The Economics of Labor & Patients' Health Outcomes: Evidence from Childbirth in Germany

Ilka Gerhardts, University of Munich (LMU) *

Jan 2024

Abstract

This paper provides novel evidence on the causal effect of non-medically indicated induction on patients' health. The analysis is based on data for Germany, where profit-oriented reimbursement schemes and acute staff shortages imply strong incentives for birth interventions, leading to induced labor being twice as common in 2022 compared to 1985. Using two years of nationwide comprehensive hospital records, the empirical design allows identifying non-random and interdependent assignment of inductions and surgical interventions. The identification exploits intervention preferences of physicians who are as good as randomly allocated to healthy first-time mothers. The results reveal evidence for increased prevalence of perineal tearing and substantially impaired neonatal fitness as the result of non-indicated birth interventions.

JEL Classification: I10, B23

Keywords: health, obstetrics, incentive schemes, physician-induced demand

*Corresponding author. Email: Ilka.Gerhardts@econ.lmu.de.

1 Introduction

"Physicians serve the health of the individual and of the population. The medical profession is not a trade. It is by nature a liberal profession." (Model Professional Code for Physicians in Germany, 1997)

"[...]das Handicap ist die moderne Geburtsmedizin, die Geburt und Schwangerschaft zur Risikoaffäre macht." [1]

More than two thousand years ago, physicians declared through the Hippocratic Oath the benefits of the sick to be the sole objective of their profession (Tyson, 2001). In contemporary Germany, the ancient ethical principles are protected by stating medical services to be not for profit. This includes obstetric care, which refers to all treatments related to childbirth.

In practice, maternity units operate under pressure by profit-oriented reimbursement schemes paired with acute staff shortages, thus facing adverse incentives for birth interventions (Scharl et al., 2019; Feige, 2008).[2] Because maternity units manage the least predictable hospital events apart from emergency care, they are burdened with substantial non-refundable standby costs for staff-intense patient monitoring (Bruns, 2017). Thereby, hospitals conceding mothers an unassisted vaginal delivery do so at a loss, a dilemma providing incentives for labor induction or surgical birth interventions (Bruns, 2017). As of 2017, 40% of German hospitals provided obstetric care without breaking even (Bruns, 2014), while the share of birth interventions reached twice the size recommended by the WHO (2015).

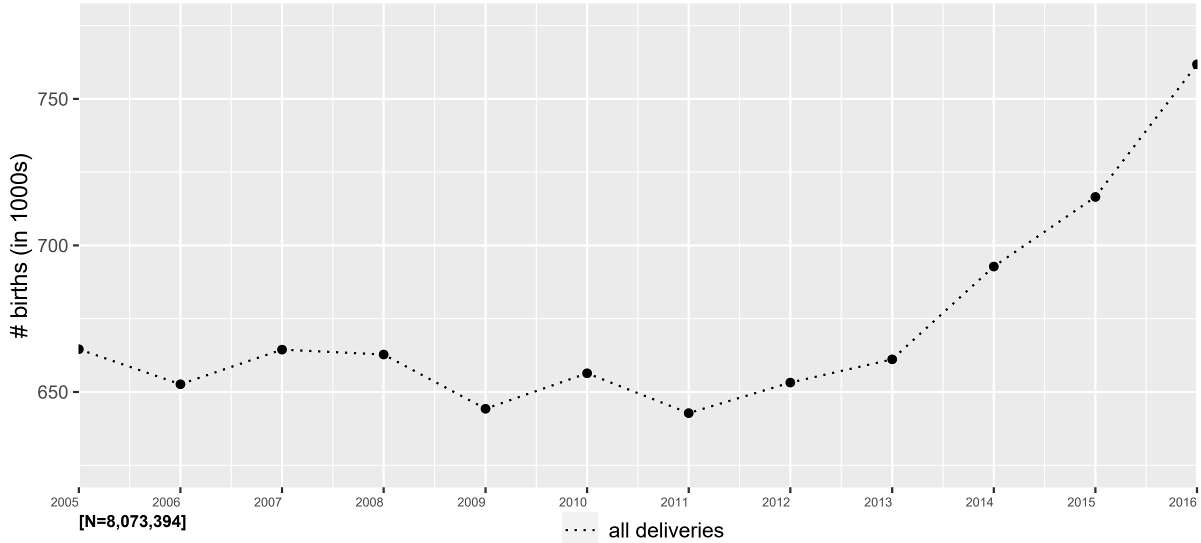
But what are the consequences of non-medically indicated induced labor for patients' health and a hospital's business operations, in particular, staffing capacities? Laying the ground for compelling causal evidence on the topic, this study applies a novel identification approach to the universe of German hospital births from 2015 and 2016. On the one hand, patients' health impacts are assessed, first and foremost in terms of 1) a severe laceration of the mother's birth canal, and 2) the neonatal APGAR score. On the other hand, the effects on a hospital's staffing capacities are primarily captured by 3) labor duration and 4) the postnatal hospital stay.

The rising birth intervention rates have triggered mostly observational evidence for physician-induced demand in the context of childbirth.[3] Very few large-scale causal studies have addressed

[1] *[...] it is modern birth medicine that renders the birth process and the pregnancy risky.* Own translation. Alfred Rockenschaub (2005). Former head of Ignaz Semmelweis Frauenklinik, Vienna. Known for ceasarian section (the surgical delivery through a mother's abdomen, henceforth: c-section) rates about 1% without inflating mortality.

[2] Since 2003, hospitals have been reimbursed based on the Diagnosis-Related-Groups (DRG) system, a flat rate-per-case scheme (Jürges and Köberlein, 2015). There is no (direct) reimbursement for inducing labor (InEK, 2021). Its economic appeal relative to spontaneous labor consists rather in reduced standby costs, i.e., non-refundable costs for staff kept in readiness but not called into action (like a surgeon for an unassisted birth). The rising trends in inductions and other main birth interventions are depicted in Figure 2.

[3] Table A.1 recaps studies targeting the causal impact of inductions.



Source: IQTIG German hospital birth records. To capture the hospitals' workload in the best possible way, 1) mothers transferred between hospitals are counted repeatedly and 2) still-born neonates are included. E.g., a mother transferred from hospital A to B who delivers a live- and a still-born twin at B is registered once in the records from A and twice in those from B. Own calculations.

Source: IQTIG German hospital birth records for 2005-2016. Own calculations.

the medical concerns associating induction with a prolonged and more harmful course of labor as well as adverse health outcomes after birth. Likewise, from an economic perspective, it remains unclear, if and to which extent a health impact of inductions aggravates staff capacity constraints due to additional patient monitoring.

Identifying the impact of non-medically indicated inductions is hard for several reasons. Among the most important ones: Interventions are likely non-randomly assigned and interdependent, e.g., choosing a pre-labor c-section foregoes induction but induction can be followed by c-section. To overcome these challenges, this chapter allows for multiple endogenous treatments.

To identify the sole and combined effects of induced labor, c-sections, and vaginal operations six instruments are considered, all of which are new to the health economics literature.

The first three instruments use variation within a given hospital and across obstetricians' preferences to perform a specific intervention. The preference for, e.g., inducing labor corresponds to the mean induction rate across an obstetrician's past deliveries. The fourth instrument exploits if a mother's predicted due date happens to be a working day or not because staff shortages are more acute on non-working days. The fifth instrument exploits if the incidence of a mother experiencing a pre-labor rupture of membranes happens during the night shift because the night shift suffers relatively more from under-staffing than the day shift. The last instrument exploits fluctuations of midwife shortages the moment a mother is admitted to a hospital.

The main findings are twofold. First, induction performed for non-medical reasons strongly impairs patients' health. Second, the adverse health effects imply a staff capacity burden easily overlooked by seminal capacity measures. As to immediate maternal health, induction makes high-degree perineal tearing 6% more likely. Specifically, induction followed by surgical intervention aggravates tearing so much that - given the distribution of single and combined inductions in our main sample - it outweighs the relief in tearing estimated for inductions alone. Besides, severe tearing due to a violent course of labor potentially requires postpartum or later-life surgery (Lydon-Rochelle et al., 2000; Gün et al., 2016; Zahn and Yeomans, 1990). In turn, birth canal surgery is associated with compromising future fertility (Halla et al., 2020; Gizzo et al., 2013; Norberg and Pantano, 2016). As to neonatal health, the detrimental effects (-2.2) found for the APGAR score, the seminal 0-10 range fitness range for newborns, exceed existing quasi-experimental findings, e.g., Lynch et al. (2019) by a factor of ten. Surgical interventions exhibit (weakly) negative health impacts, too.

By contrast, a hospital's staff capacity (measured by labor duration and the postnatal hospital stay) is weakly positively affected by induced relative to unassisted birth. Concretely, labor is estimated to shorten by 0.87 hours while a patient's postnatal hospital stay is not significantly impacted at all. In line with intuition, induction-related health compromise should translate into a staff capacity burden. Considering, e.g., just two routine health checks warranted by lower APGAR scores, a tentative back-of-the-envelope calculation suggests extra staffing costs of 11.8 million EUR p.a.[4] Finally, as expected, surgical interventions mechanically shorten labor and prolong mothers' and neonates' postnatal stays ca. 1.5 days. All in all, the labor length relief non-withstanding, the evidence points to negative intervention impacts rebounding (through impaired health) on staff capacity.

This study is the first to incorporate the endogenous and interrelated nature of all three major birth interventions. Thus, it complements the health economics literature in two main ways. First, the impact of induced versus spontaneous labor is cleanly identified. Second, it

[4]Underlying assumptions and computations are detailed in section 5.

provides a new benchmark for both, 1) surgical intervention effects identified simultaneously within the same framework, and 2) any birth intervention effect from the literature still relying on single-intervention identification.

Explicitly estimating the impact of induced versus unassisted labor is challenging. By construction, single-treatment identification defaults to comparing induced labor to any other birth mode after some waiting period, so-called expectant management. The few large-scale RCTs report mixed but predominantly positive effects of non-medically indicated induction. However, due to ethical restrictions, they are not blinded and prone to low or selective participation casting doubt on internal and external validity (Carmichael and Snowden, 2019).[5] Besides, autocorrelation of findings arises as trials of multi-side RCTs are referenced repeatedly in systematic reviews (Carmichael and Snowden, 2019).

By contrast, the limited number of large-scale quasi-experimental studies agree on weakly negative effects. Exploiting exogenous shifts in the timing of induction, Buckles and Guldi (2017), Lynch et al. (2019), and Gans and Leigh (2008) find a higher incidence of precipitous labor, birth injuries, etc. Buckles and Guldi (2017), and Jürges (2017) document null effects on c-section likelihood.

Finally, there exists a huge body of mixed observational evidence. If at all, there is some consensus on the detrimental health effects of 1) induced relative to spontaneous labor (Vahratian et al., 2005; Vrouenraets et al., 2005), and 2) induced labor after expectant management relative to induction alone (Harper et al., 2012).

This study's simultaneous identification and straightforward assessment of the relative impacts of birth interventions fills a gap in the literature. So far, Jacobson et al. (2020) provide the only study exploring the causal impact of (postponing) either inductions or c-sections but they do not identify the effects of intervening vs. not intervening at all, nor do they allow for interaction effects. They find small adverse effects on neonatal health. Despite this lack of scientific evidence, non-medically indicated induction is less restricted by medical guidelines than elective surgical procedures (DGGG, 2020b,a). The new findings contradict the marginalization of induction relative to c-section.

This study also benchmarks its multi-treatment estimates against the causal literature, thereby putting 1) the reliability of the single-treatment identification on debate, and 2) the value-added of the more involved multi-treatment identification into perspective. This is especially useful for the relatively broader evidence on c-section effects, like Card et al. (2018), Costa-Ramón et al. (2018), Costa-Ramón et al. (2019), Halla et al. (2020), and Jachetta (2016).

[5] The ARRIVE Trial (Grobman et al., 2018), a recent multi-site RCT with a global policy impact, shows a significant decrease in the likelihood of (non-/emergency) c-section (19 vs. 22%), prolonged labor (20 vs. 14 hours), and a slightly shorter maternal postnatal hospital stay. However, only 23% of eligible women participated, with roughly 10% official non-compliers. Anecdotally, Goer (2018) proposes even higher physician-induced non-compliance in the control group.

