice that compliance is not perfect, so alternatively it can be interpreted as the intention-to-treat (ITT) effect of being induced early (7-9 days after the EDD). This is a policy-relevant parameter for policy-makers considering moving the time for routine induction early for high-BMI women.

To account for non-compliance and get at the causal impact of being induced early when having a BMI close to 35, we turn to a fuzzy RD design which re-scales potential jumps at the cutoff in our key outcomes by the jump at the cutoff in the probability of being induced early (defined between 7-9 days after the EDD as specified in Sub-Section 3.2). Formally, this is equivalent to instrumenting a dummy for 'Early Induced Labor' by the jump in the guideline-induced BMI threshold:

$$y_{ihtj} = \gamma_0 + \beta_1 \text{Early Induced Labor} + g(BMI_{iht}) + \Phi' X_{iht} + \phi_h + \tau_t + \psi_m + \mu_d + \epsilon_{iht} \quad (2)$$

Here, 'Early Induced Labor' is a dummy which equals 1 if the mother giving birth was induced between 7 days to 9 days after the EDD. The fuzzy RD estimates are IV estimates of β_1 from Equation 2. It captures the local average treatment effect (LATE) at the threshold:

the effect stemming from being induced at 7-9 days after the EDD because the mother's pre-pregnancy BMI is just above the 35 cutoff.

Among those who are induced early because of mother's pre-pregnancy BMI, some will have a changed mode of labor (induced instead of spontaneous) as well as a reduced gestational length. Others will only have a reduced gestational length as they would have been induced anyway, but at a later point in time had their BMI been just below the cutoff. Any impact of treatment (early induction) may run either through reduced gestational length or a combination of changed mode of labor and reduced gestational length.

Formally, we estimate our empirical models using local polynomial regression using Calonico et al. (2017)'s -rdrobust- command. In all of the analyses, we display bias-corrected estimates with the nonparametric robust standard errors clustered at the yearly hospital level. In Appendix C.3, we show that for many of our outcomes we do not have enough power to use data-driven bandwidth methods. While the data-driven bandwidth is usually between 3-5, we opt for a slightly larger manual bandwidth of 7 and a polynomial order of one. Besides powering reasons, there is a theoretical rationale for the choice of a bandwidth of specifically 7: it ensures that all women in our sample undergo at least one glucose test to detect cases of GDM during pregnancy (see Appendix A.2 for details on high-BMI pregnancies in Denmark). Note that these specification choices are only relevant in our main specification: we indeed check that our results are robust to changes in bandwidth, kernel and polynomial order in Appendix C.5 and C.6. We also compute p-values adjusted for multiple hypothesis testing using Romano-Wolf's procedure available through the package created by Clarke et al. (2019).

Validity of our ITT estimates requires that the continuity assumption is met which ensures local variation across the 35 cutoff. This requires that in absence of the treatment, the potential outcomes would have evolved smoothly around mother's pre-pregnancy BMI equal to 35. Our Fuzzy RD design additionally requires that: (a) the BMI$\geq$35 cutoff implies a substantial increase in labor induction between 7 and 9 days after the EDD (strong first-stage), (b) the monotonicity assumption holds, and (c) the exclusion restriction holds. We are not particularly worried about the monotonicity assumption in our context (we do not find it likely that a mother who refuses labor induction at 7-9 days after the EDD because her pre-pregnacy BMI is at least 35 would have accepted that had her BMI been lower than 35). Instead, the exclusion restriction provides a bigger challenge by requiring that the 35 cutoff may only impact mothers through their changed early induction status. The following sections are devoted to addressing the main threats to the key identifying assumptions in our setting.

4.2 Threats to identification

Our key assumption for obtaining valid ITT estimates is the local randomization around the cutoff (continuity) assumption. First, it excludes manipulation of the running variable. In theory, a mother might try to manipulate her weight to be either above or below the BMI threshold in which she is more likely to be early induced. In practice, such manipulation is not likely since the mother's pre-pregnancy BMI is reported in the beginning of the pregnancy to the family GP whereas induction guidelines are typically discussed later during pregnancy with the midwife. Empirically, Figure 2 rules out manipulations of that kind.

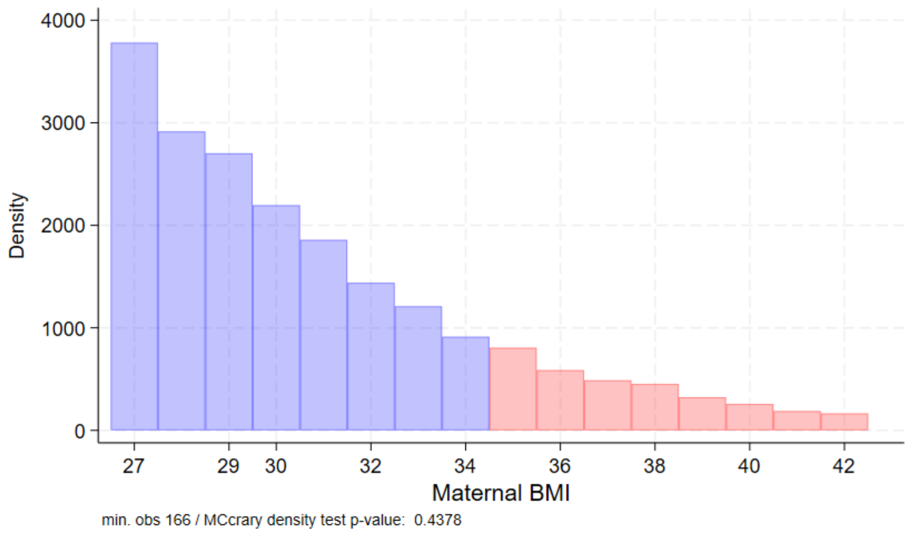


Fig. 2 Density of births with a given maternal BMI
Notes: The figure shows the density of births around the 35 maternal BMI threshold. Blue bars indicate that the maternal BMI is below the threshold, red bars that the maternal BMI is above. Thd sample includes all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later and where the mother had age<40 and no registry of gestational diabetes. While our maternal BMI data is available up to three digits, we collapsed data at the unity. A more disagregated figure is presented in in Appendix C.4.

Second, the continuity assumption also implies the absence of discontinuity in any other observables or unobservables around the threshold had labor induction at 7 days after the EDD not been implemented. First, we check for any discontinuity in our observable characteristics in Table 1. This table is overall reassuring for our design. First significant imbalances are rare and overall small. In our main sample in particular, there are no significant imbalances at the 5 % level. Nevertheless, being above the 35 threshold may be associated with a higher chance of being a first-time mother. This is especially pronounced in the EPDS sample, where we also see a slightly reduced risk of having had a previous large child with birth weight $\geq 4,5$ kg (macrosomia). Thus, we could worry that higher parity mothers are less likely to be allowed to await a spontaneous birth until 7 days after the EDD if they faced complications in prior pregnancies (e.g. prior excessive birth weight) while also having a BMI ≥ 35. If mothers with a BMI ≥ 35 are treated differently earlier in pregnancy than similar mothers with a BMI just below the cutoff, this threatens our design.

In particular, being above the 35 threshold may in some cases trigger a consultation with an obstetrician during pregnancy, additional (but uncertain) weight estimates of the child, and prophylactic epidural use (see Appendix A.2 for details). While the latter health policy does not impose a threat to the continuity assumption, it could make us unable to disentangle the impact of early induction from the impact of increased likelihood of epidural use which instead impose a threat to the exclusion restriction.

We address whether our results are driven by other health policies at the 35 cutoff (either through selection or as a consequence of epidural use) in Section 6, by exploiting the fact that other relevant health policies are in place for all high-BMI mothers and not only those who have a pregnancy going at least 7 days after the EDD. Thus, we redo our main result for a subsample of high-BMI pregnancies which do not reach 7 days after the EDD. If other health policies at the threshold are driving our main results, we expect to see discontinuities in the outcomes around the 35 BMI cutoff in this sample too and/or a discontinuity in the risk of having a planned C-section. Finally, our results are robust to excluding control variables (available from the authors), changes in bandwidth, kernel and polynomial order (Appendix C.5-C.6).

Tab. 1 Balance of the observables

	Sharp	Sharp	Sharp	Mean Control Group (main sample)
Any parent graduated college	-0.01	-0.01	0.07	0.45
	(0.031)	(0.042)	(0.058)	
Missing education data	-0.01	-0.01	0.00	0.09
	(0.018)	(0.021)	(0.024)	
Parental income, monthly (US$)	71.57	-108.97	148.72	5818.57
	(185.928)	(233.563)	(301.738)	
Missing income data	-0.02	-0.02	-0.02	0.06
	(0.016)	(0.017)	(0.019)	
Parents cohabit	0.00	0.02	0.04	0.84
	(0.021)	(0.025)	(0.032)	
Missing cohabition data	0.01	0.01	0.00	0.01
	(0.006)	(0.008)	(0.002)	
Any parent Danish	-0.00	-0.01	-0.03	0.88
	(0.018)	(0.021)	(0.020)	
Missing ethnicity data	0.01	0.01	0.00	0.01
	(0.006)	(0.008)	(0.002)	
Preeclampsia	0.00	-0.00	-0.01	0.01
	(0.007)	(0.009)	(0.007)	
Insufficient growth	-0.00	-0.00	-0.01	0.01
	(0.007)	(0.009)	(0.011)	
Prior C-section	-0.01	-0.01	0.02	0.09
	(0.016)	(0.021)	(0.033)	
Prior birth weight $\geq 4.5kg$	0.00	-0.01	-0.03*	0.02
	(0.008)	(0.009)	(0.012)	
First-time mother	0.05	0.06	0.12*	0.48
	(0.028)	(0.038)	(0.053)	
Parity data missing	0.00	0.00	0.00*	<0.01
	(0.006)	(0.008)	(0.001)	
Maternal age	-0.02	-0.14	0.18	30.15
	(0.286)	(0.355)	(0.551)	
Missing Maternal age	0.00	0.01	0.00	0.01
	(0.004)	(0.006)	(0.005)	
Missing EPDS	0.00	-0.02	-	0.70
	(0.036)	(0.049)	(.)	
Missing nurse visits	0.03	-	-0.01	0.36
	(0.041)	(.)	(0.004)	
Sample restriction	All	Nurse visit	EPDS	

Notes: Standard errors are robust and clustered at the hospital-year level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.This table displays sharp RDD estimates which are obtained using our favorite specification with polynomial order of one and an epanechnikov kernel. The bandwidth is manually set to 7 for reasons explained in subsection C.3. "Mean in the control group" refers to the average outcome for births on the left-side of the bandwidth (Maternal BMI 28-35). The set of controls is described in subsection B.1. Each cell displays the coefficient of a single regression with the observable as the outcome, the BMI as a running variable, and the cutoff at 35. In the sample restriction row, "All births" refer to all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later, whereas "Nurse visits" and "EPDS score" refer to the respective subsets of this sample for which nurse visits from the NHV and mental health data are available.

5 Results

5.1 Treatment & first stage

Women with a BMI of at least 35 but no other comorbidities are offered routine induction at 7 days after the EDD instead of at 10-13 days after. We model the intervention as being induced early (7-9 days after the EDD). Our sharp RDD-estimates of being above the BMI threshold on the risk of being induced early thus yields our first-stage estimate. One of the identifying assumptions for our fuzzy RDD design is a strong first stage. Figure 3 confirms a notable jump in the risk of early induction of almost 20 percentage points (p.p) at the BMI threshold which is not present in the placebo period when the guidelines were not yet issued. Relative to baseline, this corresponds to a 163% increase. Appendix Table B1 further shows our formal first-stage across the three different samples for our preferred specification: 16.5 p.p. in our main sample, 15.7 p.p. in the nurse visit sample, and 20 p.p in the EPDS score sample. In all cases, our estimates are statistically significant at the 1% significance level. Reassuringly, Appendix Figure B2 shows that our first-stage result are relatively stable across different bandwidth choices too. For bandwidth choices larger than 7, results on the early induction rate show a tendency to obtain larger estimates. Excluding any control variables yield an estimate close to our main results of 15.9 p.p. and including a second-order polynomial reduces the estimates to around 12.3 p.p. (results available from the authors).