Instrumented single-treatment evidence is found to deviate substantially from multi-treatment findings. Besides, single-treatment estimates differ a lot in sign, size, and significance depending on the instrument used. Therefore, with interrelated birth interventions, a multi-treatment model yields more reliable results.

This study meets key interests of public policy. First, awareness that hospital demand for intervention aggravates rather than alleviates capacity constraints is crucial to prevent snowballing effects generating even more birth interventions for non-medical reasons (Allen et al., 2006; Bonsack et al., 2014). Second, inferring some implications of inductions for subsequent fertility is likewise of principal interest: The suspected long-term effects of perineal damage counteract costly fertility incentives (parental leave policies, child allowances, etc.) and, most likely so, when the first child is born (Bruns, 2017).[6]

The remainder of the paper is structured as follows. section 2 describes the data. section 3 details the identification approach. section 4 presents estimates for the health effects of inductions while section 5 turns on a hospital's staff capacity effects. section 6 concludes by discussing the policy implications of my analysis.

2 Data

To analyze the effects of induced labor on patients' health and hospital staff constraints, this study uses nationwide mother-child level hospital records collected and cross-validated by the IQTIG institute.[7] The focus lies on the records from 2015 through 2016, for each of which the pregnancy, the entire hospital stay, and the course of delivery is documented meticulously.

First, three binary treatments represent the main birth intervention types. *Induced labor*, the intervention of principal interest, is expressed as a pooled indicator showing if any form of induction has been conducted or not. It adopts the clinical definition of induced labor, which includes, most prominently, mechanical rupture of membranes and hormonal labor stimulation by medication, but excludes minor interference like cervical ripening (see Mishanina et al., 2014, for details on induction methods like Oxytocin dose or membrane sweep).[8] Second, *Non-emergency C-section* comprises all but emergency c-sections. On the one hand, this definition does not rely on possibly strategic hospital labeling of c-sections as planned vs. spontaneous (Card et al., 2018). On the other hand, excluding emergency c-sections allows focusing on c-sections with medical

[6] On average, first births last 6.7 hours (and second births only 4.6 hours), which makes them more susceptible to intervention: all major birth interventions are way more common among first births (Table 1).

[7] The *Institut für Qualitätssicherung und Transparenz im Gesundheitswesen* is an independent scientific research institute with a legal mandate from the German Federal Ministry of Health to evaluate hospital care quality. Independent of public or private sponsorship, all officially registered hospitals are obliged to report their data for external validation and evaluation. In Germany as of 2010, only 2% of births took place outside a hospital (Kolip et al., 2012). It is a legal requirement to cite the data as follows. "Es wurden Daten aus Qualitätssicherungsverfahren gemäß §136 SGB V des Gemeinsamen Bundesausschusses verwendet."

[8] Henceforth, *unassisted* labor is defined as spontaneous labor, maybe augmented or slowed down as the delivery proceeds. Likewise, *unassisted* birth precludes any of the three main birth interventions.

scope as a treatment.[9] Third, a binary indicator captures vaginal operations in a wider sense. It combines classical vaginally operative birth assistance, i.e., by forceps, vacuum, or spatula, with episiotomy, which is a surgical cut to prevent perineal damage by spontaneous tearing. Interestingly, vaginal operations are more common (32%) than inductions (28%) or c-sections (26%) in the main analysis sample (Table 1, column (4)). Notably, despite targeting the impact of non-medically indicated interventions, none of the treatment indicators relies on reported indications possibly manipulated to justify intervention framing low-risk births as pathological (Jürges and Köberlein, 2015; Bradford et al., 2007; Kolip et al., 2012; Feige, 2008).

The dependent variables are a binary indicators for maternal health, namely the incidence of 1) high-degree perineal tearing, as well as ordinal measures of 2) the APGAR score five minutes after birth[10], 3) the hours of labor duration, and 4) the days of the (maternal and neonatal) postnatal hospital stay. *Perineal Tearing (III/IV)* encompasses also wound hematomas, reflecting either severe perineal tearing or episiotomy itself. As episiotomy is likely not randomly assigned, it is included in the treatment *Vaginal Operations.* Importantly, even though a mother might select into episiotomy while another mother would bear perineal tearing instead, the wound hematoma signals the severe course of labor for either one.

[9]As emergency c-sections refer to mortal danger, they are much harder to recode strategically, first and foremost due to stricter reporting requirements and a different workflow. Originally, emergency c-section was targeted as an outcome of induced labor but there was too little variation.

[10]The APGAR score (0-10) increases in healthy skin color and correct functioning of lungs, heart, muscles, and reflexes (Card et al., 2018).

Table 1: Characteristics of Births by Non-Missingness, Preconditions, and Birth Order

	1st & 2nd births	**with non-missing central variables**					
		1st & 2nd births		zero-precondition 1st births		zero-precondition 2nd births	
		= 1	Δ	= 1	Δ	= 1	Δ
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
maternal characteristics							
german (y/n)	0.80	0.81	-0.06***	0.81	0.00***	0.78	0.04***
single (y/n)	0.11	0.11	-0.07***	0.11	-0.01***	0.08	0.03***
low socioeconomic status (y/n)	0.80	0.81	-0.10***	0.80	0.01***	0.85	-0.05***
age	30.28	30.26	1.01***	28.16	3.08***	29.61	0.78***
bmi	24.72	24.72	0.15**	24.10	0.90***	24.43	0.33***
pre-pregnancy weight (kg)	68.78	68.79	-0.11	67.06	2.51***	68.02	0.89***
gestational age (#days)	275.17	275.27	-5.62***	279.81	-6.75***	279.24	-4.74***
prenatal care (#doctor visits)	11.14	11.13	0.18***	11.53	-0.57***	11.15	0.00
pre-care start >12th week (y/n)	0.07	0.07	0.00	0.08	-0.01***	0.08	-0.01***
neonatal characteristics							
birth weight (g)	3328.55	3332.10	-189.89***	3415.72	-126.43***	3527.25	-231.53***
hospital characteristics							
emerg. cs time >20 min (y/n)	0.01	0.01	0.01***	0.01	0.00***	0.02	0.00***
emerg. cs time <3 min (y/n)	0.00	0.00	0.00*	0.00	0.00*	0.00	0.00*
health outcomes							
emergency c-section (y/n)	0.01	0.01	0.01***	0.01	0.00	0.01	0.01***
perineal tearing (III/IV) (y/n)	0.01	0.01	0.00***	0.02	-0.01***	0.01	0.01***
APGAR score (5 min.)	9.6	9.6		9.7		9.8	
hospital capacity outcomes							
labor duration (#hours)	4.85	4.91	-3.47***	6.69	-2.68***	4.57	0.33***
maternal postnatal stay (#days)	3.45	3.44	0.57***	3.36	0.14***	2.66	0.93***
neonatal postnatal stay (#days)	3.12	3.12	-0.27***	3.19	-0.11***	2.59	0.62***
treatments							
induced labor (y/n)	0.22	0.23	-0.03***	0.28	-0.07***	0.20	0.03***
non-emergency c-section (y/n)	0.31	0.34	0.32***	0.26	0.13***	0.08	0.32***
vaginal operations (y/n)	0.20	0.21	-0.01***	0.32	-0.16***	0.11	0.11***
IV staff capacity							
non-working day due date (y/n)	0.33	0.33	0.00	0.33	0.00	0.34	0.00**
PROM 8pm-4am (y/n)	0.12	0.13	-0.02***	0.15	-0.03***	0.10	0.04***
midwife shortage at arrival [0,1]	0.57	0.57	0.29***	0.58	-0.01***	0.58	0.00
IV obstetricians' preferences							
preference induced labor [0,1]	0.23	0.23	0.00	0.23	0.00***	0.23	0.00***
preference non-emerg. cs [0,1]	0.34	0.37	0.03***	0.34	0.05***	0.23	0.16***
preference vaginal operation [0,1]	0.20	0.20	-0.01***	0.20	-0.01***	0.20	0.00**
N	1,076,763	561,572		177,215		81,620	
N obstetricians' preferences	412,228	206,199		66,916		27,457	

Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. (Differences in) means for the central analysis variables based on all births, births restricted to non-missing central variables, and 1st and 2nd births without pregnancy or birth risks ante partum. See Table A.2 and Table A.3 for details on sample and variable construction. Binary indicators *yes/no* abbreviated as *y/n*. *PROM* refers to *prelabor rupture of membranes*. *emerg. cs* is short for *emergency c-section*. For the APGAR score means but no differences are available.

Given various stages of labor are observed, *Labor duration* (in hours) can be computed in terms of the staffing-relevant period a mother has contractions, as opposed to the reimbursement-relevant period of pushing contractions (DHV and DGGG, 2020).[11] The duration of the postnatal hospital stay is measured for both, a mother and a neonate, by counting the days elapsed between the completion of delivery and hospital discharge.

For causal identification, a total of six instruments is created which are discussed in depth

[11] Detailed information on labor progress is one advantage of the IQTIG data compared to other natality records. For, Card et al. (2018) do not observe labor at all and proxy its duration by counting the hours from hospital admission to completed delivery.

in 3.3. Parsimonious baseline covariates and extensions are detailed in the notes to Table A.11. Complete variable specifications are given in Table A.2.

To exclude interventions planned for strictly medical reasons, the sample is restricted to zero-precondition births, i.e., mothers-to-be (henceforth: mothers) without any known pregnancy or birth risks antepartum, thereby focusing on normal presentation singleton pregnancies at term. Moreover, focusing on first births minimizes heterogeneous influences from prior parity experiences. Finally, conditioning on non-missing central estimation inputs yields the main analysis sample of 177,215 observations. Table A.3 depicts an overview of sample specifications. Balance Table 1 compares the central variables across samples, e.g., all first and second births versus the main analysis sample by non-/missingness. Despite confirmed high data quality overall, this matters. Hospitals cannot oblige mothers to provide non-obstetric information, and column (3) suggests, e.g., the socio-economic or marital status may be selectively missing. Table A.5 in the appendix balances all core characteristics across strata created for endogeneity and heterogeneity checks.

3 Empirical Approach

This section sheds light on the institutional background determining a hospital's incentives for physician-induced birth intervention demand and sets out the empirical strategy. After introducing a simple OLS benchmark framework, an instrumental variable strategy is developed to identify the causal impact of non-medically indicated induced labor.[12]

3.1 Institutional Background & Intervention Incentives

The identification strategy exploits supply-side incentives for induction at the hospital and the obstetrician level. Inducing a woman's labor plays a key role in hospital management, first and foremost to alleviate staff shortages and to forego standby costs the hospital is not compensated for (Bruns, 2017; Feige, 2008).[13]

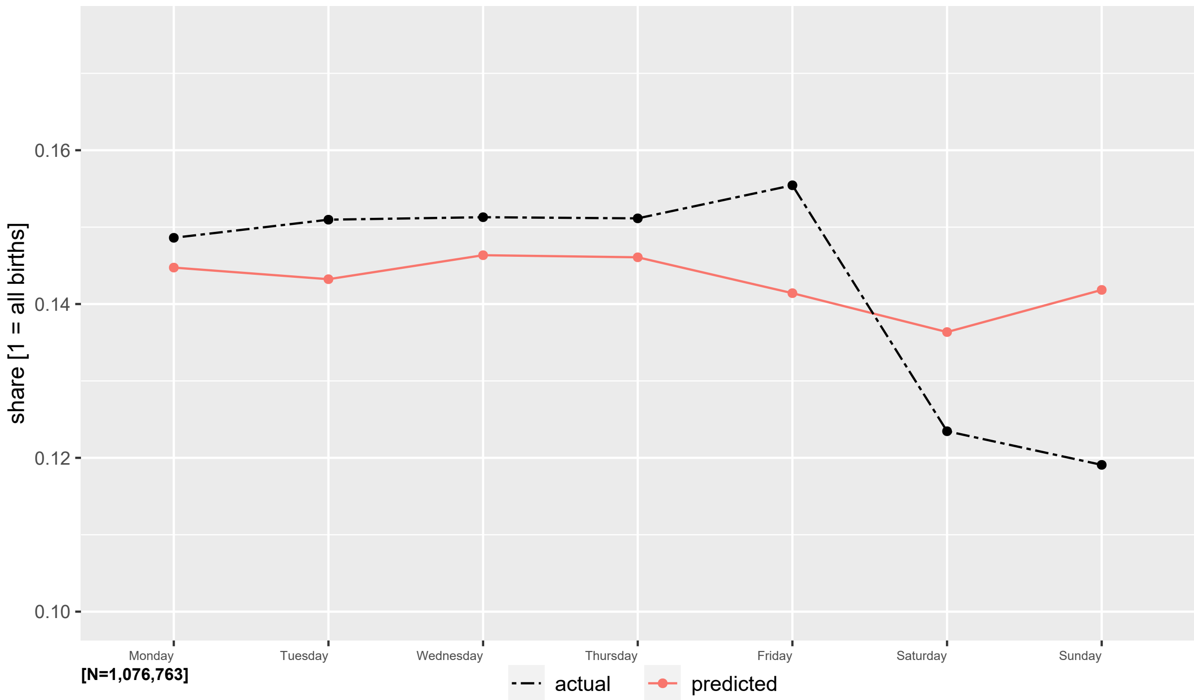
On behalf of an obstetrician, despite flat-rate pay, performing an induction could be appealing for many reasons.[14] The multitude of subjective incentives creates variation in obstetricians'

[12]Due to the highly confidential data base the empirical analysis follows a legal protocol. First, the data user commits herself to a statistical analysis plan, second, the corresponding code is run at IQTIG, and third, all output - ex-ante requested without any data insights - is released to the user after legal approval by the Gemeinsame Bundesausschuss (G-BA). Ongoing work by Gerhardts (2024) updates the analysis in response to these first findings.