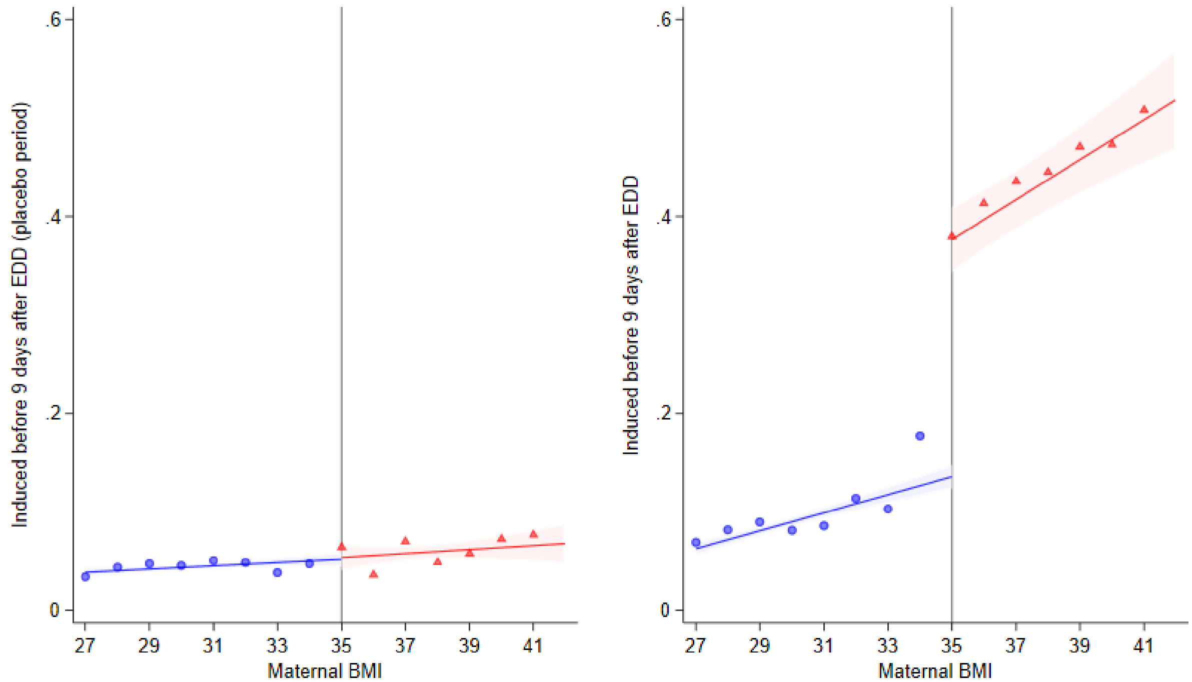


Fig. 3 First stage in placebo period (2004-2009) and when the guidelines are binding (2011-2017)

Notes: The bins are the average outcomes for each BMI unit. The lines are fitted lines from an OLS regression on all births where the mother gave birth at least 7 days after the EDD, have an age<40, and no registry of gestational diabetes. Confidence intervals are at the 90 percent level and built from heteroskedastic robust SEs clustered at the hospital-year level. The right panel sample is from the period after the early induction guidelines were issued (2011-2017), while the left panel sample is from the placebo period before the guidelines were issued (2004-2009).

Having confirmed that the risk of being induced early increases significantly at the cutoff, we next turn to the two parts of the intervention: a changed mode of labor and a reduced gestational length. Consider a mother with a BMI of 35 following the recommendations of early induction. Had her BMI been below the threshold, she may have had spontaneous onset of labor or she would have been induced anyway, but at a later point in time. In either case, gestational length would likely have been higher.

In Figure B1 in the Appendix, we first show that the ITT estimate on the probability of being induced at any point in time. We see that the baseline risk of being induced at any point in time is naturally higher, around 40-50%, but there is still a clear and significant jump at 35 indicating that the risk of being induced at any point in time increases with a little less than 10 p.p. (8 p.p. in our preferred specification) (see Appendix Table B3).Thus, having a BMI above the threshold increases the risk of induced labor at any point in time indicating an impact on the mode of labor.

When evaluating the ITT effect of moving routine induction forward to 10-13 days after the EDD instead of 14 days after, Gregersen (2024) find that the risk of labor induction (any time) similarly increases with around 10 p.p. for low-risk first-time mothers going at least 7 days after the EDD. Yet, existing RCTs reporting information on the induction rates (at any time) report a larger impact for the general population of pregnancies.[18] It makes sense that compliance with early induction is higher in an RCT setting where women volunteer to participate and thus accept assignment in either treatment or control. In contrast, the Danish national guidelines are recommendations which also describe how to follow and monitor those who prefer to await spontaneous labor instead of accepting routine induction (DSOG, 2011a). Interestingly, when we address compliance (using our fuzzy specification), we find that the probability of ever being induced increased by around 51 p.p. for those who are induced early because they have a BMI close to 35 - see the fuzzy estimates in Appendix Table B3. This estimate is very close to some of the estimates from RCTs on early induction for the general population of low-risk pregnancies (Alkmark et al., 2020; Heimstad et al., 2007), but smaller than what Place et al. (2024) reports.

Moving on to our ITT estimate on gestational length, Figure 4 shows a reduction of around 0.6 days and no discontinuity in the placebo period. Again our ITT estimate is smaller than those of existing RCTs for low-risk pregnancies: they report a reduction in gestational length of around 3 days (Alkmark et al., 2020; Gelisen et al., 2005). Yet, using our fuzzy specfciation to account for compliance by rescaling our ITT, we find that the guidelines reduce the length of the pregnancy with around 4 days through those who are induced early because of their BMI (see Table B2 in the Appendix).

[18]Specifically, the early induction group (at 7-9 days after the EDD) face around 49-53 p.p. higher risk of being induced relative to the control group (induction at either 14 or 20 days after the EDD) (Alkmark et al., 2020; Heimstad et al., 2007). In one case, the difference in induction risks seems as high as 75.7 p.p. (Place et al., 2024).

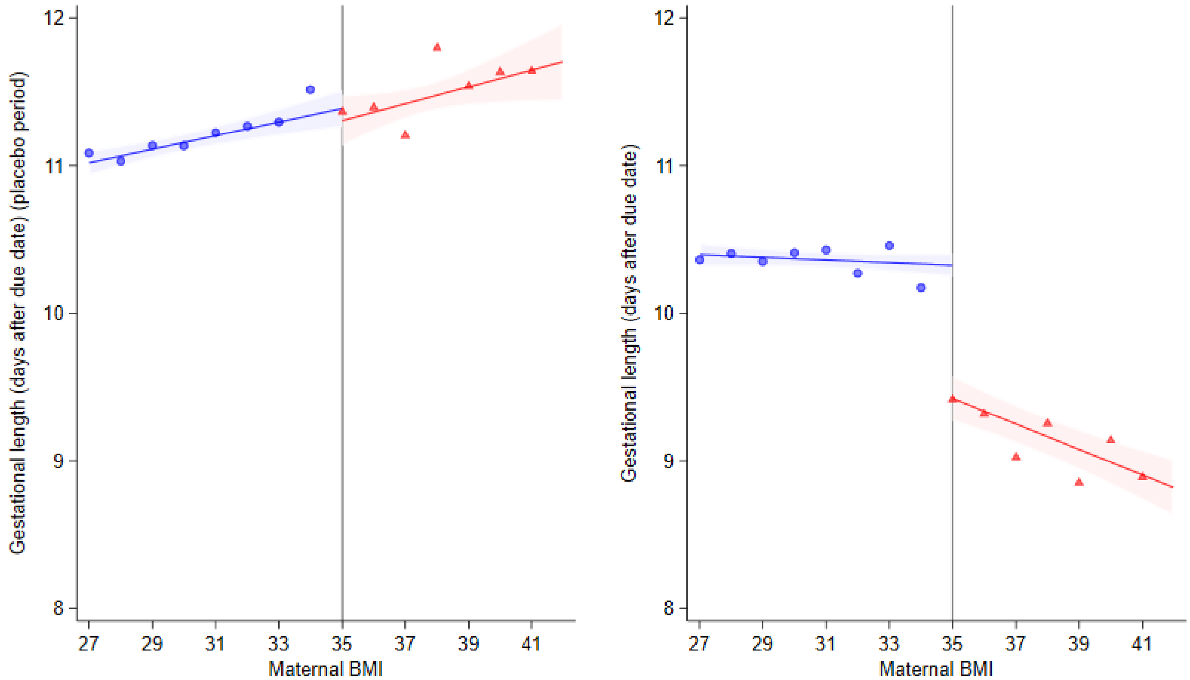


Fig. 4 Gestational length (days) in placebo period (2004-2009) and when the guidelines are binding (2011-2017)

Notes: The bins are the average outcomes for each BMI unit. The lines are fitted lines from an OLS regression on all all births where the mother gave birth at least 7 days after the EDD, have an age<40, and no registry of gestational diabetes. Confidence intervals are at the 90 percent level and built from heteroskedastic robust SEs clustered at the hospital-year level. The right panel sample is from the period after the early induction guidelines were issued (2011-2017), while the left panel sample is from the placebo period before the guidelines were issued (2004-2009).

5.2 Immediate health & birth outcomes

Table 2 presents our main results on immediate health and birth outcomes. To account for multiple testing, we compute Romano-Wolf p-values. Despite the relatively low power in our setting (see Appendix C.3), all significant regression results remain robust at least at the 10% level. Graphs showing the reduced forms are provided for cases where we find a statistically significant impact (Figure 5).[19]

Regarding the mode of delivery, we find no effect on vacuum extraction and emergency C-section, which is consistent with the majority of existing literature on moving routine induction early (detailed in Appendix A). Having a BMI $\geq$35 increases epidural use by 6.3 p.p. However, part of this effect may stem from separate guidelines on epidural use, as discussed in Section 4.2. We address this issue in Section 6. Indeed, our ITT effect on epidural use is larger than estimates found in studies of earlier routine induction for low-risk pregnancies (Alkmark et al., 2020), sometimes even concluded to have no impact (Gregersen, 2024; Place et al., 2024).

Next, we examine the impact of early induction on birth weight. We find that being above the BMI threshold results in an average birth weight reduction of 30 grams and a 2.8 p.p. reduced risk of macrosomia, defined here as a birth weight exceeding 4,5 kg. Although our ITT effect on birth weight is not statistically significant, the discontinuity in birth weight at the cutoff is clearly visible graphically and not present in the placebo period, as shown on Figure 5, panel b). While the estimate is marginally significant in the fuzzy specification, it is not robust to multiple testing. The ITT effect on macrosomia (birth weight>4.5kg) closely matches results from RCTs included in Alkmark et al. (2020) (-2.78 p.p.), but is much larger than what is found by Gregersen (2024) (-0.29 p.p). The average reduction in birth weight in our study (30 grams) is larger than what is found by Gregersen (2024) (10 grams), but smaller than reductions found in RCTs, which generally ranges from 59-65 grams (Place et al., 2024; Alkmark et al., 2020) to as high as 125 grams (Gelisen et al., 2005). Our Fuzzy

[19]Except for the complication index which is positive for less than 10 observations on each side of the cutoff.

RD estimates imply that women induced early due to a BMI slightly above 35 experience a 16 p.p. reduction in macrosomia risk and an average reduction in birth weight of 175 grams. For these women, the birth weight effects are relatively large.

Interestingly, we also observe an impact on health outcomes potentially influenced by a birth weight $\geq$ 4,5 kg[20] including lacerations (broad definition), hemorrhage (broad definition) and shoulder dystocia, a childbirth emergency where the baby's shoulder gets stuck behind the mother's pelvic bone after the head is delivered. These findings contrast to some extent with existing studies estimating ITT effects of advancing routine induction for low-risk pregnancies. Gregersen (2024) find no impact on lacerations (broad definition), and most RCTs report no impact on hemorrhages (Alkmark et al., 2020; Heimstad et al., 2007). Despite not reporting ITT estimate, Place et al. (2024) report reduced hemorrhage in their RCT including a little less than 400 paritcipants. The prevalence of shoulder dystocia is low. While no impact is found by Wennerholm et al. (2019), Gelisen et al. (2005) report a reduced risk. Moreover, a meta-analysis of four RCTs evaluating induction for suspected macrosomia (birth weight$\geq$4.5 kg) report that induction before the EDD reduces average birth weight by around 178 grams, as well as the risk of shoulder dystocia and fracture (Boulvain et al., 2016). Given that they evaluate a treatment taking place as early as three weeks prior to the EDD, it is actually rather impressive that our results indicate an average reduction of birth weight of a rather similar magnitude among those who are induced early because of their BMI. While our findings on lacerations and hemorrhage may reflect rather mild incidences, shoulder dystocia is a severe outcome and significant in our analysis. Women induced early due to a BMI just above 35 experience a reduction in shoulder dystocia risk of 3.4 p.p., which is substantial relative to the baseline mean of 1%. Because the outcome is very rare however, the corresponding plot is noisy.

Another potential complication of having a high birth weight (or shoulder dystocia) is uterus rupture. However, the risk of uterus rupture may also increase due to the induc-

[20]See (Stotland et al., 2004) for more information on the risk factors associated with a birth weight $\geq$ 4,5 kg (macrosomia).

tion process too, leading to an ambiguous predicted effect. Overall, we find no statistically significant impact but observe a consistently negative coefficient.

Shoulder dystocia can cause low fetal oxygen levels, which we cannot measure directly but may manifest as a low 5-minutes Apgar score.[21] However, the predicted effect on having an Apgar score<7 at 5 minutes is ambiguous, as it could also be influenced by the induction process itself. Similarly, we find no effect on the risk of having an Apgar<7 at 5 minutes, but the coefficient is small and positive. Similar findings are reported by Alkmark et al. (2020); Gelisen et al. (2005); Gregersen (2024); Place et al. (2024).

Turning to more severe health outcomes, , we first find a significant reduction in our measure of severe complications (a dummy capturing either stillbirth, infant mortality or Apgar score<5 at 5 minutes). The primary motivation for advancing the recommended time for routine induction was to reduce the risk of stillbirth; however, this outcome is too rare to analyze independently. While most studies are unable to detect any statistically significant impacts on mortality, the numbers generally point towards a reduction (see Appendix A for details). Caution is needed with outcomes so rare, as it is challenging to draw statistically sound conclusions on those. The same caution applies to uterus rupture, and the complication index with fewer than 10 observations on each side of the cutoff where the index is positive. While the discontinuity is significant in the placebo period, the coefficient is positive. Thus, while exercising caution, our results may tentatively suggest a reduced risk, aligning with the existing literature. Yet, no significant effects are detected on mothers' hospital nights or NICU admissions. While Alkmark et al. (2020) report a reduction in NICU admissions, most other studies do not (Heimstad et al., 2007; Gelisen et al., 2005; Gregersen, 2024; Place et al., 2024).

[21]See e.g. https://www.ncbi.nlm.nih.gov/books/NBK470427/ for general information about shoulder dystocia.