[13]Standby costs are poorly documented. Ignoring standby costs completely, the average DRG-based profits of an unassisted birth (n=100, 94 uncomplicated) are 1847-1674=173 EUR accounting for 556 EUR reimbursement-relevant staff costs (Rummel (2007)). While standby costs of just 24% (or 173 EUR) relative to 76% (or 556 EUR) would turn the profit into a loss, a more realistic standby cost estimate of up to 70% (Bruns, 2017) implies a sizable loss.

[14]Lutz and Kolip (2006) and BZgA (2005) summarize alternative forensic, demographic, economic, cultural, societal, technological, and other supply and/or demand-side incentives for inducing labor.

preferences for intervention. Given decision scope from medical guidelines (Bruns, 2017) paired with variation in capacity constraints and intervention preferences, mothers are heterogeneously exposed to physician-induced demand.



Source: IQTIG German hospital records for 1st and 2nd births in 2015-2016. There are <1% of deliveries with a hospital-corrected due date (not shown). Own calculations.

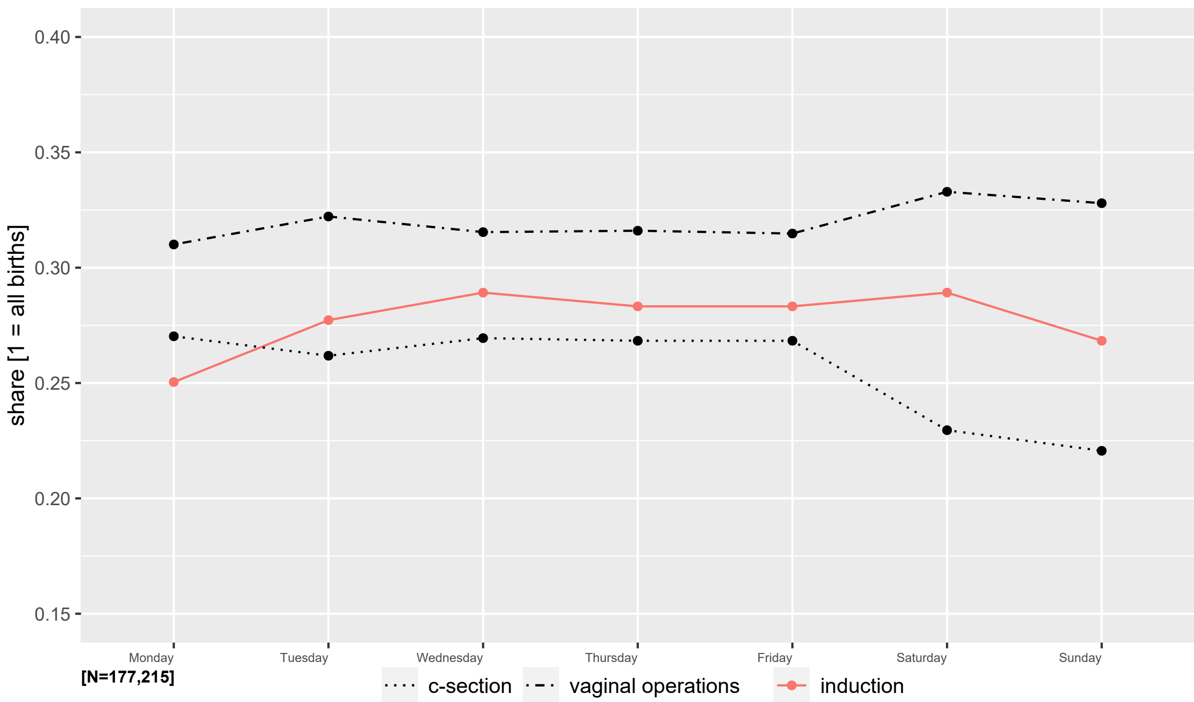
In a first step, to overcome common challenges upon visualizing physician-induced demand (Dranove and Wehner, 1994), Figure 3 plots due dates and completed births across weekdays. While due dates set out the biological benchmark distribution, actual (completed) deliveries should reflect potential man-made birth timing at a hospital, thereby causing diverging distributions. We see that most due dates are Wednesdays (used as the 100% benchmark) closely followed by Thursdays, while the fewest due dates are Saturdays (3% less than Fridays). The total range of fluctuation is limited to 7%. Intuitively, the non-uniform pattern could be driven by leisure time dependent menstruation cycles, etc.[15] Completed births somewhat follow the due date distribution from Monday through Thursday but running up to the weekend the patterns diverge. Most births occur on Fridays, followed by a drastic drop of 21% on Saturdays and - taking Fridays as the benchmark - a further drop of 3% on Sundays.

In a second step, to visualize work shift-specific intervention demand, Figure 4 and Figure 5 plot (the shares of) un-/assisted births across weekdays and hours of the day respectively. Induced *births*[16] are least frequent on Mondays, their share rises and stays up from Tuesdays

[15] The due date prediction is normed to 40 weeks from the 1st day of the last menstruation.

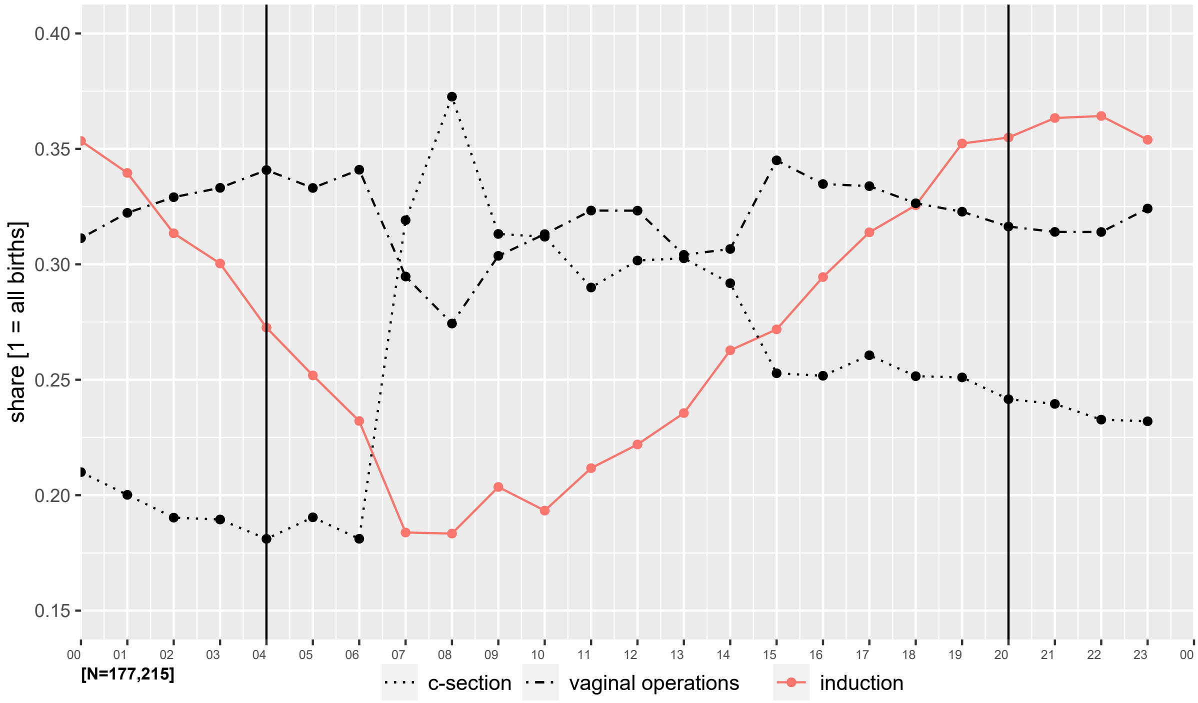
[16] To tentatively assess the timing of inductions themselves, these patterns need to be lagged by 13-17 hours (the mean interval between induction and delivery in a similar sample studied by Levine et al., 2016). For example, a

through Saturdays before dropping down on Sundays. Induced *births* are also least common between 07 and 08 am, then their share continuously increases till peaking between 09 and 10 pm, before decreasing again. The patterns of vaginal operations seem largely mirrored by c-sections, although c-section oscillate more strongly. We see the least (most) births with vaginal operations (c-sections) on Mondays, rather stable shares throughout the week, and a distinct increase (decrease) on the weekend. Similar substitution effects can be seen across daily hours, where the fewest c-sections are performed before 06 am, then they peak already at 08 am and drop drastically starting from 03 pm. Accordingly, vaginal procedures are rarest at 08 am before overtaking c-sections in frequency at 10 am again. In short, the distribution of births following any (or a combination) of the three main intervention(s) is suggestive of some non-random timing of birth assistance.



Source: IQTIG German hospital birth records for 2015-2016 restricted to the main analysis sample of zero-precondition first births (detailed in the notes to Table A.11). Own calculations.

17-hour-lag (depicted in Figure Figure A.3) yields an intervention pattern in line with previous literature where inductions are concentrated 1) on the early morning hours maximizing delivery likelihood during the day shift as well as 2) on Mondays through Fridays shifting delivery away from weekends (Halla et al., 2020; Costa-Ramón et al., 2018). However, inferring induction timing from birth timing is imprecise as the induction-birth interval depends on the induction method(s) and further (minor or major) interventions applied.



Source: IQTIG German hospital birth records for 2015-2016 restricted to the main analysis sample of zero-precondition first births (detailed in the notes to Table A.11). Own calculations.

3.2 Benchmark: Pooled OLS

As a baseline, an OLS regression with three interdependent treatments is run, thereby accounting for substitution effects and complementarities among birth interventions. Specifically, induced labor could be either substituted by a planned c-section or complemented by a spontaneous (non-emergency) c-section.[17]

Despite an ongoing debate as to whether inductions cause c-sections or not (Table A.1), their joint effect has been broadly neglected. Seminal IV studies identifying c-section impacts, either ignore (Card et al., 2018), control for (Costa-Ramón et al., 2018), or drop inductions from the sample (ibid.), none of which overcomes the problem that induction is endogenous, too. Jacobson et al. (2020) distinguish unassisted, induced vaginal, and c-section deliveries, thereby mechanically mixing the impact of failed inductions with that of a c-section alone. By contrast, this new model supplements the main interventions by a single cumulative induction-plus-surgery indicator.

[17]Equation 3, the originally targeted model allows all two- and three-way intervention interactions. However, in practice, after *failed* vaginal operations a spontaneous (but non-emergency) c-section is medically only feasible before the fetus descends too far into the birth canal. In the main sample of 177,215 births, there are 45 doubly surgical (and 15 triple intervention) cases causing extreme multicollinearity issues. Follow-up work by Gerhardts (2024) settles on an intermediate model that dismisses rare interactions from the original specification but still manages to disentangle induction followed by c-section (9%) vs. induction followed by vaginal procedures (8% of birth modes, see Table A.8).

$$Y_m = \beta_0 + \beta_1 * IL_m + \beta_2 * CS_m + \beta_3 * VO_m + \beta_4 * IL_Surgery_m$$
$$+ \underset{1\times k}{\mathbf{x}'_m} \underset{k\times 1}{\delta} + \underset{1\times s}{\lambda} \underset{s\times 1}{\mathbf{1}} + v_m \quad (1)$$

where

covariates $\mathbf{x}_m \equiv \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_{k-1} \end{bmatrix}$, sets of fixed effects $\lambda \equiv \begin{bmatrix} \lambda_1 & \dots & \lambda_s \end{bmatrix}$

- Y_m: outcome of mother (or her neonate) m
- $IL_m,\ CS_m,\ \ VO_m \in \{0,1\}$: =1 if mother m has an induction (and maybe other interventions), a non-emergency c-section (dto.), or vaginal operations (dto.) respectively; 0 else
- $IL\_Surgery_m$: =1 if mother m has an induction followed by a non-emergency c-section, vaginal operations, or both; 0 else
- details on pre-determined controls, fixed effects, and cluster-robust standard errors given below Table A.11

OLS estimates for induction health impacts vary substantially in terms of size, sign, and significance (Bonsack et al., 2014; Coatesid et al., 2020; Axt-Fliedner et al., 2004). This study accounts for interrelated birth interventions and a parsimonious set of core controls inspired by Card et al. (2018). Related estimation designs by Card et al. (2018), Halla et al. (2020), Buckles and Guldi (2017), Jürges (2017), Schulkind and Shapiro (2014a), and Costa-Ramón et al. (2018) are also cautious about adding many different sets of fixed effects even though non-linear trends, seasonality of births across months, weekdays, and daily hours, as well as hospital-specific intervention effects, are well-known. Accounting for binary outcomes and possible error correlation up to the county level (due to mothers selecting into hospitals), cluster-robust standard errors are reported.

Nevertheless, the OLS benchmark regression likely yields biased estimates as interventions are non-randomly assigned. For an up-/downward bias, i.e., over-/understating the adverse health effects, inductions without medical advantages needed to be concentrated on mothers with worse/better expected health outcomes. Adverse health outcomes are consistently negatively correlated with socioeconomic status (Jeong et al., 2020). For zero-precondition first-time mothers of lower socioeconomic status more doctor visits are registered despite the belated start of prenatal care, see Table A.4, which hints at curative rather than preventive appointments. In the literature, the correlation between socioeconomic status and inductions varies across countries (Carter et al., 2020).[18] Intuitively, a concentration of inductions on women with lower socioeco-

[18]One key risk factor reflected in socioeconomic status is nutrition quality (Wolfe et al., 2011). Zero-precondition births exclude mothers with severe obesity and the core controls account for (even non-linear effects of) height and weight, and BMI. However, BMI is just a (noisy) proxy for dietary quality, i.e., even a slim person could have bad eating habits.

nomic status could be rationalized by, e.g., physician-induced demand increasing in information asymmetries, thereby overstating an adverse health effect. Vice versa, an adverse health effect would be underestimated if mothers with high socioeconomic status got more high-tech medical care and thus more exposure to false positives about the fetus' well-being. Comparing the unconditional means across samples, zero-precondition first births to mothers with lower socioeconomic status are equally often induced as the main sample (but more prone to c-sections), see Table A.4. In line with prior literature (O'Dwyer et al., 2013; Carter et al., 2020), we further see inductions to be centered on slightly older mothers but less common among single mothers (23%). By contrast, c-sections are more frequent (around 27%) for both groups relative to the overall sample (21%). This is suggestive of (un-)observed differences likewise associated with birth outcomes even among zero-precondition first-time mothers.

3.3 Instrumental Variable Estimation (IV)

To resolve self-selection into multiple, possibly combined interventions, the three major interventions and induction followed by surgical intervention are instrumented.[19] Two alternative sets of three instruments each are discussed, first in terms of exogeneity, followed by relevance.[20] A brief remark on monotonicity concludes the identification discussion.

3.3.1 Instrument Exogeneity & Exclusion Restriction

1. A Set of Instruments based on Obtetricians' Intervention Preferences

   This set of instruments uses an obstetrician's preferences to perform inductions, c-sections, and vaginal operations (similar to Bhuller et al. (2020) in another context). Preferences are measured via the obstetrician's average rate of performing the respective intervention in all past deliveries. The idea is that an obstetrician has both, institutional decision scope on whether to offer intervention as well as influence on a mother's consent due to physician-patient information asymmetries.

   The first requirement of the exclusion restriction is random obstetrician assignment. Addressing concerns about mothers selecting into a hospital for its intervention reputation warrants including hospital fixed effects.[21]Next, considering within-hospital randomness, restricting the sample to zero-precondition first births is important. On the one hand,

[19]Whether birth interventions are endogenous (and IV estimation preferable) is not tested formally because the Wu-Hausman Test/ Durbin Score does not adapt easily to this set-up. Results from a workaround exploring the regression-based approach reported after Stata's *ivregress* command (Cameron & Trivedi 2005) are available upon request.

[20]Discussing exogeneity, "as good as randomly assigned conditional on (core) covariates and fixed effects" is shortened to "randomly assigned" for the sake of simplicity.

[21]This was done only for the original model Equation 3 in Table A.13.

zero-preconditions rule out skill-based obstetrician assignment, i.e., the matching of (unobserved) medical skills to heterogenous maternal health records. On the other hand, first-birth mothers are less likely to request a specific obstetrician based on prior experiences. However, learning about physician-specific intervention histories mothers might try to pick an obstetrician matching their preferences. Considering the organizational workload of German hospitals such selection seems unlikely but not impossible.

Therefore, the subsample of mothers rejected by one and transferred to another hospital is analyzed because those mothers could neither choose the hospital nor the obstetrician. Table A.5 shows transferred women to be more often single, older, and of higher socio-economic status, delivering relatively lower birth weight babies, and experiencing a lot more inductions or (even emergency) c-sections. Likewise, the subsample of mothers not presented to an obstetrician during the prenatal period is checked on. Albeit overall more similar to the main sample, these mothers are way less likely to be induced (Table A.5).[22]

The second requirement of the exclusion restriction is that a given obstetrician's intervention preference may influence birth outcomes only through altered intervention likelihood. Therefore, a problematic scenario would be, e.g., a mother staying with her assigned obstetrician but refusing to collaborate with him upon learning of his preference for intervention, thereby provoking an emergency c-section. Probably more salient in this context is the gatekeeper problem (Maestas et al., 2013) meaning that obstetrician assignment could be a packaged treatment including intervention preferences but also systematic skill differences. For instance, an obstetrician's high c-section preference might result in less experience and fewer skills in handling vaginal deliveries. Reassuringly, comparing unassisted and induced labor relies on a similar skill set. Nevertheless, the sample of (otherwise low-risk) first-time mothers delivered pre-arrival to hospital could be insightful. In this sample, unassisted delivery is the default with induction (c-section) rates as low as 9% (18%, see Table A.5) independent of an obstetrician's intervention preference (which is limited to influencing the type of intervention). Therefore, the estimated impact of a high induction preference should reveal other potentially influential characteristics specific to these obstetricians.[23]

2. A Set of Instruments based on Hospital Staff Capacity

Instrument: Midwife Shortages upon Maternal Arrival at a Hospital

The idea is that a mother arriving at a hospital where all midwives are busy is more likely to not get assigned a midwife at all, which in turn makes her more prone to induction.

[22] The corresponding subsample regressions were only run for the original model Equation 3, see Table A.12 and Table A.15.

[23] Sample-specific reduced form regressions - available upon request - were only run based on Equation 3. Another promising subsample to test intervention preference-dependent medical skills consists of (otherwise low-risk) first-time mothers suffering pre-/eclampsia. The high blood pressure condition provokes seizures and ranges among the strongest medical indications for induction or surgical delivery. However, the sample size was <10.

The first requirement of the exclusion restriction is that mothers do not selectively arrive at a hospital in response to within-day minute-wise fluctuations in midwife shortages arising from ongoing deliveries there. Despite maternal midwife and hospital selection, random assignment seems plausible as a mother usually cannot observe current midwife shortages at a hospital, and much less so before being admitted herself. To capture unpredictable variation, midwife shortages are defined as the share of *current* deliveries without midwives.[24] The newly arriving mother herself is excluded from the shortage measure. Otherwise, if she went into labor pre-admission, her choice to bring a midwife along or not would bias the measure. Addressing concerns arising from a mother selecting into an, e.g., high-quality hospital guaranteeing a midwife to each patient upon arrival, motivates the inclusion of hospital fixed effects. Furthermore, subsample regressions for hospitals forbidding in-patient midwives rule out pre-determined mother-midwife constellations unaffected by whichever midwife shortages prevail upon hospital admission.[25]

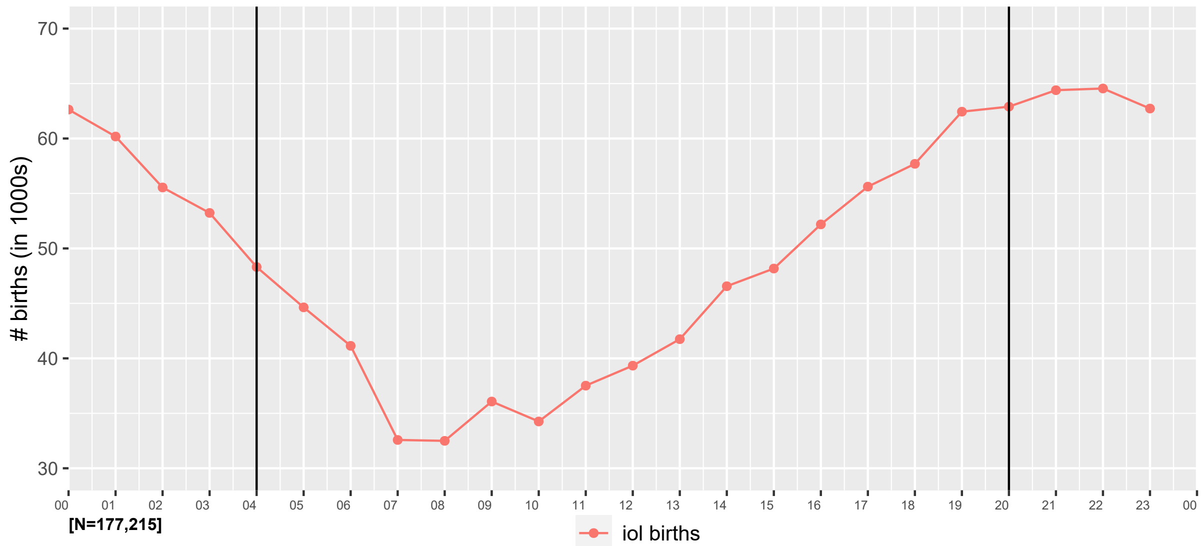
The second requirement of the exclusion restriction is that facing a given midwife shortage may influence birth outcomes only through altered intervention likelihood. To meet this condition, the midwife shortage prevailing (not the midwife assignment itself) upon arrival should not determine whether, e.g., a mother arranges to get certain anesthesia she would not have asked for otherwise.

Instrument: Pre-labor Membrane Rupture during a Hospital's Night Shift

The idea is that staff shortages are relatively more acute at night making scheduling of births more attractive, especially after a membrane rupture requiring intense monitoring otherwise. Figure 6 plots the within-day distribution of induced *births* confirming the extent to which inductions are used to schedule births around the clock. Figure 7 represents a close-up of two groups, namely all births following a pre-labor membrane rupture and a subset of those that were also induced. We see that the two groups co-move to some extent, although membrane ruptures oscillate five times as strongly within a day. Both groups reach their minima around noon, five hours later than induced births overall (Figure 6). Thus, pre-labor membrane ruptures shape part of the induction allocation beyond obstetricians' control.