Tab. 2 Results on immediate health & birth outcomes

	Sharp	Fuzzy	Obs	Mean Control Group	RW pvalue (fuzzy spec.)
Epidural	0.063**	0.339**	16342	0.46	0.0396
	(0.029)	(0.133)			
Emergency C-section	-0.003	-0.020	16342	0.06	0.999
	(0.012)	(0.054)			
Vaccuum extraction	-0.004	-0.016	16342	0.08	0.999
	(0.015)	(0.065)			
Birth weight≥4,5 kg	-0.028*	-0.163**	16342	0.08	0.0594
	(0.015)	(0.069)			
Birth Weight (grams)	-30.051	-176.401*	16342	3850.19	0.2772
	(22.892)	(102.183)			
Shoulder dystocia	-0.006*	-0.034**	16342	0.01	0.0990
	(0.003)	(0.016)			
Lacerations	-0.047*	-0.256**	16342	0.29	0.0792
	(0.025)	(0.112)			
Lacerations (3nd degree and higher)	0.002	0.008	16342	0.03	0.999
	(0.009)	(0.042)			
Hemorrhage	-0.070**	-0.345***	11405	0.57	0.0099
	(0.030)	(0.115)			
Complications index	-0.003**	-0.014**	16342	< 0.01	0.0594
	(0.001)	(0.006)			
Uterus Rupture	-0.002	-0.011	16342	< 0.01	0.9802
	(0.003)	(0.015)			
Share sent in NICU	0.002	0.007	16342	0.06	0.999
	(0.014)	(0.063)			
Apgar <7	0.001	0.006	16342	0.01	0.999
	(0.005)	(0.021)			
Hospital nights (after birth)	-0.009	-0.102	16342	2.18	0.999
	(0.097)	(0.439)			
Covariates	Yes	Yes			

Notes: Standard errors are robust and clustered at the hospital-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table displays RDD estimates which are obtained using our favorite specification with polynomial order of one and epanechnikov kernel. The bandwidth is manually set to 7 for reasons explained in subsection C.3. Coefficient plots showing the robustness of significant results to varying bandwidths, polynomial order, and kernel can be found in Appendix Table C.5. "Mean in the control group" refers to the average outcome for births on the left-side of the bandwidth (Maternal BMI 28-35). The set of controls is described in Sub-Section B.1. For the hemorrhage outcome, the number of observations is lower than for the other outcomes because we drop births before 2013 to account for a change in the definition of hemorrhage in 2012. Last column reports the Romano-Wolf p-value for the fuzzy RD specification, computed with all outcomes in the table.

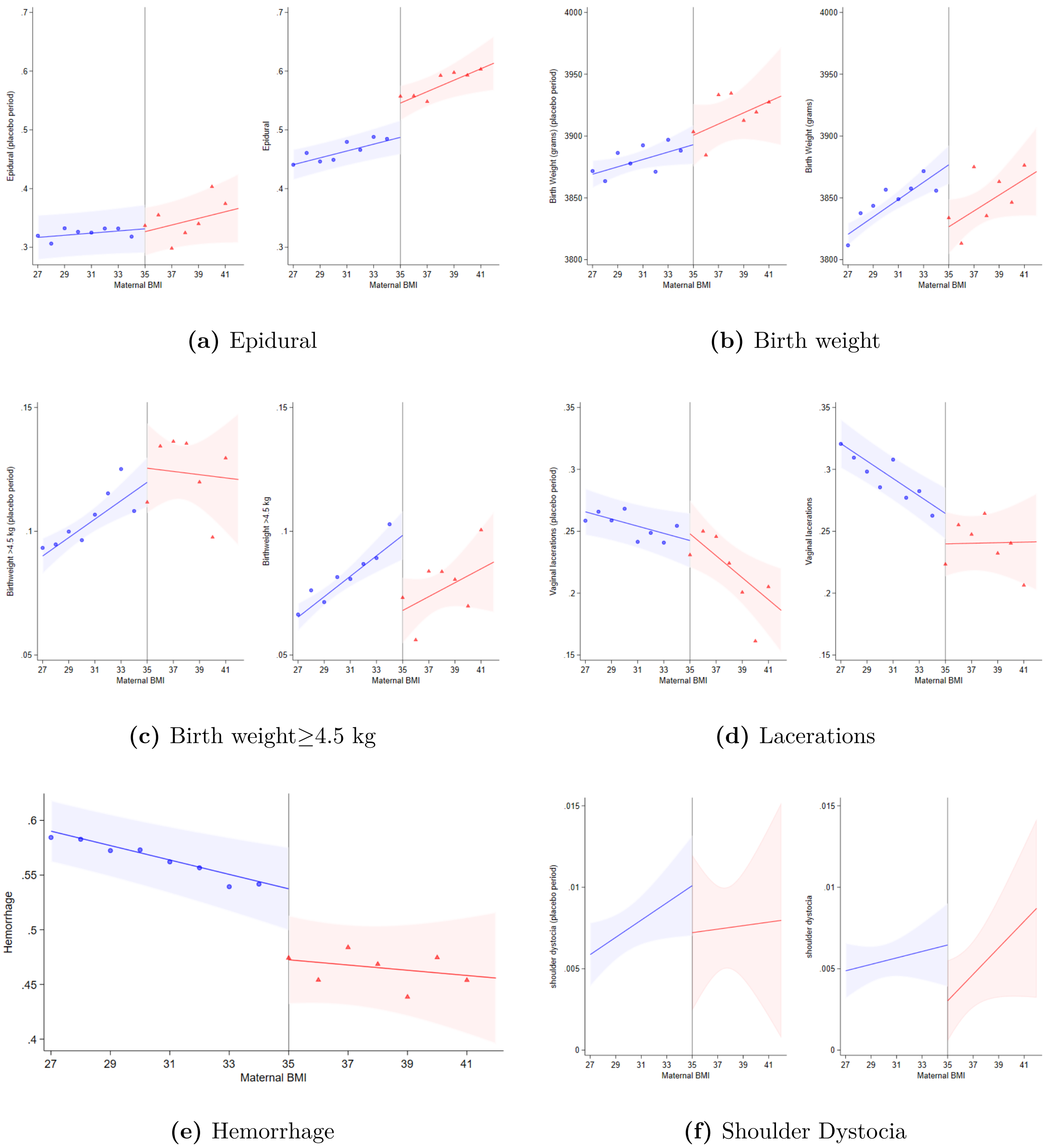


Fig. 5 Discontinuity in certain immediate health outcomes around the 35 BMI cutoff

Notes: Bins are average outcomes for each BMI unit and are only plotted provided a minimum number of obs. Lines are fitted lines from an OLS regression on all births where the mother gave birth at least 7 days after the EDD, have an age<40, and no registry of gestational diabetes. Confidence intervals are at the 90 percent level and built from heteroskedastic robust SEs clustered at the hospital-year level. Samples in both panels consist of all births in Denmark that occurred at 7 days after the EDD or later. Right panel sample is from the period after the early induction guidelines were issued (2011-2017), while left panel sample is from the placebo period (2004-2009). For hemorrhage, placebo period is not shown because outcome definition changed in 2012 - see Section 6 for an alternative placebo test.

5.3 Short-term health (mental health, GP & nurse visits, social outcomes)

In this section, we examine whether the beneficial immediate health effects extend over a longer period. Table 3 summarizes our main results, while Figures 6-7 present the reduced form results graphically for significant results. Again, all significant regression results remain robust at least at the 10% level when adjusting for multiple hypothesis testing using Romano-Wolff p-values.

We find evidence of reduced healthcare usage during the first year of life due to early induction. Specifically, being induced early reduces nurse visits (both total and universal) by around two during the first year. While marginally significant in the regression, the result on *total* nurse visits is not convincingly distinguishable in the plot (see Appendix Figure B3), and is not robust to adding a higher polynomial order (see coefficient plot C.6). However, the effect on *universal* visits is highly statistically significant, stable across specifications (see coefficient plot C.6) and the discontinuity is fairly clear in Figure 6. In many cases, nurses are typically contacted prior to the GP (i.e. to schedule additional visits because of concerns regarding the child's development, problems with breastfeeding etc.). By contrast, GPs need to be contacted in more severe cases requiring prescription medication or specialists referrals. One might suspect that induced mothers are simply substituting nurse visits with GP visits due to more serious conditions (e.g., postpartum infections due to induction). However, the small and insignificant effect on additional nurse visits—those scheduled when the mother needs extra care— does not support the hypothesis that induced mothers develop more conditions. Moreover, although insignificant, the effect on GP visits is also negative (Fuzzy:-2 GP Visits). Appendix Figure B5 also shows a suggestive drop in GP visits. Overall, our findings point to a reduction in healthcare usage during the first year of life, suggesting that the benefits of early induction for high-BMI mothers may extend beyond birth.

Next, we show early induction improves maternal mental health as evidenced by lower EPDS scores. Our ITT estimates indicate that the likelihood of having a high EPDS score

(>11) reduces by around 6 p.p., and the overall EPDS score reduces by approximately 0.9 points corresponding to roughly 24% of a standard deviation. Mothers induced early due to having a BMI just above 35 show a substantial reduction relative to the baseline: more than 27 p.p. reduced risk of having a high EPDS score and an average reduction of more than 4 points - see results on the fuzzy specifications. These findings provide a possible explanation for the reduction in nurse visits: mothers with improved mental health may require less support from nurses. Consistent with this hypothesis, the effect on the number of both total and universal nurse visits is sizable and highly significant in the sample for which we have EPDS data (see Appendix B4). The coefficient for additional nurse visits becomes negative in this alternative sample.

Our results do not support the thesis that the adverse associations between a mother's birth experience and being induced are causal (see Appendix A). Instead, our findings on improved mental health align with results from a RCT on low-risk pregnancies, which document that the majority of women preferred induction at 7 days after the EDD rather than awaiting spontaneous labor (Heimstad et al., 2007). For a population of low-risk first-time mothers, Gregersen (2024) finds no ITT effect on a proxy for mental health, but her measure may capture more severe cases.

Finally, we find no evidence on any effect on social outcomes such as labor supply and fertility up to 2 years after birth. Similarly, Gregersen (2024) finds no impact of moving routine induction forward on fertility for a broader population of low-risk first-time mothers. In contrast, avoidable C-sections may cause a reduction in fertility also reflected in a temporary increase in employment (Halla et al., 2020).

Tab. 3 Short-term outcomes

	Sharp	Fuzzy	Obs	Mean Control Group	RW pvalue (fuzzy spec.)
EPDS>11	-0.058**	-0.273**	4999	0.062	0.0693
	(0.025)	(0.120)			
EPDS Score (0-30)	-0.882**	-4.150**	4999	4.961	0.0693
	(0.402)	(1.876)			
Nurse visit (Total)	-0.377*	-2.198*	10457	6.249	0.0891
	(0.216)	(1.132)			
Universal nurse visit (0-5)	-0.445***	-2.592***	10457	4.585	0.0198
	(0.743)	(0.635)			
Additional nurse visits	0.020	0.129	10457	1.687	0.9703
	(0.193)	(0.991)			
GP Visits (child)	-0.432	-2.554	16342	17.815	0.5941
	(0.501)	(2.278)			
Share Employed (Mother, 1 year post birth)	0.031	0.180	11315	0.840	0.5446
	(0.023)	(0.125)			
Share Employed (Mother, 2 years post birth)	0.018	0.087	11315	0.850	0.9703
	(0.020)	(0.110)			
Prob. birth two years after	-0.005	-0.028	11315	0.031	0.5941
	(0.012)	(0.064)			
Covariates	Yes	Yes			

Notes: Standard errors are robust and clustered at the hospital-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table displays RDD estimates which are obtained using our favorite specification with polynomial order of one and an epanechnikov kernel. The bandwidth is manually set to 7 for reasons explained in subsection C.3. "Mean in the control group" refers to the average outcome for births on the left-side of the bandwidth (Maternal BMI 28-35). The set of controls is described in subsection B.1. Coefficient plots showing the robustness of significant results to varying bandwidths, polynomial order, and kernel can be found in Appendix Table C.5. The number of observations varies across outcomes because of data availability and because for social outcomes variables 2 years after birth, the years 2016 and 2017 were dropped. Last column reports the Romano-Wolf p-value for the fuzzy RD specification, computed with all outcomes in the table.

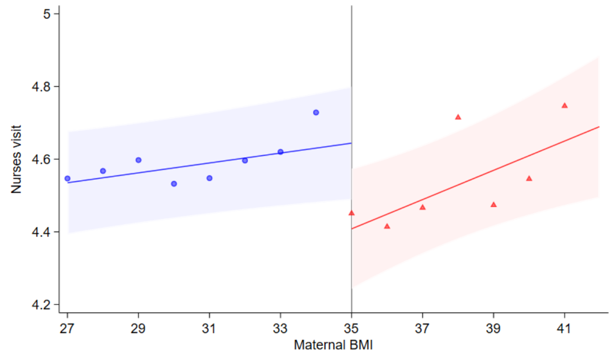


Fig. 6 Impact of the intervention on the number of universal nurse visits (first year after birth)
Notes: The bins are the average outcomes for each BMI unit. The lines are fitted lines from an OLS regression on all births. Confidence intervals are at the 90 percent level and built from heteroskedastic robust SEs clustered at the hospital-year level. The sample consists of all births in Denmark between 2011 and 2017 where the mother gave birth at least 7 days after the EDD, have an age<40, and no registry of gestational diabetes as well as being a part of the Nurse visit sample with a total of n=11,225. No placebo period is shown because the data from the NHV is not available before 2011. An alternative placebo test is however provided in Section 6.