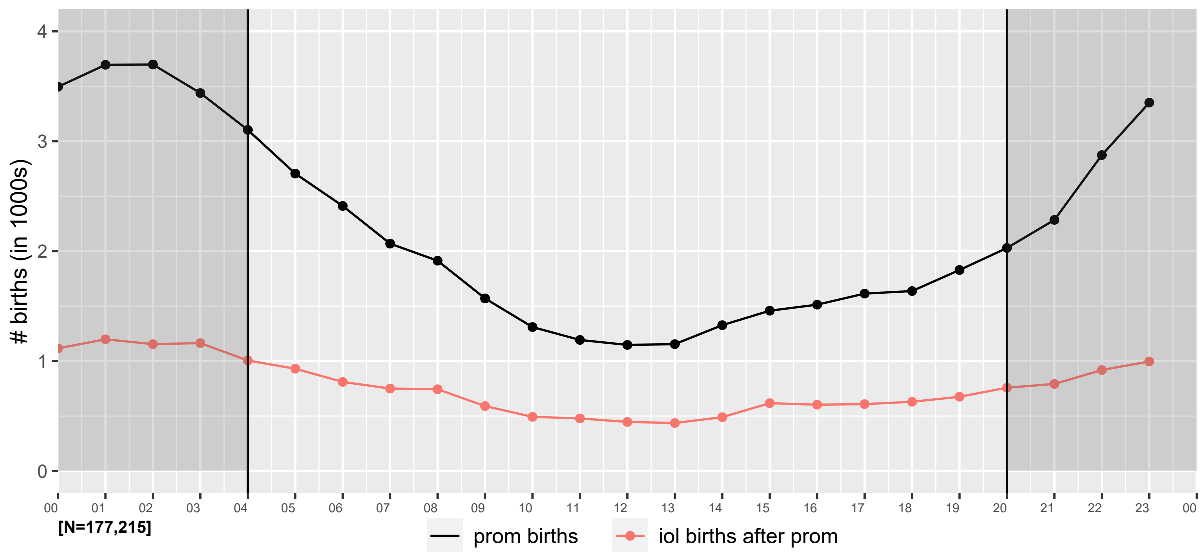
[24] Midwives are commonly assigned to several mothers simultaneously. Therefore, to capture acute and unpredictable shortages, the instrument is defined in terms of ongoing deliveries, i.e., the most care-intense periods, instead of counting midwives not assigned to a mother yet.

[25] This was done only for the original model Equation 3 in Table A.13.



Source: IQTIG German hospital birth records for 2015-2016 restricted to the main analysis sample of zero-precondition first-births (detailed in the notes to Table A.11). ***iol births*** refer to induced births. Own calculations.

Figure 7: Births with Pre-Labor Membrane Ruptures & Subsequent Inductions Across Daily



Source: IQTIG German hospital birth records for 2015-2016 restricted to the main analysis sample of zero-precondition first-births (detailed in the notes to Table A.11). ***prom births*** refer all to births following a prelabor membrane rupture, ***iol births after prom*** refers to a subsample of these births that are also induced. Own calculations. Please note that this figure is zoomed-in by 10 compared to Figure 6.

The first requirement of the exclusion restriction is that a given mother's membrane rupture falls randomly into the hospital's day or night shift. If within-day timing of membrane ruptures was randomly allocated to mothers, we would expect it to be comparable across maternal strata. Figure A.4 and Figure A.5 rule out influential daily working or exercising routines by plotting the daily distributions of births following membrane ruptures stratified by a mother's employment and fitness status.[26] Optimally, the timing of *membrane ruptures* (not subsequent births) should be compared as the endogenous allocation of interventions contaminates birth timing. Reassuringly, a few spikes around midday non-withstanding, even the daily distributions of births are well aligned across maternal strata.

The second requirement of the exclusion restriction is that a membrane rupture happening in either shift may impact birth outcomes only through changed intervention likelihood. That means mothers should react similarly to the rupture, e.g., by entering into the hospital as soon as possible instead of waiting for the morning to come. This seems plausible out of fear for the fetus' well-being. It is also feasible because the emergency ambulance is covered by public insurance in Germany.

Instrument: Due Date on a Non-working Day

The idea is that staff shortages are more likely on weekends and holidays, which makes scheduling labor onset for mothers due on these days relatively more appealing to a hospital.

The first requirement of the exclusion restriction is that neither parents nor physicians influence the due date's non-working day status (random assignment). In practice, the condition implies conceiving parents should not be targeting a non-working day for birth based on the due date prediction formula. Given the due date is criticized for poor precision, such a rationale seems unlikely, even if parents had non-working day preferences. However, maternal characteristics like specific working habits could influence the onset of menstruation cycles during the week. Reassuringly, across socio-economic status, no specific due date patterns emerge in Figure A.2. Moreover, a gynecologist predicting the due date may not change it upon noticing a birth is due on, e.g., a Sunday. Likewise, hospitals may correct the due date prediction only for medical reasons and not to justify early-on interventions. Reassuringly, the share of mothers with a hospital-corrected due date is negligible (well below 1%).

The second requirement of the exclusion restriction is that having been assigned a non-working day due date may influence birth outcomes only through altered intervention likelihood. This condition implies that, e.g., a mother due on a Sunday may not educate herself about options of anesthesia fearing a tougher birth experience due to Sunday-specific understaffing.[27]

For conditional random instrument assignment, Figure A.1 explores unconditional correlations of instruments and maternal characteristics. Many well-known patterns emerge,

[26] Studying membrane ruptures at $\geqslant$37 weeks of gestation implies German state-mandated maternity protection has been mitigating differential impacts of daily stress at work for about eight weeks already.

[27] Anasthesia like epidurals correlate with stalled labor, emergency c-sections, and severe perineal tearing (Tammaa et al., 2007).

e.g., the intuitive overlap between instruments. Moreover, obstetricians' intervention preferences (and staffing constraints) relate to a mother's age, her bmi, whether she has a her own midwife etc., all of which strongly motivates the inclusion of core controls. Nevertheless, to put the correlations into perspective, obstetricians' preferences and staff capacity indicators are much less specific to maternal and hospital strata than the share of interventions themselves (see Table A.4 and Table A.5). All in all, besides controlling for the observable differences, placebo tests are warranted to assess the presence of unobservable differences possibly introducing endogeneity into the framework.

3.3.2 Instrument Relevance & First-Stage Results

Each set of instruments gives rise to the following system of first-stage equations.

$$\underset{t\times 1}{\mathbf{t}_m} = \underset{t\times z}{\mathbf{\Gamma}} \underset{z\times 1}{\mathbf{z}_m} + \underset{t\times k}{\mathbf{\Phi}} \underset{k\times 1}{\mathbf{x}_m} + \underset{t\times s}{\mathbf{\Lambda}} \underset{s\times 1}{\mathbf{1}} + \underset{t\times 1}{\epsilon_m} \quad (2)$$

Notation builds on Equation 1. There are $1, ..., t$ treatments, $1, ..., k-1$ covariates, and $1, ..., s$ sets of fixed effects observed for mother m, while $1, ..., z$ instruments are defined as

$$\text{either } \mathbf{z}_m \equiv \begin{bmatrix} InducedLaborPref\ (ILP_m) \\ CSectionPref\ (CSP_m) \\ VaginalOperPref\ (VOP_m) \\ ILP_m * CSP_m * VOP_m \end{bmatrix}, \text{ or } \mathbf{z}_m \equiv \begin{bmatrix} DueDateNoWorkday\ (DN_m) \\ MembraneRuptureNight\ (RN_m) \\ MidwifeShortage\ (MS_m) \\ DN_m * RN_m * MS_m \end{bmatrix}$$

- $DN_m \in \{0,1\}$: =1 if the due date is a weekend day or public holiday, 0 else
- $RN_m \in \{0,1\}$: =1 if mother m has a pre-labor membrane rupture between 8 pm to 4 am, 0 else
- $MS_m \in [0,1] = \begin{cases} 0 \text{ if \#current deliveries at that hospital} = 0 \\ \frac{\text{\#current deliveries at that hospital without a midwife}}{\text{\#current deliveries at that hospital}} \text{ else} \end{cases}$
- ILP_m, CSP_m, $VOP_m \in [0,1]$: mean prior rate of inductions, non-emergency c-sections, and vaginal operations of obstetrician treating mother m

Multi-treatment First-stage Results

As classical weak instruments statistics are not applicable in this setting, underidentification is tested instead.[28] Obstetricians' intervention preferences identify Equation 2, i.e., underidentifi-

[28]Intuitively, given multiple endogenous variables, the standard first-stage F-statistic could fail as follows. Assuming a just-idenitfied model, in which one instrument is predictive of several endogenous variables, while another instrument is barely predictive for any of them. Then, in both first-stages, the F-statistic would be high, even though one of the endogenous variables would be only weakly identified. Sanderson and Windmeijer (2016) provide a cluster-robust underidentification test for multiple endogenous treatments by running the Sargan-Hansen J-Test for overidentification (Cameron and Miller, 2015) in auxiliary regressions. Conventional extensions of weak instrument statistics handle either multiple endogenous treatments, e.g., the Anderson-Rubin test (Chernozhukov et al., 2009), or cluster-robust standard errors, e.g., the Montiel-Olea-Pflueger F-statistic (Andrews and Stock, 2018) but not both.

cation is rejected by a p-value of <0.001 for all treatments (Table A.6).[29] Adding to this, more intuitively than statistically, the first stage of the combined intervention *Induction + surgery* (shown in the lower panel of Table 2) confirms strongly significant positive correlations with all preference-based instruments.[30]

Single-treatment First-stage Results

Treating either induction or (non-emergency) c-section as the only treatment all preference-based instruments are relevant for both interventions, only induction preference is irrelevant for c-section (Table 2). As expected, induction (c-section) preference predicts induction (c-section) most strongly. Despite likely omitted variable bias from left-out rivaling interventions, the estimates are quite in line with intuition, e.g., they show complementarities between induction and surgical intervention preferences, as well as substitution effects between c-section and vaginal operation preferences.

Among staff capacity-based intruments, only those referring to night shift constraints predict induction (albeit with opposing signs). This is in line with the co-movement depicted in Figure 7.[31] All instruments but the due date's non-working day status, which is never relevant anyways[32], strongly predict c-section. A significant positive correlation with midwife shortages is a common finding in the literature. BZgA (2005) and Jacobson (1993) attribute this to midwives being often more patient and more proficient in conservative obstetric skills than obstetricians. Usually, hospitals have their own midwives and/ or tolerate so-called in-patient midwives to be brought along by the mothers. However, due to severe midwife shortages, it becomes increasingly difficult to find an in-patient midwife during pregnancy (Bruns, 2017). Figure 8 visualizes the trend over time in midwife shortages plotting the shares of midwife types over time.

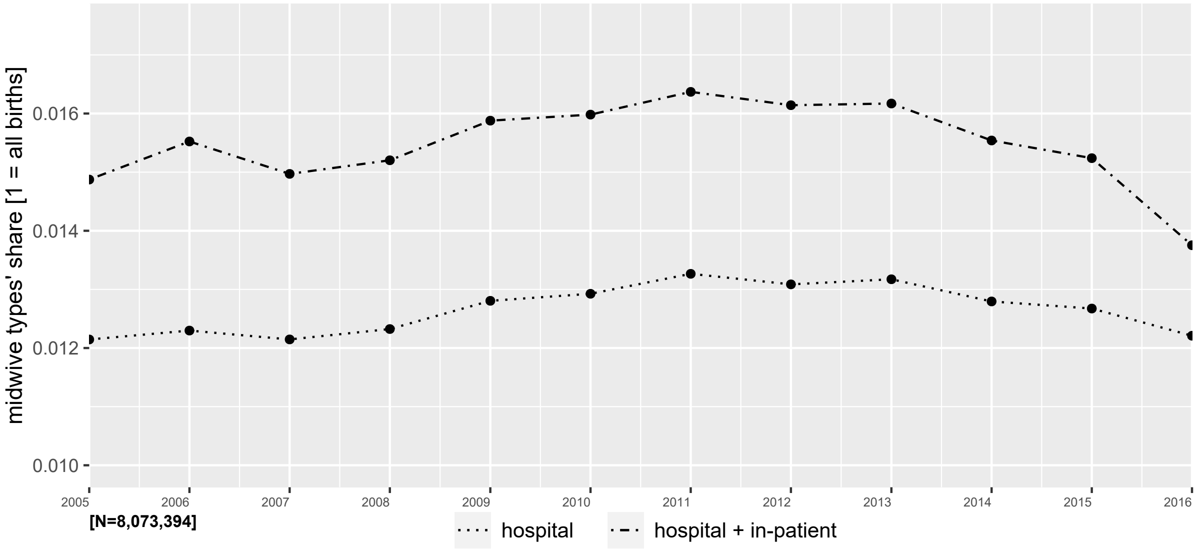
Added-variable plots in the appendix (see Figure A.9, Figure A.10, and Figure A.11) explore the residual correlation (reflected in the slope of the regression line) between a given instrument and all three main intervention types netting out maternal core characteristics. When employed

[29]By contrast, for model Equation 3 featuring all intervention interactions underidentification is never rejected, neither for the main nor the interacted intervention treatments (Table A.8).Staff capacity-based instruments even fail to identify the parsimonious multi-treatment model (Equation 1).