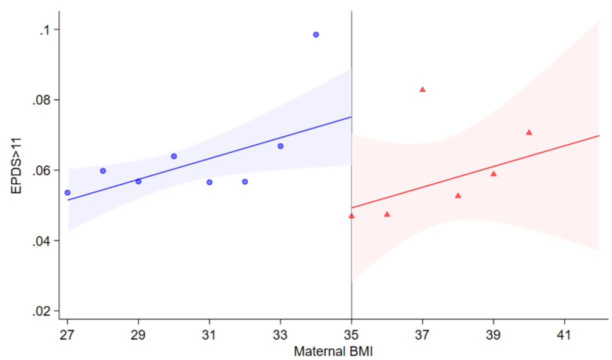


Fig. 7 Impact of the intervention on maternal mental health (EPDS score>11)
Notes: The bins are the average outcomes for each BMI unit. The lines are fitted lines from an OLS regression on all births. Confidence intervals are at the 90 percent level and built from heteroskedastic robust SEs clustered at the hospital-year level. The sample consists of all births in Denmark between 2011 and 2017 where the mother gave birth at least 7 days after the EDD, have an age<40, and no registry of gestational diabetes as well as information on EPDS score with a total of, with a total of n=4999. No placebo period is shown because the data from the NHV is not available before 2011. An alternative placebo test is however provided in Section 6. Appendix B4 presents the same figure using the continuous EPDS score as the outcome, although this approach may be less psychometrically appropriate.

6 Disentangling early induction BMI-based guidelines effects from that of other BMI-based guidelines

The goal of this section is to disentangle the impact of the early induction guidelines from that of other guidelines potentially affecting mothers with a BMI just above 35.

In general, overweight mothers are advised to lose weight prior to becoming pregnant and to reduce their weight gain during pregnancy (see Appendix A.2 for detailed information on overweight pregnancies in Denmark). Thus, on the one hand, healthcare professionals may not substantially change behavior towards mothers with a BMI just above the 35 threshold which speaks in favor of our local randomization identification. Nevertheless, some hospitals schedule a talk with a obstetrician during pregnancy and additional checks focusing on estimating the child's size when a mother's BMI is above the exact 35 threshold. Moreover, epidural may be used prophylactic for women above this threshold too. Thus, we need to rule out than any differences in health outcomes are explained by these other types of health policies.[22]

The additional checks during pregnancy may impose a threat to the continuity assumption if they lead to the discovery of underlying health problems necessitating a birth prior to 7 days after the EDD. In that case, women with BMI of at least 35 included in our main sample would be positive selected relative to women just below the cutoff. Increased epidural prophylactic is not a concern for the discontinuity assumption, but it may make it impossible to disentangle potential health effects stemming from increased epidural use from health effects stemming from the early induction policy thereby invalidating the exclusion restriction.

In order to address these concerns, we exploit the fact that these other health policies effective at the 35 threshold are in place for all pregnancies and not just the ones who pass 7 days after the EDD. Thus, we repeat our main analysis on variables which were significant in

[22]Having a high BMI may restrict access to fertility treatment through the public health care system, but the exact BMI threshold varies across hospitals (between 29 and 40) and across age - and different risk groups too. Thus, the threshold for access to fertility treatment do not coincide with our BMI threshold but exists both above and below the 35 cutoff.

the sharp specification, in a placebo sample of pregnancies giving birth prior to 7 days after the EDD. If any of our results are not driven by the early induction policy, our outcomes should be affected significantly in the same direction for this alternative sample. If our results are driven by positive selection, then, in this placebo sample, we would instead expect mothers above the threshold to have worse health outcomes. We additionally study differences in elective C-section rates to address one potential source of possible positive selection of women with a BMI above the threshold into our sample: talking with an obstetrician during pregnancy could lead to a higher risk of elective C-sections because of the discovery of underlying health issues or increased awareness of risks associated with vaginal births. An elective C-section typically takes place prior to the EDD and such women would thus not be a part of our main sample.

Table 4 below shows the reduced form (sharp) estimates for all outcomes which we found were significantly impacted by earlier routine induction for high-BMI women (including universal nurse visits studied in Appendix), together with results on elective C-section. While the coefficient on elective C-section is positive, it is small and highly statistically insignificant. Thus, we do not find evidence of a positive selection at the exact 35 threshold in our main sample driven by talking with an obstetrician during pregnancy (and then leading to a higher risk of having an elective C-section prior to 7 days after the EDD).

Reassuringly, none of the affected outcomes from our main analysis is statistically affected at the 5 percent level in the sample of mothers who give birth earlier than 7 days after the EDD suggesting that our results are not driven by other health policies taking place at the 35 threshold, not even using epidural prophylactic because of BMI. Moreover, the magnitudes of the estimated coefficients for epidural, gestational length (past the EDD), birth weight, shoulder dystocia, EPDS score and nurse visits (both total and universal) are all much smaller than the ones we found in our main result section. While insignificant, the coefficient on our complications index in the placebo sample is actually similar in magnitude to our main result. Thus, the only result which we could be concerned about is that on the complications index (including Apgar score<5 at 5 minutes and stillbirth).

	Sharp	Sharp	Obs (placebo/main sample)
Elective CS	0.008	-	44065/16342
	(0.010)	-	
Gestational length (Days)	0.012	-0.675***	44065/16342
	(0.503)	(0.145)	
Epidural	0.022	0.063**	44065/16342
	(0.020)	(0.029)	
Birth weight $\geq$ 4.5 kg	0.005	-0.028*	44065/16342
	(0.006)	(0.015)	
Lacerations	-0.015	-0.047*	44065/16342
	(0.015)	(0.025)	
Shoulder dystocia	0.000	-0.006*	44065/16342
	(0.002)	(0.003)	
Hemorrhage	-0.001	-0.070**	30811/11405
	(0.020)	(0.030)	
Complications index	-0.003	-0.003**	44065/16342
	(0.003)	(0.001)	
Nurses Visits (Total)	0.068	-0.377*	27642/10457
	(0.159)	(0.216)	
Universal nurses visits	0.048	-0.445***	27642/10457
	(0.102)	(0.135)	
EPDS>11	-0.031*	-0.058**	12578/4999
	(0.016)	(0.025)	
EPDS Score	-0.039	-0.882**	12578/4999
	(0.250)	(0.402)	
Days after EDD	<7	≥7	
	(Placebo)	(Main Sample)	
Covariates	Yes	Yes	

Notes: Standard errors are robust and clustered at the hospital-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table displays sharp RDD estimates which are obtained using our favorite specification with polynomial order of one and an epanechnikov kernel. The bandwidth is manually set to 7 for reasons explained in subsection C.3. The set of controls is described in subsection B.1. The number of observations varies across outcomes because of data availability. The elective C-section coefficient is not shown in the main sample, for it is only used for a placebo regression in the alternative sample.

7 Conclusion

The number of pre-pregnancy overweight and obese women is rising in many countries. While pre-pregnacy BMI is correlated with adverse birth outcomes for mothers and children, little causal evidence exists on whether health policies during pregnancy based on pre-pregnancy BMI is beneficial or not. We study the effect of Danish obstetric guidelines recommending early routine induction for mothers with a pre-pregnancy BMI of at least 35 (7 days after their expected due date instead of 10-13 days after) using a regression discontinuity design. Our institutional setting implies that we only consider high-BMI women with no other major comorbidity. This is an interesting setting to study the impact of a policy based solely on pre-pregnancy BMI.

Our findings are threefold. First, early induction for high-BMI women improves immediate and first-year health outcomes. It reduces the share of newborns with birth weight of at least 4,5 kg. Consistently, the intervention reduces the occurrence of certain neonatal complications which can be triggered by a too-high birth weight: shoulder dystocia, hemorrhage, and vaginal lacerations. The main argument favoring early induction is a reduced mortality risk. While we find that early induction for high-BMI mothers reduces the joint risk of having an Apgar score$<$5 at 5 minutes and a measure of child mortality, we are cautious of the robustness of this result. First, the very low incidence of this outcome makes it harder to identify jumps exactly at the cutoff. Second, this result does not hold in all robustness checks, and we cannot know whether it is explained by residual selection or by mere noise around the cutoff.

Our second main result is evidence that the benefits of early induction for high-BMI mothers may persist beyond birth. Specifically, we find suggestive evidence that healthcare usage is lower during the first year of life for early induced mothers, and that the intervention reduces maternal postpartum depression risks. However, our third and final set of results shows no impact on mothers' labor supply and/or fertility responses the first two years after birth.

The overall beneficial health effects contrast with the more mixed evidence found for lower-risk pregnancies. Thus, our study provides causal evidence of beneficial effects of providing early induction for mothers with a BMI close to 35. This is an important policy result given the growing number of obese mothers-to-be and the broader debate surrounding the medicalization of childbirth.

References

Alkmark, M., J. K. Keulen, J. C. Kortekaas, C. Bergh, J. Van Dillen, R. G. Duijnhoven, H. Hagberg, B. W. Mol, M. Molin, J. A. van der Post, et al. (2020). Induction of labour at 41 weeks or expectant management until 42 weeks: A systematic review and an individual participant data meta-analysis of randomised trials. *PLoS medicine 17*(12), e1003436.

Almond, D., J. J. Doyle Jr, A. E. Kowalski, and H. Williams (2010). Estimating marginal returns to medical care: Evidence from at-risk newborns. *The quarterly journal of economics 125*(2), 591–634.

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

Bassler, D., M. Briel, V. M. Montori, M. Lane, P. Glasziou, Q. Zhou, D. Heels-Ansdell, S. D. Walter, G. H. Guyatt, S.-. S. Group, et al. (2010). Stopping randomized trials early for benefit and estimation of treatment effects: Systematic review and meta-regression analysis. *Jama 303*(12), 1180–1187.

Behrman, J. R. and M. R. Rosenzweig (2004). Returns to birthweight. *Review of Economics and statistics 86*(2), 586–601.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2007). From the cradle to the labor market?

the effect of birth weight on adult outcomes. *The Quarterly Journal of Economics 122*(1), 409–439.

Boulvain, M., O. Irion, and J. G. Thornton (2016). Induction of labour at or near term for suspected fetal macrosomia. *Cochrane Database of Systematic Reviews* (5).

Bryan, M., J. Roberts, C. Sechel, et al. (2019). The effect of mental health on employment: Accounting for selection bias. Technical report, HEDG, c/o Department of Economics, University of York.

Burgos, J., L. Rodríguez, B. Otero, P. Cobos, C. Osuna, M. d. M. Centeno, J. C. Melchor, L. Fernández-Llebrez, and T. Martínez-Astorquiza (2012). Induction at 41 weeks increases the risk of caesarean section in a hospital with a low rate of caesarean sections. *The Journal of Maternal-Fetal & Neonatal Medicine 25*(9), 1716–1718.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal 17*(2), 372–404.

Card, D., A. Fenizia, and D. Silver (2018). The health effects of cesarean delivery for low-risk first births. Technical report, National Bureau of Economic Research.

Card, D., A. Fenizia, and D. Silver (2023). The health impacts of hospital delivery practices. *American Economic Journal: Economic Policy 15*(2), 42–81.

Chen, C., X. Xu, and Y. Yan (2018). Estimated global overweight and obesity burden in pregnant women based on panel data model. *PloS one 13*(8), e0202183.

Clarke, D., J. P. Romano, and M. Wolf (2019). The Romano-Wolf multiple hypothesis correction in Stata.

Coates, D., A. Makris, C. Catling, A. Henry, V. Scarf, N. Watts, D. Fox, P. Thirukumar, V. Wong, H. Russell, et al. (2020). A systematic scoping review of clinical indications for induction of labour. *PLoS One 15*(1), e0228196.

Costa-Ramón, A., M. Kortelainen, A. Rodríguez-González, and L. Sääksvuori (2022). The long-run effects of cesarean sections. *Journal of Human Resources 57*(6), 2048–2085.

Cox, J. and J. Holden (2003). *Perinatal mental health: A guide to the Edinburgh Postnatal Depression Scale (EPDS).* Royal College of Psychiatrists.

Dahlen, H. G., C. Thornton, S. Downe, A. De Jonge, A. Seijmonsbergen-Schermers, S. Tracy, M. Tracy, A. Bisits, and L. Peters (2021). Intrapartum interventions and outcomes for women and children following induction of labour at term in uncomplicated pregnancies: A 16-year population-based linked data study. *BMJ open 11*(6), e047040.

Daskalakis, G., D. Zacharakis, M. Simou, P. Pappa, S. Detorakis, S. Mesogitis, and A. Antsaklis (2014). Induction of labor versus expectant management for pregnancies beyond 41 weeks. *The Journal of Maternal-Fetal & Neonatal Medicine 27*(2), 173–176.

Daysal, N. M., M. Trandafir, and R. Van Ewijk (2015). Saving lives at birth: The impact of home births on infant outcomes. *American Economic Journal: Applied Economics 7*(3), 28–50.

DSOG (2003). Den overvægtige gravide . `https://web.archive.org/web/20050121020122fw_/http://www.dsog.dk/files/overvaegtige%20gravide.htm`.

DSOG (2007a). Behandling af gestationel diabetes mellitus. `https://www.dsog.dk/obstetrik`.

DSOG (2007b). FOETUS MAGNUS SUSPICIO (DO366). `https://www.dsog.dk/obstetrik`.

DSOG (2010). Kliniske retningslinier for diabetesbehandling ved graviditet hos kvinder med kendt diabetes diabetes (Type 1 og Type 2) før graviditeten. `https://www.dsog.dk/obstetrik`.

DSOG (2011a). *Graviditet efter termin.* DSOG.

DSOG (2011b). Graviditet efter termin. https://www.dsog.dk/obstetrik.

DSOG (2014). Igansættelse: af fødsel. https://web.archive.org/web/20220120225251/http://static.squarespace.com/static/5467abcce4b056d72594db79/546e7748e4b0d969a4f6cf10/546e7746e4b0d969a4f6cc5a/1395263134000/PP-med-2014.pdf?format=original.

DSOG (2015). Monitorering af gravide efter gestationsalder 41+0. https://www.dsog.dk/obstetrik.

DSOG (2017). Den overvægtige gravide . https://www.dsog.dk/obstetrik.

DSOG (2018). Foetus magnus suspicio (do366). https://www.dsog.dk/obstetrik.

Figlio, D., J. Guryan, K. Karbownik, and J. Roth (2014). The effects of poor neonatal health on children's cognitive development. *American Economic Review 104*(12), 3921–3955.

Frank, R. G. and E. Meara (2009). The effect of maternal depression and substance abuse on child human capital development. Technical report, National Bureau of Economic Research.

Gelisen, O., E. Caliskan, S. Dilbaz, E. Ozdas, B. Dilbaz, and A. Haberal (2005). Induction of labor with three different techniques at 41 weeks of gestation or spontaneous follow-up until 42 weeks in women with definitely unfavorable cervical scores. *European Journal of Obstetrics & Gynecology and Reproductive Biology 120*(2), 164–169.

Gelman, A. and G. Imbens (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics 37*(3), 447–456.

Goldberger, A. S. (2008). Selection bias in evaluating treatment effects: Some formal illustrations. In *Modelling and Evaluating Treatment Effects in Econometrics*. Emerald Group Publishing Limited.

Gregersen, M. K. (2024). Earlier routine induction of labor - consequences on mother and child morbidity. *Health Economics*.

Halla, M., H. Mayr, G. J. Pruckner, and P. Garcia-Gomez (2020). Cutting fertility? Effects of cesarean deliveries on subsequent fertility and maternal labor supply. *Journal of Health Economics*, 102325.

Hedegaard, M., Ø. Lidegaard, C. Skovlund, L. Mørch, and M. Hedegaard (2015). Perinatal outcomes following an earlier post-term labour induction policy: A historical cohort study. *BJOG: An International Journal of Obstetrics & Gynaecology 122*(10), 1377–1385.

Hedegaard, M., Ø. Lidegaard, C. W. Skovlund, L. S. Mørch, and M. Hedegaard (2014). Reduction in stillbirths at term after new birth induction paradigm: Results of a national intervention. *BMJ open 4*(8), e005785.

Heimstad, R., P. R. Romundstad, J. Hyett, L.-Å. Mattsson, and K. Å. Salvesen (2007). Women's experiences and attitudes towards expectant management and induction of labor for post-term pregnancy. *Acta obstetricia ET gynecologica scandinavica 86*(8), 950–956.

Heimstad, R., E. Skogvoll, L.-Å. Mattsson, O. J. Johansen, S. H. Eik-Nes, and K. Å. Salvesen (2007). Induction of labor or serial antenatal fetal monitoring in postterm pregnancy: a randomized controlled trial. *Obstetrics & Gynecology 109*(3), 609–617.

Hildingsson, I., A. Karlström, and A. Nystedt (2011). Women's experiences of induction of labour–findings from a swedish regional study. *Australian and New Zealand journal of obstetrics and gynaecology 51*(2), 151–157.

Hirani, J. L.-J., H. H. Sievertsen, and M. Wüst (2020). Missing a nurse visit. Technical report, IZA Discussion Papers.

Hoe, T. P. (2022). Does hospital crowding matter? evidence from trauma and orthopedics in england. *American Economic Journal: Economic Policy 14*(2), 231–262.

Jungensen, D. R. (2011). Visitation ved højt bmi. *Jordmoderforeningen* (05). `https://jordemoderforeningen.dk/tidsskriftsartikel/visitation-ved-hojt-bmi/`.

Kenyon, S., L. Middleton, M. Skrybant, and T. Johnston (2019). When to induce late term pregnancies. *BMJ 367*.

Keulen, J. K., A. Bruinsma, J. C. Kortekaas, J. Van Dillen, P. M. Bossuyt, M. A. Oudijk, R. G. Duijnhoven, A. H. Van Kaam, F. P. Vandenbussche, J. A. Van Der Post, et al. (2019). Induction of labour at 41 weeks versus expectant management until 42 weeks (index): multicentre, randomised non-inferiority trial. *bmj 364*.

Kreiner, C. T. and H. H. Sievertsen (2020). Neonatal health of parents and cognitive development of children. *Journal of Health Economics 69*, 102247.

Lou, S., L. Hvidman, N. Uldbjerg, L. Neumann, T. Jensen, J. Haben, and K. Carstensen (2019, 09). Women‘s experiences of postterm induction of labor: A systematic review of qualitative studies. *Birth 46*.

Maibom, J., H. H. Sievertsen, M. Simonsen, and M. Wüst (2021). Maternity ward crowding, procedure use, and child health. *Journal of Health Economics 75*, 102399.

Maimburg, R. D. (2020). Induction of labour in earlier gestational weeks-reflections on evidence and clinical practice. *Sexual & Reproductive Healthcare: Official Journal of the Swedish Association of Midwives 23*, 100487–100487.

Muglu, J., H. Rather, D. Arroyo-Manzano, S. Bhattacharya, I. Balchin, A. Khalil, B. Thilaganathan, K. S. Khan, J. Zamora, and S. Thangaratinam (2019). Risks of stillbirth and neonatal death with advancing gestation at term: A systematic review and meta-analysis of cohort studies of 15 million pregnancies. *PLoS medicine 16*(7), e1002838.

Peters, L. L., C. Thornton, A. De Jonge, A. Khashan, M. Tracy, S. Downe, E. I. Feijen-de Jong, and H. G. Dahlen (2018). The effect of medical and operative birth interventions on child health outcomes in the first 28 days and up to 5 years of age: A linked data population-based cohort study. *Birth 45*(4), 347–357.

Place, K., L. Rahkonen, A. Tekay, K. Väyrynen, M.-R. Orden, M. Vääräsmäki, J. Uotila, K. Tihtonen, K. Rinne, K. Mäkikallio, et al. (2024). Labor induction at 41+ 0 gestational weeks or expectant management for the nulliparous woman: The finnish randomized controlled multicenter trial. *Acta Obstetricia et Gynecologica Scandinavica 103*(3), 505–511.

Rydahl, E., E. Declercq, M. Juhl, and R. D. Maimburg (2019). Routine induction in late-term pregnancies: follow-up of a danish induction of labour paradigm. *BMJ open 9*(12), e032815.

Rydahl, E., L. Eriksen, and M. Juhl (2019). Effects of induction of labor prior to post-term in low-risk pregnancies: A systematic review. *JBI Database of Systematic Reviews and Implementation Reports 17*(2), 170.

Schwarz, C., M. M. Gross, P. Heusser, and B. Berger (2016). Women s perceptions of induction of labour outcomes: Results of an online-survey in germany. *Midwifery 35*, 3–10.

Shetty, A., R. Burt, P. Rice, and A. Templeton (2005). Women's perceptions, expectations and satisfaction with induced labour—a questionnaire-based study. *European Journal of Obstetrics & Gynecology and Reproductive Biology 123*(1), 56–61.

Sievertsen, H. H. and M. Wüst (2017). Discharge on the day of birth, parental response and health and schooling outcomes. *Journal of health economics 55*, 121–138.

Smith-Nielsen, J., S. Matthey, T. Lange, and M. S. Væver (2018). Validation of the edinburgh postnatal depression scale against both dsm-5 and icd-10 diagnostic criteria for depression. *BMC psychiatry 18*(1), 1–12.

Spillane, E. (2020). Induction of labour the debate. *British Journal of Midwifery 28*(3), 142–143.

SSI (2009). ANBEFALINGER FOR SVANGREOMSORGEN. Technical report, The Danish Health Authorities (Sundhedsstyrelsen).

SSI (2013). ANBEFALINGER FOR SVANGREOMSORGEN. Technical report, The Danish Health Authorities (Sundhedsstyrelsen).

Stommes, D., P. Aronow, and F. Sävje (2023). On the reliability of published findings using the regression discontinuity design in political science. *Research & Politics 10*(2), 20531680231166457.

Stotland, N., A. Caughey, E. Breed, and G. Escobar (2004). Risk factors and obstetric complications associated with macrosomia. *International journal of gynecology & obstetrics 87*(3), 220–226.

Tonei, V. (2019). Mother's mental health after childbirth: Does the delivery method matter? *Journal of Health Economics 63*, 182–196.

Vinter, C. A., M. H. Tanvig, P. Damm, K. V. Naver, K. R. Andreasen, L. L. T. Andersen, S. Liest, S. Lunde, D. M. Jensen, and K. M. Renault (2012). Overvægtige gravide og komplikationer i relation til graviditet og fødsel. *Ugeskrift for Læger 174*(16), 1079–82.

Von Hinke Kessler Scholder, S., N. Rice, and E. Tominey (2019). Mental health around pregnancy and child development from early childhood to adolescence.

Waldenström, U., I. Hildingsson, C. Rubertsson, and I. Rådestad (2004). A negative birth experience: prevalence and risk factors in a national sample. *Birth 31*(1), 17–27.

Wennerholm, U.-B., S. Saltvedt, A. Wessberg, M. Alkmark, C. Bergh, S. B. Wendel, H. Fadl, M. Jonsson, L. Ladfors, V. Sengpiel, et al. (2019). Induction of labour at 41 weeks versus expectant management and induction of labour at 42 weeks (swedish post-term induction study, swepis): multicentre, open label, randomised, superiority trial. *bmj 367*.

WHO (2011). *WHO recommendations for induction of labour.* World Health Organization.

Wolff, S. L., I. Lorentzen, A. P. Kaltoft, H. Schmidt, M. M. Jeppesen, and R. D. Maimburg (2016). Has perinatal outcome improved after introduction of a guideline in favour of routine induction and increased surveillance prior to 42 weeks of gestation?: A cross-sectional population-based registry study. *Sexual & Reproductive Healthcare 10*, 19–24.

Zizzo, A. R., I. Kirkegaard, A. Pinborg, and N. Ulbjerg (2017). Decline in stillbirths and perinatal mortality after implementation of a more aggressive induction policy in post-date pregnancies: A nationwide register study. *Acta Obstetricia et Gynecologica Scandinavica 96*(7), 862–867.

Online Appendix

A Background

A.1 Existing literature on early routine induction

When possible, we compare our results to studies which are more plausible causal - especially studies reporting ITT estimates which are more directly comparable to the parameter we can causally estimate. While there is a lack of causal studies on whether having a high pre-pregnancy BMI justifies even earlier routine induction (Coates et al., 2020), we can compare our results to studies evaluating early induction for a general population of low-risk pregnancies. It is well-known that the risk of stillbirth increases with gestational length, especially after the EDD[23]. Generally, the discussion of early induction revolves around whether the benefits of induction earlier than 14 days after the EDD outweighs the costs (see Gregersen (2024) for details).

For the general population of low-risk pregnancies, there exists four RCTs evaluating routine induction at 7-9 days after the EDD instead of at 14 days after (Wennerholm et al., 2019; Keulen et al., 2019; Gelisen et al., 2005), and one comparing routine induction at 7 days after the EDD instead of at 12-15 days after (Place et al., 2024). Except for Place et al. (2024), all report ITT estimates.

Notably, a large randomized control trial in Sweden (SWEPSIS) was shut down early in 2020 because of increased mortality rates in the control group (no early induction) leading to much fewer participants than planned (Wennerholm et al., 2019). Yet, their contribution to the body of evidence makes it unlikely that setting up new RCTs would be ethically justifiable (Kenyon et al., 2019). With six deaths in the control group and no deaths in the early induction group, Wennerholm et al. (2019) cautiously report that early induction may significantly reduce the risk of stillbirth. With the exception of SWEPSIS, none of the other

[23] See e.g. Muglu et al. (2019) for a recent review.

RCTs find any statistically significant reduction of mortality risks, but the numbers always point in the direction of a reduction.

Analyzing combined individual data from Wennerholm et al. (2019) and Keulen et al. (2019), Alkmark et al. (2020) report a total reduction in mortality risks, an overall decreased risk of severe adverse perinatal outcomes, admission to the NICU, meconium-stained amniotic fluid and hypertension, but no impact on the rate of C-section. Birth weight and the risk of macrosomia were reduced too, and the use of pain relief during labor increased. The beneficial effects are driven by first-time mothers. Doing a standard meta-analysis adding the study by Gelisen et al. (2005), Alkmark et al. (2020) further confirm a reduction in their measure of severe adverse perinatal outcomes and no impact on the risk of C-sections, but do not report results for remaining outcomes. Because of early termination of the SWEPSIS study, Alkmark et al. (2020) may also suffer from the risk of overestimating treatment effects (Bassler et al., 2010).

The RCT by Gelisen et al. (2005) includes only 600 individuals, but find a reduced risk of macrosomia, shoulder dystocia, meconium-stained amniotic fluid and meconium aspiration syndrom in the early induction group. Remaining outcomes were similar across treatment and control group. The more recently published RCT by Place et al. (2024) was already running at the time SWEPSIS was terminated. They analyze results for a little less than 400 participants, and find a lower risk of instrumental delivery, having a large bleeding, and a birth weight$\geq$ 4,000 grams. They report no deaths and thus do not analyze mortality.

The slightly older review by Rydahl et al. (2019), includes one of the above RCTs (Gelisen et al., 2005) together with a RCT from Norway comparing induction at 9 days after the EDD with induction at 20 days after the EDD (Heimstad et al., 2007). Additionally, they consider two empirical evaluations of different procedures (either inductions at 8 days after the EDD or alternatively 14 to 15 days after the EDD) (Burgos et al., 2012; Daskalakis et al., 2014) as well as three pure correlational studies. The incidence of mortality is very rare, so even in their meta-study they lack statistically power to analyze it but the numbers points towards a reduction. They overall find a bit more mixed evidence on morbidity.[24] Yet, the empirical evaluations face different problems regarding causal interpretation. Burgos et al. (2012) use a before and after design and are thus not able to separate cohort effects from the procedure of early induction. Daskalakis et al. (2014) evaluate two different procedures applied at the same hospital depending on physicians' preference. Thus, causal interpretation requires random allocation between patients and physicians which is likely not met in an empirical setting. The RCT by Heimstad et al. (2007) provides ITT estimates, and find no severe differences between control and treatment group, but the early induction group is more likely to experience precipitate labor and less likely to have a baby with birth weight<4,000 grams or oligohydramnios.