[30]The other three first-stages - referring to each main intervention at a time - are likewise very strong and available upon request. However, the decisive criterion is the underidentfication value, not the single nor the joint significance of the first stage estimates.

[31]Absent other complications, a pre-labor membrane rupture does not imply maternal or fetal compromise. Therefore, medical guidelines state to monitor the mother closely for at least 12 hours before inducing labor (Mylonas and Friese, 2015; DGGG, 2006). However, the risk of infection increases while waiting for labor to start, and fear for the fetus' well-being probably drives down induction refusals.
Using instruments involving pre-labor membrane ruptures raises concerns about impacting a rather limited share of observations at all. 30% of zero-precondition first-time mothers experience a membrane rupture (which is way above the overall mean of 8% given by IQTIG (2017)), out of which 15% occur during the night shift (8 pm till 04 am). At the same time, given an average yearly induction rate of 28% in the sample period (Schwarz et al., 2016), many mothers with zero instrument status do have their labor induced. This warrants subsample regressions exclusively for mothers with pre-labor membrane ruptures.

[32]Intuitively, the low precision of the due date counteracts instrument relevance. Still, as the single best predictor of the natural birth date, the due date represents a hospital's benchmark for treatment decisions.



Source: Destatis data downloaded from `https://www.statistischebibliothek.de` for the years 2005 through 2016. Own calculations.

in a single-treatment approach, all but one instruments capture birth intervention dynamics.

Table 2: First-Stage Effects of Non-Medically Indicated Birth Interventions

	Staff capacity shortages					Obstetricians' preferences for			
		non-working day status of		during night shift					
Instrument	midwife shortage upon admission	predicted due date	actual weekday of birth	prelabor membrane rupture	arrival at hospital	non-emergency c-section	induced labor	vaginal operations	ind. labor x c-section x vaginal oper.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Induced labor	-0.0003	-0.0022	0.0032	0.0574***	-0.2195***	0.1080***	0.2692***	0.0397**	
	(0.0029)	(0.0022)	(0.0022)	(0.0035)	(0.0028)	(0.0085)	(0.0178)	(0.0179)	
Non-emerg. c-section	0.0493***	-0.0015	-0.0397***	-0.0400***	-0.1014***	0.7308***	0.0439	-0.0675**	
	(0.0034)	(0.0022)	(0.0024)	(0.0028)	(0.0022)	(0.0089)	(0.0305)	(0.0304)	
Induction + surgery						0.2521***	0.1285***	0.1520***	1.2266***
						(0.0090)	(0.0151)	(0.0147)	(0.1693)
N	177,215	177,215	177,215	177,215	177,215	66,916	66,916	66,916	66,916

Notes: *p<0.1; **p<0.05; ***p<0.01. Linear models based on IQTIG birth records for Germany 2015-2016 restricted to the sample of zero-precondition first births (see notes to Table A.11). Each row represents a birth intervention treatment to be instrumented by one (upper panel) or multiple (lower panel) instruments at a time.

4 Maternal & Neonatal Health Effects

This section discusses how non-medically indicated induction (possibly augmented by surgical intervention) impacts patients' immediate health. First, novel multi-treatment IV estimates are presented, then single-treatment equivalents relate to the literature. Finally, a channel-based outlook sketches possible longer-run impacts.

4.1 Multi-treatment IV results

Each column of Table 3 is dedicated to a distinct outcome. The upper four rows belong to the same specification (Equation 1) showing the jointly estimated impacts of all instrumented treatments, one below the other, conditional on core controls.[33] The last rows show induction estimates from separate regressions, i.e., row (5) excludes controls, and row (6) excludes the combined intervention *Induction + surgery*.[34]

Maternal Health

Following column (1) perineal laceration incidence is 6% more likely for induced (and possibly invasive) than unassisted delivery.[35] The estimated beneficial effect of induction alone is counteracted by inductions entailing a more violent course of labor and additional intervention.[36] The impact of induction alone is stable to leaving out core controls. Modeling main intervention effects only, no significant effect of induced labor (in any combination vs. unassisted delivery) emerges. Thus, disentangling the main and combined effects seems key for identification. In the literature, induction-*caused* perineal damage lacks explicit reporting (see, e.g., Jürges, 2017)

[33]Strictly speaking, all main effects are identified relative to unassisted delivery and deliveries suffering two-fold surgery, i.e., vaginal operations followed by c-section which occurred in the whole sample only 45 times. Assume discarding rare treatment combinations was irrelevant in terms of omitted variable bias and multicollinearity. Then, the *main* interventions' interpretation should be stable to using *Induction + surgery* instead of two- and threefold interactions.
Originally, it was intended to estimate a model using three main birth intervention treatments and all their possible interactions (Equation 3). However, the corresponding results indicated multicollinearity issues and underidentification. The IV (main and subsample) estimates are reported for completeness acknowledging that no causal insights arise from this evidence (Table A.11, Table A.12. The coefficients estimated using very weak instruments are likely more biased than their OLS analogs and inflated standard errors might render significant relationships insignificant (Angrist and Pischke, 2009).).

[34]The last row's identified effects differ, e.g., the impact of induced labor (possibly supplemented by surgical intervention) is compared to any birth mode not involving induction.

[35]Given the distribution of inductions alone vs. augmented by surgery (Table A.8), the overall effect of induction in our sample can be computed as (+0.60) ppt * 0.17 = 0.102 ppt net off (-0.36) ppt * 0.28 = (-0.1008) ppt yielding (+0.0012) ppt (residual increase), relative to a sample mean of 0.02 ppt.

[36]The (only existing and imperfect) OLS benchmark (Table A.11, column (2)) is based on a distinct model (Equation 3). and predicts a rise in perineal tearing around 0.003 ppt (15% relative to the sample mean) after induction without surgery. Notably, in this study, even the interpretation of IV and OLS effects from the same model differs. A sample of zero-precondition mothers includes inductions with debatable medical advantages, which could explain relatively more positive health effects estimated by OLS. Figure A.7 and Figure A.8 show the frequency of intervention indications by medical severity, which seem stable across weekdays but responsive to hours of the day.

preventing a direct comparison at this stage.[37] While older medical guidelines do list inductions among risk factors of severe tearing (Tammaa et al., 2007), more recent ones refer to an evidence gap about its impact (DGGG, 2020c).

Neonatal Health

Column (6) of Table 3 depicts a strong and stable negative intervention impact on the APGAR score five minutes (or ten minutes likewise - not shown -) post-birth, first and foremost due to induced labor (-2.2 points or 23%) but also in response to a non-emergency c-section (-0.92 points). As recapped in Table A.1, seminal empirical evidence is concentrated on immediate neonatal health outcomes. Abstracting from limited comparability with single-treatment models[38], the new evidence contradicts findings by Jürges and Köberlein (2015) or Jacobson et al. (2020) and exceeds the tiny negative induction impacts suggested by Lynch et al. (2019) or Schulkind and Shapiro (2014b). Besides, the new findings speak against a positive c-section effect on the APGAR score of about 0.5 points established by Card et al. (2018) while lining up perfectly with the decrease estimated by Costa-Ramón et al. (2018). Thus, this study adds large-scale quasi-experimental evidence on adverse health effects to a highly unreconciled evidence base.

Placebo Effects

To uncover potential instrument endogeneity, this model is regressed on a battery of placebo outcomes, none of which yields significant estimates. Column (5) of Table 3 shows that all coefficients associated with the placebo outcome *Prenatal care starting >12th week* are insignificant. The rationale is that interventions happening at the delivery may not *lead to* events earlier throughout the pregnancy. Other placebo candidates tested are a fetus' 1) sex, and 2) innate disability, as well as a mother's 3) alcohol or cigarette abuse during pregnancy, 4) employment status, and 5) psychological or social problems.

[37] To simplify interpretation, the (relative) impact of a (non-emergency) c-section on perineal lacerations - though interpretable through a potential outcomes framework as used by Card et al. (2018) based on Abadie and Kennedy (2003) - is not discussed.

[38] Simultaneous identification within the same framework puts intervention impacts naturally into perspective to each other. By contrast, next to internal validity problems (most prominently, omitted variable bias from left-out rivaling interventions) comparing single-treatment estimates across different studies hinges on each study's external validity.

Table 3: Health & Capacity Multi-Treatment IV Effects of Non-Medically Indicated Birth Interventions

	patient health	hospital staff capacity			placebo	literature link
Dependent variable	perineal tearing (III/IV)	labor duration (# hours)	postnatal hospital stay (# days)		1st prenatal care >12th week	APGAR score (10 min.)
			mother	neonate		
	(1)	(2)	(3)	(4)	(5)	(6)
Induced labor	-0.3632*	-27.4370**	-0.7281	-0.3623	0.0939	-2.1522*
	(0.2035)	(11.6755)	(1.7541)	(2.1989)	(0.2621)	(1.2566)
Non-emergency c-section	-0.1512*	-17.0140***	1.8026**	1.7198*	0.0366	-0.9223*
	(0.0854)	(4.8519)	(0.7406)	(0.9354)	(0.1074)	(0.4928)
Vaginally operative procedures	-0.1391*	-10.0060**	1.2411*	1.5796*	0.0659	-0.7214
	(0.0836)	(4.6532)	(0.7103)	(0.8841)	(0.1043)	(0.4902)
Induced labor + surgery	0.5998**	40.0837***	-0.6750	-1.5735	-0.2079	2.3148
	(0.3044)	(17.2384)	(2.6272)	(3.2927)	(0.3837)	(1.7955)
Induced labor (no controls)	-0.3883*	-28.0768**	-0.5209	-0.2269	0.4036	-2.1525*
	(0.2063)	(11.7495)	(1.7291)	(2.1598)	(0.3017)	(1.2394)
Induced labor (main effects only)	0.0223	-1.6727	-1.1627***	-1.3736***	-0.0397	-0.6639***
	(0.0207)	(1.0939)	(0.2981)	(0.3351)	(0.0394)	(0.1878)
Mean (dependent variable)	0.02	6.8	3.4	3.2	0.077	9.7

Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births (N=177,215). Instrumenting by intervention preferences creates a subsample of births with non-missing obstetrician id (N=66,916). Sample and variable creation detailed in Table A.2 and Table A.3. The model builds on Equation 1, thereby using four endogenous treatments simultaneously - the main intervention indicators plus a binary indicator that is one for any induction followed by either c-section or vaginal operations or both (*Induced labor + surgery*). Jointly instrumented through obstetricians' intervention preferences and their triple interaction, the estimates of all four treatments are reported in rows (1)-(4). Besides, for induced labor, coefficients estimated by two other regressions (without controls in row (5), and excluding *Induced labor + surgery* in row (6)) are reported. Apart from row (5), regressions include as core controls the year of delivery, a mother's age, her region of origin (7 categories), her socioeconomic status (6 categories), and her single status (yes/no), where categorical variables enter as sets of binary indicators. Moreover, continuous measures are created for maternal height (as cubic), maternal weight at the beginning of the pregnancy (as cubic), and maternal BMI. Each column corresponds to a distinct outcome. Robust standard errors clustered by 3-digit zip codes of maternal residence. Means are available for the full sample of zero-precondition first-births.