Several studies evaluate the introduction of new obstetric guidelines in Denmark in 2011. Specifically, they evaluate introducing surveillance at 7 days after the EDD together with routine induction at 10-13 days after the EDD instead of 14 days after for women age<40 and BMI<35 (Hedegaard et al., 2014, 2015; Wolff et al., 2016; Zizzo et al., 2017; Rydahl et al., 2019; Gregersen, 2024). Here we limit attention to the studies which address time trends in some way. When accounting for time trends, none of these studies find any evidence of a statistical significant reduction in mortality due to earlier routine induction.[25] Notice, that these studies evaluate a different treatment and population than what we do.

Wolff et al. (2016) find an increased risk of being admitted to the NICU (borderline significant) and a decreased risk of having a large baby, but no impact on remaining outcomes. Using an interrupted time-series analysis, Rydahl et al. (2019) find a large increased risks of uterus rupture. At the time of implementation, they also find a smaller increase in epidural use as well as a smaller decrease in the risk of augmentation of labor, but both levels off over

[24]Focusing on the two RCTs and two empirical evaluations, they find that early induction affected morbidity in the following ways: increased risk of the child's PH value<7, C-section because of a failure to progress (but not overall C-Section risk), and precipitate labor, but reduced risk of meconium-stained amniotic fluid and oligohydramnious). (Some of) the included three cohort studies additionally showed increased risk of uterus rupture, overall risk of C-section + C-section after previous vaginal birth, instrumental delivery, labor dystocia, and chorioamnionitis.

[25]We refer to the Appendix in Gregersen (2024) for a detailed review.

time.

Using a flexible difference-in-difference strategy, Gregersen (2024) find no evidence on any (severe) morbidity effects despite a significant increase in the induction rate of more than 32% for first-time mothers with BMI<35 and age<40. Effectively, she is estimating ITT effects. She is unable to analyze the risk of uterus ruptures because of too few incidences in her sample, suggesting that Rydahl et al. (2019)'s result on uterus rupture may be driven by mothers with a previous C-section. While the main results show a reduced risk of having a large baby and increased use of epidural anesthesia, neither result is robust.

In contrast to most medical papers, Gregersen (2024) additional follow mothers and children up to three year after birth and find no effect on the child's health care usage, mother's fertility or a proxy for mother's mental health during the first year after birth. A few medical papers show that either induced labor or induced birth are associated with adverse child health in the medium to longer-term (Dahlen et al., 2021; Peters et al., 2018) or find that the risk of being diagnosed with cerebral palsy within 3 years after birth is reduced in years with more liberal labor induction policies (Hedegaard et al., 2015), but they do no use a causal identification strategy.

A part of the discussion regarding earlier routine induction relates to mothers' birth experience which is harder to evaluate. In a RCT, Heimstad et al. (2007) provide ITT estimates and report evidence favoring routine induction at 7 days after the EDD instead of continued monititoring and awaiting spontaneous birth. Not only did the majority of mothers prefer to be induced at 7 days after the EDD, but the majority of mothers who were induced early had a positive experience and would even prefer the same management in a future pregnancy. Mothers exposed to early induction had shorter duration of labor but experienced more intense and frequent contractions too. However, in other studies, induction of labor was associated with less satisfaction, a less positive birthing experience, and a higher concern for the newborn's safety (Shetty et al., 2005; Waldenström et al., 2004; Hildingsson et al., 2011). In an online survey, the majority of women with induced labor would prefer not to be induced in the future (Schwarz et al., 2016). Moreover, they felt they lacked both information and support in their decision-making. These points are supported by a review of qualitative studies which points to a need to provide better information on the induction process as well as to help women adjust their birth expectations when inducing labor (Lou et al., 2019).

A.2 High-BMI pregnancies in Denmark

The aim of this section is to describe general guidelines and background information relevant for high-BMI pregnancies in Denmark. Overweight and obese pregnancies are associated with increased maternal and child health risks. While the cause is unknown, metabolic issues and insulin resistance are possible explanations (SSI, 2013). Specifically, DSOG (2017) lists increased risks of pregnancy complications such as hypertension, preeclampsia and gestational diabetes (GDM) along with unfavorable outcomes such as miscarriage, very preterm birth ($\leq$28 weeks), and even perinatal death. Moreover, children born from overweight mothers face increased risks of congenital malformations and a higher likelihood of ending up being overweight in the long term. Not only do birth interventions such as labor induction, instrumental delivery, and C-section happen more frequently, but the risks associated with C-section such as infections, a large bleeding and anesthesia complications are higher too. The risk of having a large baby (macrosomia) increases with maternal BMI, which carries both overlapping and separate risks[26]. Detecting large babies is linked with uncertainty - especially for overweight women (DSOG, 2007b, 2018). Guidelines effective during our sample periods favors a planned C-section when the estimated birth weight$\geq$4,8 kg (4,5 kg in case of any type of diabetes (DM)).

Throughout the study period, the risks associated with being overweight and obese are recognized and addressed in general guidelines (DSOG, 2003; SSI, 2009, 2013). The poten-

[26] For the child, DSOG (2017) list risks such as shoulder dystocia, plexus brachialis lesions, asphyxia, and stillbirth. Maternal risk includes a large bleeding (>500 ml), a severe laceration (degree 3 and 4) and the risk of an emergency C-section. Average hospitalization (for both mother and child) are longer too.

tial benefits of giving birth at a larger hospital with access to specialized care are already highlighted by the first DSOG task force focusing on overweight in 2003 (DSOG, 2003) and advised by some gynecologists in 2012 (Vinter et al., 2012) and certain hospitals in 2011 (Jungensen, 2011). However, this did not become a part of the national guidelines until 2017 (DSOG, 2017).

During our sample period, overweight pregnancies are offered counseling regarding reducing weight gain during pregnancy as well as encouraged to engage in at least 30 minutes of daily physical activity (SSI, 2009). While weight loss during pregnancy is not recommended, ensuring a healthy lifestyle during pregnancy and losing weight prior to becoming pregnant were advised. Pregnancies with a BMI$\geq$35 might receive additional checks and surveillance including additional ultrasound scans to help estimate the size of the fetus as well as the fetus's position in the womb[27] (SSI, 2009). Yet, the determination of these factors becomes more challenging and less precise for overweight and obese women (SSI, 2013). Moreover, pregnancies with a BMI$\geq$35 may be referred to a consultation with an obstetrician during pregnancy. This consultation may involve discussing potential birth complications and the possibility of using epidural anesthesia prophylactically (SSI, 2009). Several guidelines related to the risks and consequences of being overweight or dealing with suspected macrosomia emphasize that an overall evaluation of individual risks and preferences should be considered when determining the preferred birth mode and best plan of action for each pregnancy (DSOG, 2018, 2017). In particular, prior pregnancy and birth history should be taking into account when deciding the best course of action.

Throughout the study period, a BMI $\geq$ 27 meant that the woman would undergo at least one glucose test to detect possible cases of GDM. If diagnosed with GDM, special guidelines for surveillance as well as recommended time and mode of birth will be implemented (DSOG, 2007a). Similarly, special guidelines are effective for pregnancies with pre-existing

[27] Procedures may vary across hospitals. In 2012, using an ultrasound scan at weeks 35-38 to obtain a weight estimate of the child for overweight pregnancies was standard procedure at some hospitals (Vinter et al., 2012).

DM (DSOG, 2010). In general, guidelines recommend induction no later than at the EDD/7 days after the EDD when insulin treatment is required/not required. Yet, if the weight estimate is above 4 kg/4,5 kg even early induction/a planned C-section is recommended.

A final health policy relevant to our study is restricted access to fertility treatment based on BMI. Most fertility clinics restrict treatment to women with BMI below a certain threshold (sometimes at 35, but it varies between 29 and 40 across clinics and age-groups).

B Additional Data Information

B.1 Controls: pre-birth parental background variables and other indications for induction before the EDD

We use the LPR to recover measures of certain pregnancy complications including prior C-section (ICD-10:O82), insufficient fetal growth (0365), prior large baby (birth weight≥4,5 kg), having a multiple birth, first time-mother, or preeclampsia (O11.9). To avoid bad control issues, we only consider conditions registered prior to the EDD.

Asides from the LPR, the set of background variables we use to lead balances of observable tests and adjust the RD estimates are also extracted from administrative register data from Statistics Denmark. This data includes parental background characteristics such as a dummy if either parent graduated from college, monthly joint parental income in 2024 US dollars, maternal age, a dummy which equals one if either parent was born in Denmark, a dummy for whether parents cohabit, as well dummies for missing information on any of the control variables.

From the MFR (birth register), we also extract measures of the hospital, of the day of the week of birth, and of the month of birth to build fixed-effects. Hospital fixed-effects are especially important to account for differences in practices across hospitals.

B.2 Descriptives in our main sample

To understand how risk profiles vary across the BMI cutoff, Table 5 shows descriptive statistics for our main sample of pregnancies going at least 7 days past the EDD, both in total (column 1), and split by mothers with a BMI < 35 (column 2), and those with BMI ≥ 35 (column 3).

Mothers with a BMI ≥ 35 exhibit an increased likelihood of experiencing certain complications and adverse birth outcomes such as macrosomia (birth weight≥4,5 kg), admission to the NICU, and longer hospital stays. They also more likely to have a high risk of postpartum depression (EPDS score > 11) and to belong to a lower socioeconomic status, as evidenced by lower parental income, education level, and employment status. However, certain birth outcomes are also better for this group of mothers as the risk of hemorrhage, lacerations, and shoulder dystocia are all lower.

These descriptive differences are hard to interpret. At least three different forces are at play. First, ceteris paribus, higher maternal BMI is widely acknowledged as a risk factor associated with worse outcomes (Coates et al., 2020) (see Appendix 3.1 for details). Secondly, since we focus on births taking place at least 7 days after the EDD, our population is positively selected. Remember that severe complications or risk-factors indicate induction or a planned C-section prior to the EDD. Thirdly, mothers above the BMI cutoff are almost 20 p.p. more likely to be induced, than those below. This intervention can improve or deteriorate neonatal and maternal health or may be a result of unobserved underlying health differences. Hence, absent of an identification strategy, it is impossible to explain the observable differences in neonatal health.

Tab. 5 Variable means, births taking place at least 7 days week after the EDD (2011-2017)

	All births	BMI<35	BMI>=35	t-test
Any parent college	0.558	0.567	0.403	0.000
Parental income, monthly (dkk)	6448.586	6503.880	5515.404	0.000
Any parent Danish	0.867	0.866	0.881	0.001
Parents are cohabiting	0.828	0.829	0.806	0.000
First-time mother	0.513	0.514	0.484	0.000
Maternal age	30.533	30.570	29.899	0.000
Preeclampsia	0.007	0.007	0.011	0.010
Insufficient growth	0.017	0.017	0.014	0.039
Prior C-section	0.067	0.066	0.085	0.000
Prior large fetus	0.013	0.012	0.021	0.000
Induced labor at any gestational age	0.397	0.384	0.618	0.000
Induced 7 and 9 days after EDD	0.079	0.063	0.346	0.000
Epidural	0.431	0.425	0.519	0.000
Elective C-Section	0.006	0.006	0.010	0.012
Emergency C-section	0.048	0.048	0.056	0.012
Vaccuum extraction	0.098	0.099	0.070	0.000
Gestational age (days)	290.225	290.269	289.480	0.000
Birth weight	3777.937	3774.956	3828.238	0.000
Apgar <7	0.005	0.005	<0.01	0.698
Birthweight $\geq$ 4.5 kg	0.056	0.056	0.072	0.000
Share sent in NICU	0.049	0.048	0.058	0.002
Hospital Nights	2.125	2.115	2.298	0.000
Shoulder dystocia	0.007	0.007	<0.01	0.618
Lacerations	0.329	0.333	0.254	0.000
Lacerations 3rd degree higher	0.034	0.034	0.026	0.001
Hemorrhage	0.511	0.517	0.414	0.000
Complication index	0.002	0.002	<0.01	0.893
EPDS>11	0.056	0.056	0.060	0.529
EPDS Score	4.844	4.829	5.090	0.008
Nurses visits	4.596	4.600	4.531	0.029
GP Visit (1st year)	17.005	16.934	18.200	0.000
Mother employed 2 years post birth	0.870	0.874	0.803	0.000
Prob.another child two years post birth	0.023	0.023	0.022	0.677
Observations	91743	86611	5132	

Notes: The following table displays the mean values of various variables for all births in Denmark that reached at least 7 days after the EDD. It also includes the mean values for mothers with a BMI $<$ 35, those with a BMI $\geq$ 35, and the p-value from a two-tailed t-test comparing the two subgroups.

C Additional results

C.1 Treatment & First Stage (robustness)

Tab. B1 First stage: ITT effect on early labor induction (7-9 days after the EDD) across samples

	(1)	(2)	(3)
Sharp	0.157***	0.151***	0.193***
	(0.025)	(0.031)	(0.046)
Covariates	Yes	Yes	Yes
(Effective) Observations	16342	10457	4999
Mean Control Group	0.096	0.099	0.102
Sample	All births	Nurses visits	EPDS Score

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Standard errors are robust and clustered at the hospital-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.This table shows RD estimates from a sharp RDD with being induced 7-9 days after the EDD as the outcome, the mother's BMI as the running variable, and 35 as the cutoff. The bandwidth is manually set to 7 for reasons explained in subsection C.3. "Mean in the control group" refers to the average outcome for births on the left-side of the bandwidth (Maternal BMI 28-35). The set of controls is described in subsection B.1. In the sample restriction row, "All births" refer to all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later, whereas "Nurse visits" and "EPDS score" refer to the respective subsets of this sample for which nurse visits from the NHV and mental health data are available.