4.2 Single-treatment IV results

Table 4 reports single-treatment IV estimates (conditional on core controls) for all instruments newly proposed in this study as well as some state-of-the-art instruments from the literature.[39]

Maternal Health

Using obstetricians' preferences-based instruments, having an induced (and potentially surgical) birth is estimated to increase severe perineal tearing incidence between 0.04 to 0.33 ppt relative to any birth mode not involving induction, thereby comprising not only unassisted births but even pre-labor c-sections. In line with intuition, the multi-treatment estimate (using the same instruments jointly and measuring the impact of induced vs. unassisted delivery, both of which are challenging to the perineum) is much smaller. While the impact of c-sections is of secondary interest in the context of high-degree tearing, the much smaller range (-0.05 to 0.02 ppt) somewhat puts the estimates' stability into perspective. Finally, using staff capacity-related instruments or a simple OLS model, no significant impacts are found.

Neonatal Health

Using three of five instruments at a time, induced labor predicts a significant decrease in the APGAR score (0.14-2.39 points) compared to not inducing labor. Larger impacts result from obstetricians' preferences-based instruments, the upper bound of which lines up with the preferred estimate of Table 3. The OLS benchmark, half the size of the IV lower bound, is significantly negative, too. When predicting the APGAR score by c-section as the only treatment the pattern is less stable, i.e., two staff capacity instruments predict a positive impact (up to 0.33 points), one capacity and one preference-based instrument suggest negative impacts of the same size, and finally, one capacity and one preference-based instrument fail to detect significant impacts at all. The naive OLS model suggests an impact overall similar to that of induction.

Placebo Tests

Interventions are interrelated, thereby responding (more or less strongly) to the same instruments. This easily turns the left-out interventions (among other candidates) into omitted variables (or bad controls, see, e.g., Costa-Ramón et al. (2018) controlling for (and stratifying by) induction upon targeting c-section effects) biasing the single-intervention IV model. Depending on which placebo outcome is chosen different drivers of endogeneity can be detected.

[39] From Table 2, we know that out of six proposed instruments, five identify either induction or c-section and four identify either one at a time.

Table 4: Health & Capacity Single-Treatment Effects of Non-Medically Indicated Induction vs. C-section

	maternal health	hospital staff capacity			placebo	literature link	
Dependent variable	perineal tearing (III/IV)	labor duration (# hours)	postnatal hospital stay (# days)		1st prenatal care >12th week	APGAR score (10 mins.)	
			mother	neonate			
	(1)	(2)	(3)	(4)	(5)	(6)	N
Induced labor (not instrumented)	0.0008	-1.0590**	0.1849***	0.1268***	0.0015	-0.0737***	177,215
	(0.0009)	(0.0449)	(0.0091)	(0.0123)	(0.0014)	(0.0047)	
Nightly prelabor membrane rupture	-0.0152	-0.4690	-0.6826***	-0.6857***	-0.0737**	0.0477	177,215
	(0.0180)	(0.7056)	(0.1800)	(0.2222)	(0.0307)	(0.0881)	
Arrival at hospital during night shift	-0.0045	-4.5616***	1.0108***	0.8340***	0.0183***	-0.1393***	177,215
	(0.0037)	(0.1487)	(0.0415)	(0.0476)	(0.0060)	(0.0175)	
Obstet.'s preference vaginal operations	0.3332*	20.0184	11.6739**	13.4012**	0.1361	-1.6002	66,916
	(0.1933)	(13.4764)	(5.6347)	(6.4334)	(0.2568)	(1.1091)	
Obstet.'s preference for c-section	0.1564***	-36.9029***	9.6169***	7.3097***	-0.1771***	-2.3933***	66,916
	(0.0269)	(3.0332)	(0.8717)	(0.7897)	(0.0453)	(0.2520)	
Obstet.'s preference for inductions	0.0361*	-1.6921	-0.4765*	-0.6667**	-0.0366	-0.7348***	66,916
	(0.0203)	(1.2240)	(0.2868)	(0.2981)	(0.0364)	(0.1799)	
Non-emergency c-section (not instrumented)	-0.0333***	-4.9841***	1.2523***	1.0687***	-0.0025*	-0.0946***	177,215
	(0.0008)	(0.0897)	(0.0153)	(0.0181)	(0.0015)	(0.0061)	
Nightly prelabor membrane rupture	0.0218	0.6731	0.9797***	0.9841***	0.1058**	-0.0685	177,215
	(0.0259)	(1.0245)	(0.2361)	(0.2993)	(0.0444)	(0.1260)	
Arrival at hospital during night shift	-0.0097	-9.8733***	2.1879***	1.8051***	0.0396***	-0.3010***	177,215
	(0.0079)	(0.3050)	(0.0864)	(0.1017)	(0.0129)	(0.0383)	
Obstet.'s preference vaginal operations	-0.1959*	-11.7677**	-6.8624*	-7.8778*	-0.0800	0.9553	66,916
	(0.1134)	(5.0723)	(3.8674)	(4.1733)	(0.1418)	(0.8008)	
Obstet.'s preference for c-section	0.0231***	-5.4556***	1.4217***	1.0806***	-0.0262***	-0.3536***	66,916
	(0.0034)	(0.2148)	(0.0535)	(0.0697)	(0.0065)	(0.0298)	
Midwife shortage upon arrival	-0.0489***	-20.8249***	4.5028***	4.6815***	-0.0152	0.3324***	177,215
	(0.0157)	(1.1067)	(0.3331)	(0.3758)	(0.0298)	(0.1132)	
Weekday of delivery non-working day	-0.0190	-10.7459***	1.1859***	1.3024***	-0.0229	0.2304**	177,215
	(0.0196)	(0.7736)	(0.1834)	(0.2386)	(0.0332)	(0.1017)	
Mean (dependent variable)	0.02	6.8	3.4	3.2	0.077	9.7	

Notes: *p<0.1; **p<0.05; ***p<0.01. models based on IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births (N=177,215). Instrumenting by

Therefore, it is not surprising that instruments yield significant impacts of induction (or c-section) on belated prenatal care (and other placebo outcomes). Taking this evidence and the reasons laid out in section 3 together, single-treatment IV models should be interpreted with caution.

4.3 Channels of Longer-run Induction Impacts

Severe Perineal Tearing & Future Surgery

A growing body of observational literature raises concerns about *Perineal tearing (III/IV)* impairing maternal future health and fertility outcomes. Especially debated is the strong association with future surgery, most prominently due to 1) recurrence of tearing (Priddis et al., 2013; Woolner et al., 2019), 2) subsequent c-section on request (O'Donovan and O'Donovan, 2018; Størksen et al., 2015; Ryding et al., 2016; Smarandache et al., 2016; Garthus-Niegel et al., 2014; Pang et al., 2008; Tschudin et al., 2009; Woolner et al., 2019), as well as (3) avoidance of future pregnancies (Priddis et al., 2013) or even infertility (Jolly et al., 1999; Gottvall and Waldenström, 2002), which is especially critical as severe tearing is centered on first-time mothers (DGGG, 2020c). However, due to the lack of quasi-experimental multi-treatment evidence, this channel forbids causal longer-run inference of induction impacts.

Induction & Subsequent C-section: A Thought Experiment

But to which extent are inductions burdening the health care system by causing c-sections?[40] For a back-of-the-envelope quantification within the scope of this study, let's assume the share of mothers with induction-debited c-sections to be 7%.[41] On the one hand, induced and unassisted labor can fail alike, which breaks the direct link between c-section as an outcome of labor trials. On the other hand, the calculation is conservative in accounting only for immediately provoked c-sections, thereby excluding higher-order parity c-sections (maybe resections[42]) requested due

[40]Rummel (2007) computes hospital profits (based on DRG cost-rates among n=100 mostly uncomplicated births) for a c-section (3,843 EUR [reimbursement] - 2,385 EUR [reimbursement-relevant hospital costs] = 1,458 EUR) vs. an unassisted vaginal birth (1,847 EUR - 1,674 EUR = 173 EUR). Thus, c-section-specific additional reimbursement amounts to 3,843 EUR - 1,847 EUR = 1,996 EUR.
Using register data of *BARMER GEK*, a major German public health fund in 2010, Kolip et al. (2012) find average reimbursement costs of 1,520 EUR vs. 2,680 EUR for vaginal and c-section delivery respectively implying 1,160 EUR additional reimbursement for c-section. Despite their differences, the studies agree on a sizable extra financial burden of c-sections for the health insurance system.

[41]In our sample of healthy mothers, 79% of inductions happen on non-medical grounds (Figure A.7 and Figure A.8). The share of elective inductions is derived from the composition of indications (i.e., summing up explicit maternal requests and incompletely specified risks) reported for inductions in our sample. While indications may be strategically coded (Jürges and Köberlein, 2015), intuitively, hospitals should rather understate elective interventions implying a lower-bound cost estimate. For simplicity, imposing the share observed for induced on all assisted zero-precondition first-births, 79% * 9% (the share of c-sections immediately following induction) = 7% of healthy mothers experience both, induction and c-section without a medical indication ex-ante. Per 1000-women, drawing on Rummel (2007), we get 1,000 * 7% * 1,400 EUR (hospital profit per c-section) = 99,500 EUR vs. 1,000 * 7% * 2,000 EUR (additional health care burden per c-section) = 140,000 EUR.

[42]With a share of 23.6% in 2010 already (Kolip et al., 2012), resection has been the most popular indication for c-section for more than a decade.

to traumatic past induction. Expressed per 1000-women, this yields 99,500 EUR profits for the hospital while burdening the health care system with 140,000 EUR. As of 2019, across healthy first-time mothers, this implies 34 million hospital profits vs. 47.8 million losses for the public health care system.[43]

5 Hospital Staff Capacity Effects

This section discusses the impact of non-medically indicated induction and/ or surgical intervention on a hospital's staff capacities. Adopting a structure similar to section 4, the focus lies on new causal evidence from a parsimonious multi-treatment IV model, which is put into perspective by seminal single-treatment models from the literature. Some back-of-the-envelope calculations assessing the system-wide health care impact conclude.

5.1 Multi-treatment IV Results

Do the adverse intervention effects established in section 4 rebound from patients' impaired health onto hospitals' staff capacity constraints? Columns (2) to (4) of Table 3 based on Equation 1 quantify staff capacity impacts via two key measures of a hospital's monitoring workload.[44]

Labor Duration

For inductions (possibly augmented by surgery) average labor shortens by close to 1 hour.[45] The pure induction effect is stable to leaving out core controls. Using main effects alone, no significant effect of induction emerges. Surgical interventions mechanically cut labor short by 10 (vaginal operations) to 17 (c-section) hours. All in all, interventions produce overarching favorable effects for staff capacity absorbed by *Labor Duration*.[46]

Postnatal Hospital Stay

Neither sole nor combined induction impact the postnatal hospital stay significantly. Notably, the accompanying standard errors are unusually large. Not including controls, standard errors are slightly smaller and the estimated coefficients decrease (in absolute terms) by around one third. The main-effects-only model yields highly significant negative impacts that shorten the

[43] Kolip et al. (2012) suggest 5% of first-time mothers suffer preconditions. Using 2019 as base year, this equals 341,000 zero-precondition first births among 359,000 first-time mothers from Germany. The actual sample size used in this study deviates due to a different sample period, a (stricter) definition of zero-preconditions, and some missings in non-mandatory maternal background information.

[44] Results of the originally intended model (Equation 3) are not discussed.

[45] Drawing on the intervention shares among zero-precondition first-time mothers (Table A.8), we find some 0.28 * (-27.4) h = -7.7 hours shorter labor due to induction alone vs. a (0.09 + 0.08) * (+40.1) h = 6,8 hours prolongation, which yields (-0.9) hours (-13%) of foregone labor experienced by a representative mother.

[46] The health impact of labor length is ambiguous. On the one hand, longer labor causes longer pain and exhaustion. On the other hand, shorter labor might come at the cost of severe tearing or even phenomena like precipitate deliveries. Viewing shorter labor in the light of worse tearing (section 4) speaks of a hastened birth experience and disadvantage for the patient.

maternal (neonatal) hospital stay by 1.2 days or 35.3% (1.4 days) following induction in any combination relative to any not induced birth.