Tab. B2 gestational length (Days past the EDD)

	Sharp	Sharp	Sharp	Fuzzy	Fuzzy	Fuzzy
Gest. length	-0.675***	-0.640***	-0.987***	-4.050***	-4.137***	-4.924***
	(0.145)	(0.174)	(0.251)	(0.641)	(0.882)	(1.047)
(Effective) Observations	16342	10457	4999	16342	10457	4999
Mean Control group	10.372	10.367	10.335	10.372	10.367	10.335
Sample	All	Nurse visits	EPDS Score	All	Nurse visits	EPDS Score

Notes: Standard errors are robust and clustered at the hospital-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table shows RD estimates from a sharp RDD with gestational length in days (past the EDD) as the outcome, the mother's BMI as the running variable, and 35 as the cutoff. The bandwidth is manually set to 7 for reasons explained in subsection C.3. "Mean in the control group" refers to the average outcome for births on the left-side of the bandwidth (Maternal BMI 28-35). The set of controls is described in subsection B.1. In the sample restriction row, "All births" refer to all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later, whereas "Nurse visits" and "EPDS score" refer to the respective subsets of this sample for which nurse visits from the NHV and mental health data are available.

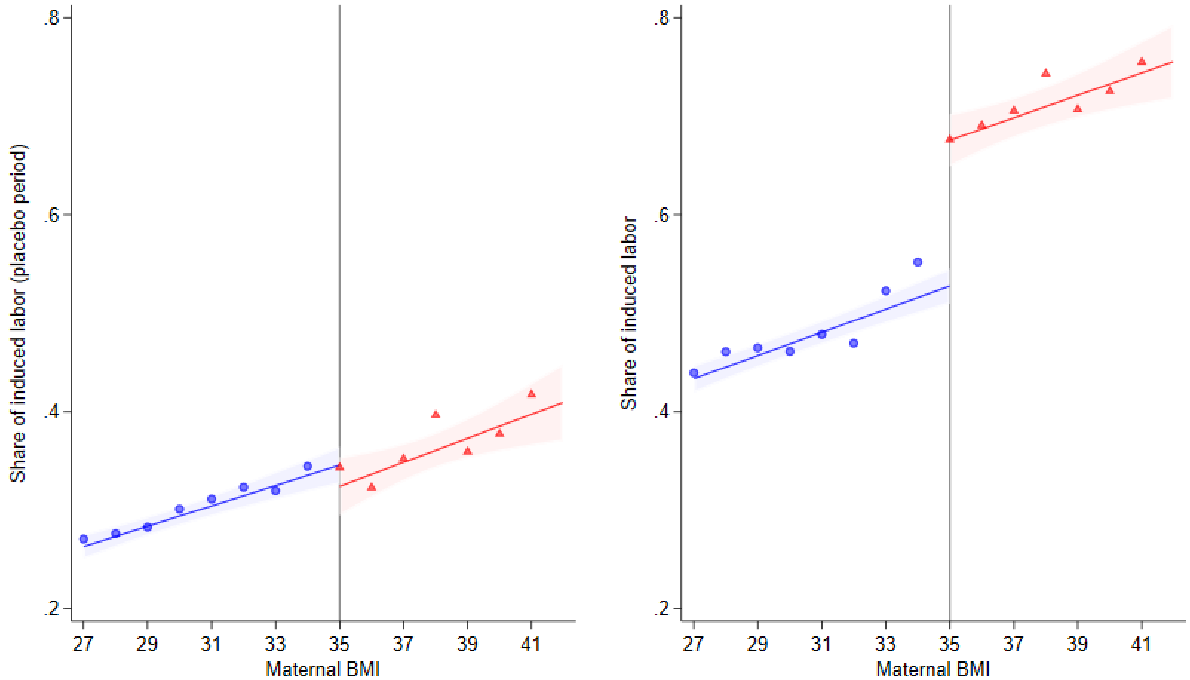


Fig. B1 ITT effect on being ever induced in placebo period (2004-2009) and when the guidelines are binding (2011-2017)

Notes: The bins are the average outcomes for each BMI unit. The lines are lines of fit from an OLS regression on all observations. Confidence intervals are at the 90 percent level and built from heteroskedastic robust SEs clustered at the hospital-year level. The right panel sample is from the period after the induction guidelines were issued (2011-2017), while the left panel sample is from the placebo period before the guidelines were issued (2004-2009).

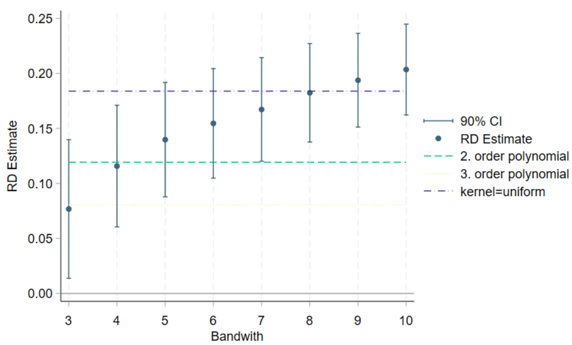
Tab. B3 Ever induced

	Sharp	Sharp	Sharp	Fuzzy	Fuzzy	Fuzzy
Ever induced	0.078***	0.066**	0.058	0.508***	0.455***	0.308*
	(0.026)	(0.032)	(0.043)	(0.116)	(0.154)	(0.168)
(Effective) Observations	16342	10457	4999	16342	10457	4999
Mean Control Group	0.462	0.463	0.473	0.462	0.463	0.473
Sample	All	Nurse visits	EPDS Score	All	Nurse visits	EPDS Score

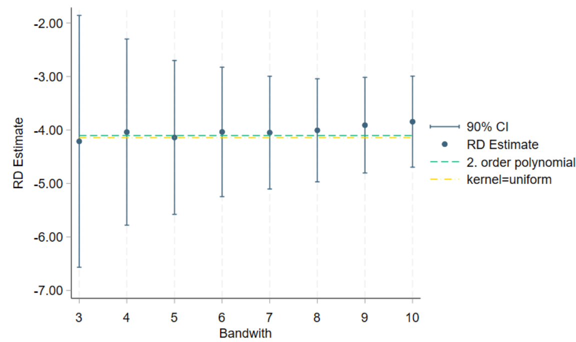
Notes: Standard errors are robust and clustered at the hospital-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.This table shows RD estimates from a sharp RDD with labor induction induced at any time as the outcome, the mother's BMI as the running variable, and 35 as the cutoff. The bandwidth is manually set to 7 for reasons explained in subsection C.3. "Mean in the control group" refers to the average outcome for births on the left-side of the bandwidth (Maternal BMI 28-35). The set of controls is described in subsection B.1. In the sample restriction row, "All births" refer to all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later, whereas "Nurse visits" and "EPDS score" refer to the respective subsets of this sample for which nurse visits from the NHV and mental health data are available.

Fig. B2 Effects on early induction and gestational length across different bandwidth, polynomial order, and kernel choices, and samples

(a) Induced at 7-9 after the EDD



(b) Gestational length (days after the EDD)



Notes: The two figures are coefficient plots showing RD estimates from a RDD with labor induction (left panel) and gestational length measured in days after the EDD (right panel) as outcomes and BMI as the running variables. All estimates are from a sharp RDD on the right panel (as it is the first stage) and a fuzzy RDD on the right panel. Point estimates with higher polynomial order and different kernel are computed with a bandwidth of seven, as in our main specification. Standard errors are robust and clustered at the hospital-year level. Sample is the main one (all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later).

C.2 Effect on Healthcare usage across samples

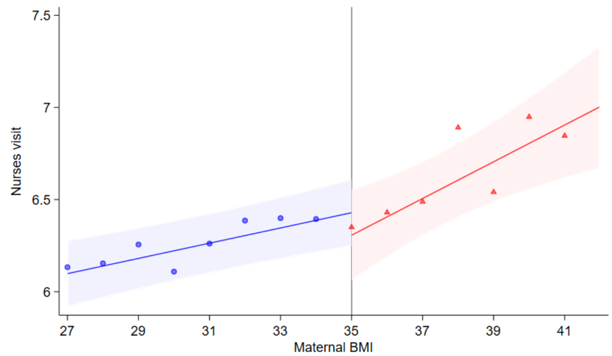


Fig. B3 Impact of the intervention on the number of total nurse visits (first year after birth)

Notes: The bins are the average outcomes for each BMI unit. The lines are lines of fit from an OLS regression on all observations. Confidence intervals are at the 90 percent level and built from heteroskedastic robust SEs clustered at the hospital-year level. The sample consists of all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later, with a total of n=11,225. No placebo period is shown because the data from the NHV is not available before 2011. An alternative placebo test is however provided in Section 6.

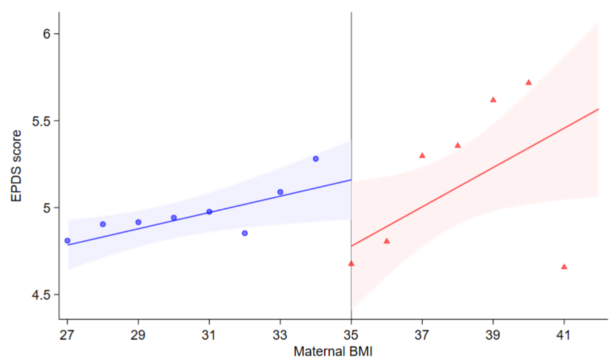


Fig. B4 Impact of the intervention on maternal mental health (EPDS score)

Notes: The bins are the average outcomes for each BMI unit. The lines are lines of fit from an OLS regression on all observations. Confidence intervals are at the 90 percent level and built from heteroskedastic robust SEs clustered at the hospital-year level. The sample consists of all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later, with a total of n=4999. No placebo period is shown because the data from the NHV is not available before 2011. An alternative placebo test is however provided in section 6.

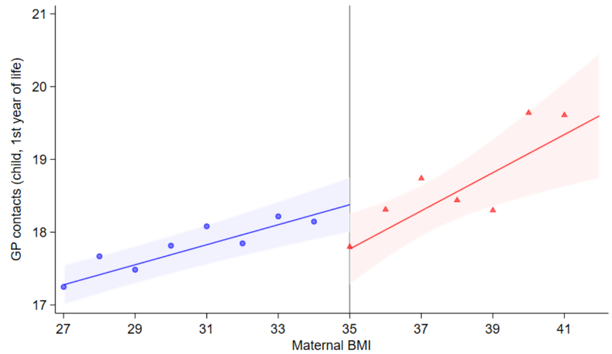


Fig. B5 Impact of the intervention on the number of GP Visits (1st year of life)
Notes: The bins are the average outcomes for each BMI unit. The lines are lines of fit from an OLS regression on all observations. Confidence intervals are at the 90 percent level and built from heteroskedastic robust SEs clustered at the hospital-year level. The sample consists of all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later, with a total of n=4999. No placebo period is shown because the data from the NHV is not available before 2011. An alternative placebo test is however provided in section 6.

Tab. B4 Impact on healthcare usage (Nurse visits and EPDS samples)

	Fuzzy	Obs	Fuzzy	Obs
Nurses visits (total)	-2.198*	10427	-3.589***	4994
	(1.132)		(1.309)	
Universal nurses visit	-2.592***	10427	-2.240***	4994
	(0.743)		(0.635)	
Additional nurses visits	0.129	10427	-1.593	4994
	(0.991)		(1.311)	
GP Visits (child)	-2.554	10427	-5.064	4999
	(2.278)		(3.791)	
Covariates	Yes		Yes	
Sample	Nurse visits		EPDS Score	

Notes: Standard errors are robust and clustered at the hospital-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.This table shows RD estimates from a sharp RDD with being induced 7-9 days after the EDD as the outcome, the mother's BMI as the running variable, and 35 as the cutoff. The bandwidth is manually set to 7 for reasons explained in subsection C.3. "Mean in the control group" refers to the average outcome for births on the left-side of the bandwidth (Maternal BMI 28-35). The set of controls is described in subsection B.1. In the sample restriction row, "All births" refer to all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later, whereas "Nurse visits" and "EPDS score" refer to the respective subsets of this sample for which nurse visits from the NHV and mental health data are available.

C.3 Power Analysis and picking the appropriate bandwith

Since the RDD is a data-demanding design (Stommes et al., 2023; Goldberger, 2008) and the group of mothers with a pre-pregnancy BMI≥35 is rather small, we here address the power issue in detail. In Figure B7 below, we show the share of all ongoing pregnancies with a BMI above the 35 threshold to highlight our power issue. Specifically, out of all pregnancies passing the EDD, only a relatively low share has a BMI of at least 35. Once again, remember that mothers with many types of comorbidity are induced or receive a planned CS prior to the EDD.

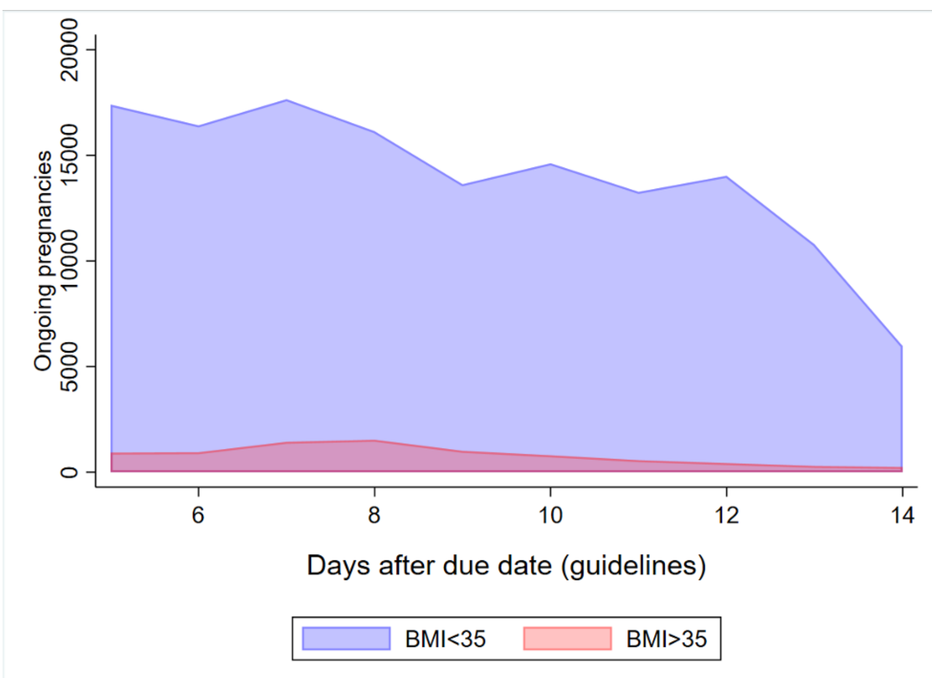


Fig. B6 Total ongoing births by days after the due date and BMI category

Notes: The figure shows the total ongoing births by days after the due date and BMI category (BMI<35 or BMI $\geq$ 35). The sample includes all births in Denmark between 2011 and 2017 with a gestational length of more than four days past the EDD.

In Table B5, we standardize our outcomes to compute the probability to detect a strong effect (SMD$\geq$ 0.8) when alpha equals 0.1 for all of the outcomes in our main analysis. The first column displays power results with Calonico et al. (2017)'s optimal data-driven bandwidth algorithm. It is between 3 or 6 depending on the outcomes. Strikingly, the probability to detect a strong effect is below the 80 percent cutoff for all of our outcomes but the severe complications index. This is despite two generous hypotheses regarding the alpha and the size of the effect. Opting for an optimal data-driven bandwidth can only lead to underpowered analyses in our setting. Hence, in the second column of the power table, we enlarge the bandwidth and set it to seven for every outcome. This allows analyses for almost all of our outcomes to be reasonably powered. However, increasing the polynomial order by one renders all tests underpowered again. These power results motivate two central choices in the rest of our analysis.

Tab. B5 Power analysis: probability to detect an effect of 0.8 SD

	Data-driven bandwith	Bandwith=7 (p=1)	Bandwith=7 (p=2)
Epidural	0.273	0.972	0.717
Apgar <7	0.544	0.982	0.699
Share sent in NICU	0.381	0.981	0.736
Birth weight≥4.5 kg	0.401	0.981	0.675
Hospital Nights	0.569	0.999	0.889
Uterus Rupture	0.496	0.996	0.780
Birth Weight (grams)	0.508	0.999	0.845
Lacerations	0.397	0.993	0.773
Hemorrhage	0.326	0.981	0.672
Shoulder dystocia	0.602	1.000	0.969
Emergency CS	0.780	0.996	0.765
Lacerations	0.397	0.993	0.773
Lacerations(3nd degree and higher)	0.508	0.999	0.879
Vaccuum extraction	0.449	1.000	0.877
Complications index	0.999	1.000	1.000

Notes: The table shows the probability to detect an effect of 0.8 SD for various outcomes and varying specification choices. Sample is the main one containing all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later.

First, in our central specifications, we set the bandwidth to seven rather than using a data-driven bandwidth. It is necessary for our analyses to be sufficiently powered. Naturally, since our manual bandwidth is superior to the optimal bandwidth, this choice increases the risk of bias. However, the various balancing tests (see Table 1) and placebo tests (see sections 5 and 6) we conducted speak against major bias and in favor of the local randomization hypothesis. Moreover, when the outcomes are available before the guidelines, we also check there is no discontinuity before the introduction to the guidelines. Together, these robustness checks suggest the probability our results are cofounded is minimal, despite us picking a bandwidth slightly higher than the optimal one. Moreover, using coefficient plots (Appendices C.5 and C.6), we make sure that the coefficients are stable across specifications when the bandwidth is increased or decreased.

Second, we set the polynomial order to one in our central specifications. This is again for power reasons. Again, to ensure more complex functional forms of the function linking the running variable to the outcomes are not mistaken for discontinuities, we also check that the coefficients are stable when increasing the polynomial order using coefficient plots. In those coefficient plots, however, we never set the polynomial order above 2, since Gelman and Imbens (2019) showed higher order polynomials implied noisy estimates and poor coverage of confidence intervals.

C.4 Bunching analysis

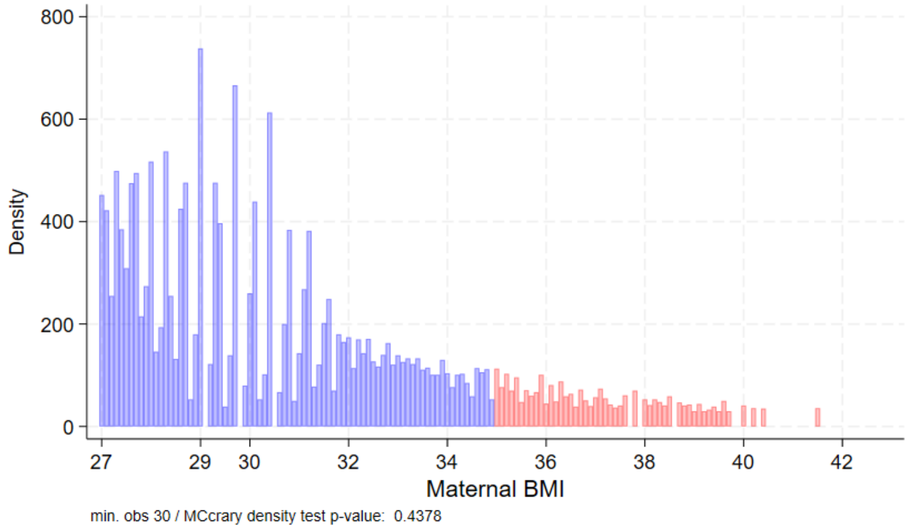
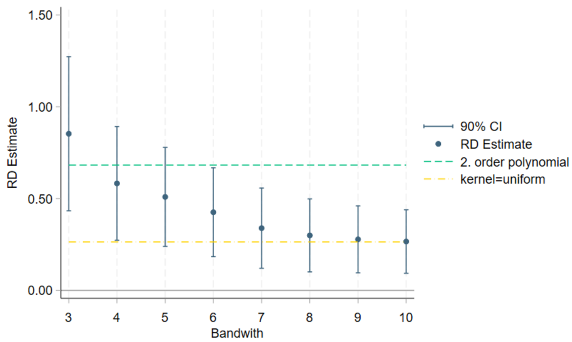


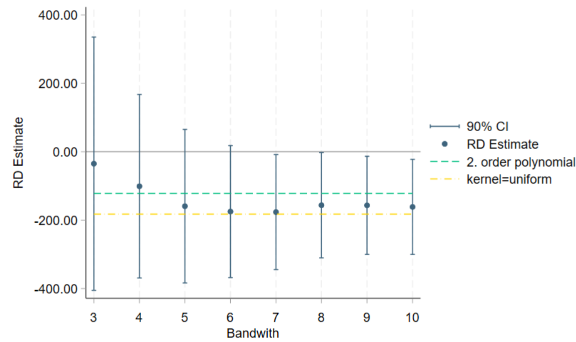
Fig. B7 Density of births with a given maternal BMI

Notes: The Figure shows the density of births around the 35 maternal BMI threshold. Blue bars indicate that the maternal BMI is below the threshold, red bars that the maternal BMI is above. Sample is the main one containing all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later. While our maternal BMI data is available up to three digits, we collapsed data at the first digit to abide with our provider rules. All bins with less than 30 obs are deleted.

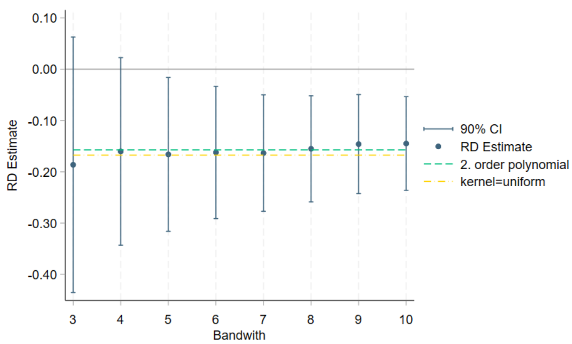
C.5 Coefficient plots for immediate outcomes



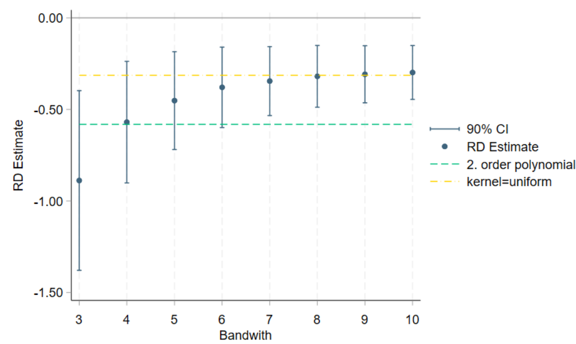
(a) Epidural



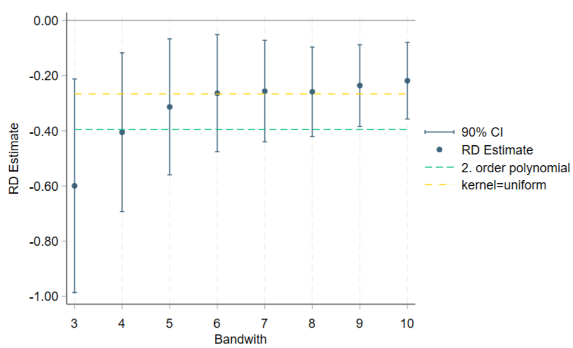
(b) Birth weight (grams)



(c) Birth weight≥4.5 kg



(d) Hemorrhage

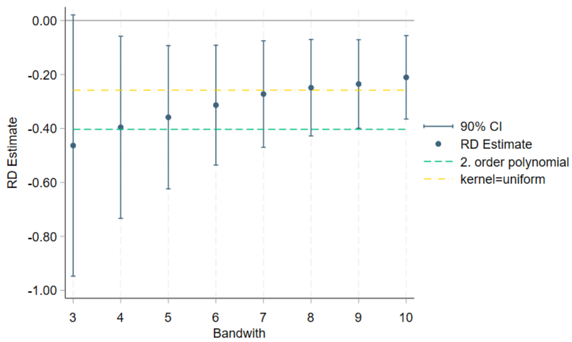


(e) Lacerations

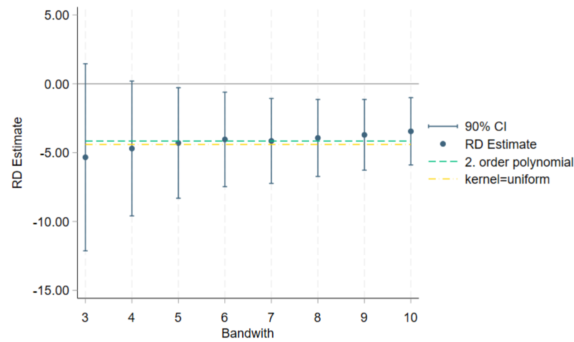
Fig. B8 Coefficient plots for immediate health outcomes

Notes: The figures are coefficient plots showing RD estimates from a fuzzy RDD with various outcomes and varying specifications. Point estimates with higher polynomial order and different kernel are computed with a bandwidth of seven, as in our main specification. Controls are included in all specifications. Standard errors are robust and clustered at the hospital-year level. Sample is the main one (all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later).

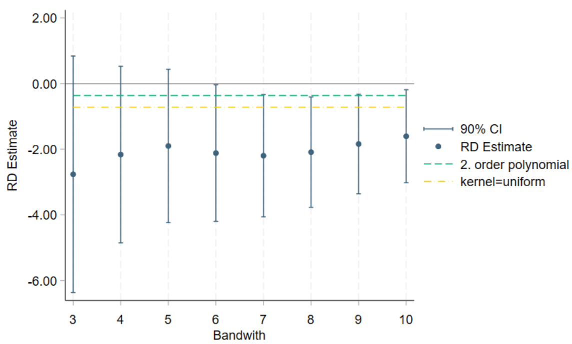
C.6 Coefficient plots for short-run outcomes



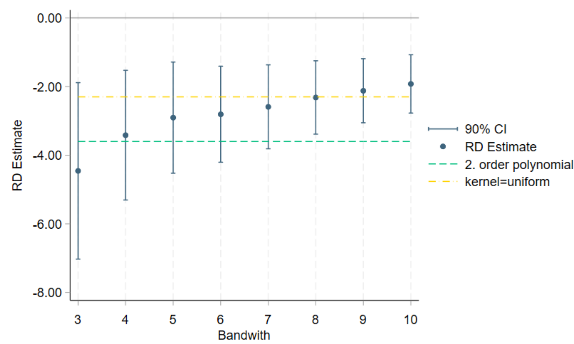
(a) EPDS>11



(b) EPDS score



(c) Nurse visits (Total)



(d) Universal Nurse visits

Fig. B9 Coefficient plots for various neonatal and first-year health outcomes

Notes: The figures are coefficient plots showing RD estimates from a fuzzy RDD with various outcomes and varying specifications. Point estimates with higher polynomial order and different kernel are computed with a bandwidth of seven, as in our main specification. Controls are included in all specifications. Standard errors are robust and clustered at the hospital-year level. Sample is the main one (all births in Denmark between 2011 and 2017 that occurred at 7 days after the EDD or later).