The hospital stay of mothers and neonates is prolonged after births involving c-section (ca. 1.8 days for both) or vaginal operations (1.2 and 1.6 days). Thus, surgical interventions alone drive adverse health effects mirrored in additional patient monitoring. This might explain (part of) the marginalization of inductions relative to surgical birth interventions reflected in inductions' less restricted usage on non-medical grounds.

5.2 Single-treatment IV results

Regressing hospital staff capacity outcomes on induction modeled as single treatments (Equation 3) yields mostly significant but volatile estimates across instruments.

Labor Duration

According to column (2) of Table 4 induction (relative to any other birth mode) is estimated to significantly shorten labor duration between -36.9 to -4.6 hours. The wide range relies on just two (out of five) instruments. Upon instrumenting by obstetricians' preferences precision is a problem. Especially for vaginal operations' preference standard errors explode across all staff capacity outcomes. The corresponding OLS estimate predicts a significant but relatively modest decrease in labor length by 1 hour, well aligned with the (conceptionally different) total impact the multi-treatment specification Equation 1 yields.

For all but one instrument, non-emergency c-section is estimated to significantly shorten labor between -20.8 to -5.5 hours. Centered around 10 hours, the range encompasses the -17 hour decrease found by the multi-treatment estimate for c-section (not mixed with trial of labor and compared to unassisted birth). Finally, the OLS estimate is close to the IV lower bound estimate (5 hours).

Postnatal Hospital Stay

The estimated impact of induction on patients' postnatal hospital stay is always significant. The predicted positive impacts (accompanied by inflated standard errors) reach up to 13 extra days in the hospital. Depending on the instrument some much smaller negative effects emerge. The corresponding OLS estimates are negligible in size and the multi-treatment model (Table 3) finds no significant induction-related effects at all.

For c-sections, the most precise single-treatment estimates suggest 1 - 5 additional days in the hospital encompassing the multi-treatment estimate of 1.8 days shown in Table 3. The only negative (and very imprecise) coefficients emerge when instrumenting by vaginal operations' preference (-7 days). OLS estimates are consistently positive and close to 1 additional day in the hospital for both, mothers and neonates.

5.3 Back-of-the-Envelope Calculation: Compound Costs for Intervention-Related Monitoring

The DGGG (2020a) states that elective inductions' financial impact on the health sector has not yet been established. As for (elective) c-sections, Feige (2008) mentions 100 million EUR annual reimbursement burden. Based on Table 3, this section sketches German hospitals' system-wide *staffing capacity burden* originating from the major birth interventions performed to relieve their *staffing constraints.*

Labor Duration

Per 1000-women, non-medically indicated interventions forego ca. 8,500 hours of labor or >1.6 million staffing costs.[47] Considering 341,000 healthy first-time mothers in Germany as of 2019 (subsection 4.3), foregone labor saves 547 million EUR staffing costs, a hazardous misalignment of maternal and hospital interests.

Postnatal Stay

Per 1000-women-and-neonates, additional costs for a prolonged stay of ca. 1,800 days implies a financial burden of 551,000 EUR. For all zero precondition first-time deliveries, this amounts to 188 million EUR.[48]

Adverse Health Effects & Implicit Staffing Burden

Focusing on a hospital's short-run capacity costs as reflected in the main outcomes of this study yields a computation conservative in several ways. First, using mean cost rates associated with uncomplicated births abstracts from potential adverse health outcomes requiring not just prolonged but also *more intense* monitoring like, e.g., neonatal ventilation. Second, explicitly ignoring many other cost types is bound to underestimate the full costs by far.

The German guideline system proposes hospital care procedure workflows, thereby mapping staffing obligations to adverse health conditions. Among healthy first-births, the mean APGAR score is 9.7 (sd 0.76) and the estimated decrease due to induction is 2.15. A score of <8 warrants additional testing already (GNPI, 2022). Still conservative, we assume two additional basic tests

[47] 1000 * (28% * (-27.44) h (induction) + 26% * (-17.01) h (c-section) + 32% * (-10.00) h (vaginal operations) + 17% * (+40.08) h (induction-plus-surgery)) = (-8492.2) hours. Following Rummel (2007), this implies (2/3[h] * 50 EUR + 1[h] * 40 EUR) = 622,761 EUR reimbursable costs => -2,075,871 EUR total monitoring costs [100%, incl. non-reimbursable] drawing on Bruns (2017), 1.6 M (79%) of which are presumably non-medically indicated (subsection 4.3).

[48] Per 1000-women, 1000 * (26% * +1.80 days (c-section) + 32% * +1.24 days (vaginal operations)) = +864.8 days. Per 1000-neonates, 1000 * (26% * (+1.72) days (c-section) + 32% * (+1.58) days (vaginal operations)) = (+952.8) days. For both jointly, (+1817.6) days * (30 EUR + 1/2h * 50 EUR) = 99,968 reimbursable costs [=30%] => 333,227 EUR [100%] total costs (Bruns, 2017). Proxying accommodation base costs of 200 EUR/day = 363,520 EUR, total staffing + accommodation costs = 696,747 EUR, out of which 551 k EUR 79% (see subsection 4.3) arise from non-medically indicated interventions.

performed per mother's labor induced on non-medical grounds. Then, among 341,000 healthy first-time mothers testing due to non-medically indicated induction entails total hospital costs of 11.8 million EUR.[49] Moreover, if the results of these routine tests confirmed neonatal adaptation anomalies more involved testing and care procedures would follow (GNPI, 2022). Modest in absolute values, this exemplary induction impact channeled by *two routine tests* is already close to 12% of the annual burden attributed to avoidable c-section procedures as a whole. As soon as intensive care measures come into play, costs rise astronomically (Almond et al., 2005).

Taken together, the naive computations have shown 1) that (weakly) favorable induction effects on seminal hospital capacity measures do not rule out substantial negative impacts on staffing costs working through more subtle channels. Besides we learn, 2) how rapidly a hospital's costs diverge from the cost it is compensated for. This in turn incentivizes more intervention without medical reason fueling a snow-balling effect that, in the long run, suggests adverse impacts on hospitals and patients alike.

6 Discussion

Non-medically indicated labor induction is a viral topic around the world. This chapter provides novel evidence for the role of inductions performed to alleviate staff capacity constraints in German hospitals. The estimations shown here are based on a new identification approach that uses exogenous variation in obstetricians' intervention preferences, ruling out key concerns of endogeneity through a battery of placebo tests. The main results document that induced vs. unassisted labor 1) provokes severe birth canal lacerations and lower APGAR scores, which rebound on staff capacity via 2) additional examinations and monitoring.

A framework incorporating the endogenous and interrelated nature of the three major birth interventions is pioneer work in the field. To begin with, interactions isolate successful and failed inductions entailing surgical intervention. Next, regarding the marginalization of induction relative to surgical interventions, simultaneous impact identification makes it trivial to compare the effects to each other. Last, methodologically cleaner than prior literature, the framework benchmarks evidence from single-treatment identification.

Tentatively sketching *some* likely follow-up costs for the public health care system touches upon the unresolved link between inductions and subsequent (c-section) surgery. Apart from the overall health impact explored here, ongoing work by Gerhardts (2024) focuses on heterogeneous impacts across different types of mothers. If induction-related lower APGAR scores were centered on mothers with, e.g., lower socioeconomic status, this would impair neonatal health

[49] Applying current cost rates (KBV, 2020) per 1000-women, we get 1000 * 28% (induction rate) * 79% (non-medically indicated) * (17 EUR for pulse oximetry + 30 EUR for an electrocardiogram)= 10,4 k EUR [30%] => 34.6 k EUR [100% incl. non-reimbursable monitoring (Bruns, 2017)] total hospital costs for testing due to inductions performed on non-medical grounds.

and cognitive development disproportionately. Furthermore, examining intervention effects at low-quality versus small hospitals will shed light on a disputed reason for centralizing maternity care.

Finally, discussing the (adverse) health effects of birth interventions in the light of the professional ethical ideal stated in the very first citation, goes beyond the scope of an economics study.

A Appendix

Table A.1: Quasi-/Experimental Evidence on Non-medically Indicated Induction

Outcome	**Study authors**	**Design**	**Health Impact** positive (+)/ neutral (=)/ negative (-)
Maternal			
Labor progress	Buckles et al (2017)	IV elective delivery policy (n=410,459)	(−) precipitious labor more likely (4.6 x) [< 39 weeks]
	Jacobson et al (2020)	Reduc.Form holiday effect (n=4,599)	(=) labor complications
Blood loss	Saccone et al (2015)	SRMA of 5 ARRIVE trials (n=844)	(+) -58 mL
Surgical delivery: C-section (CS) & vaginal operations	Buckles et al (2017)	IV elective delivery policy (n=410,459)	(=) CS [< 39 weeks]
	Jürges (2018)	DID parental leave policy (n=565,000)	(=) emergency CS
	Alfirevic et al (2009)	SRMA of 61 ARRIVE trials (n=12,819)	(+/−) less failed labor (8.4%:53.8%)/ more epidurals
	Wood et al (2014)	SRMA of 31 ARRIVE trials (n=12,166)	(+) fewer CS [w/o membrane rupture]
	Saccone et al (2015)	SRMA of 5 ARRIVE trials (n=844)	(=) CS
	Mishanina et al (2014)	SRMA of 157 ARRIVE trials (31,085)	(+) CS less likely (-12%) [≥ 39 weeks]
	Sanchez et al (2003)	SRMA of 16 ARRIVE trials (n=6,588)	(+) fewer CS (20.1%:22.0%) [at 41 weeks]
	Middleton et al (2018)	SRMA of 27 ARRIVE trials (n=11,738)	(+/−) fewer CS/ more vaginal operations [≥ 39 weeks]
	Gülmezoglu et al (2012)	SRMA of 21 ARRIVE trials (n=8,749)	(+) fewer CS [≥ 39 weeks]
	Caughey et al (2009)	SRMA of 9 ARRIVE trials (n=6,138)	(+) fewer CS [at 41 weeks]
	Dare et al (2018)	SRMA of 12 trials (n=6,814)	(=) CS [w membrane rupture]
	Miller et al (2015)	ARRIVE trial (n=162)	(=) CS
	Wennerholm et al (2009)	SRMA of 13 ARRIVE trials (n=5,920)	(+) fewer CS [≥ 41 weeks]
	Sotiriadis et al (2019)	SRMA of 5 ARRIVE trials (n=7,261)	(+) fewer CS
	Grobman et al (2018)	ARRIVE trial (n=6,106)	(+) fewer CS (18.6%:22.2%)
Infection	Dare et al (2018)	SRMA of 4 trials (n=445)	(=) uterine [w membrane rupture]
	Dare et al (2018)	SRMA of 9 trials (n=6,611)	(=) placental [w membrane rupture]
Neonatal			
Infection	Dare et al (2018)	SRMA of 12 trials (n=6,406)	(=) [w membrane rupture]
Birth injury	Buckles et al (2017)	IV elective delivery policy (n=410,459)	(−) more likely (8 x) [< 39 weeks]
APGAR score (5 minutes postpartum)	Saccone et al (2015)	SRMA of 5 ARRIVE trials (n=844)	(=)
	Sanchez et al (2003)	SRMA of 16 ARRIVE trials (n=6,588)	(=) [at 41 weeks]
	Middleton et al (2018)	SRMA of 16 ARRIVE trials (n=9,047)	(+) fewer APGAR <7 [≥ 39 weeks]
	Lynch et al (2019)	RDD Baby Bonus (n=1,862)	(−)
	Jürges (2018)	DID parental leave policy (n=565,000)	(=)
	Jacobson et al (2020)	Reduc.Form holiday effect (n=4,599)	(=)
	Schulkind et al (2014)	NatEx tax benefit (n=44,389)	(−) fewer normal APGAR scores at antedated birth
Birth weight	Buckles et al (2017)	IV elective delivery policy (n=410,459)	(−) -251 g [< 39 weeks]
	Lynch et al (2019)	RDD Baby Bonus (n=1,862)	(−) not postponing birth, <2500 g more likely
	Jürges (2018)	DID parental leave policy (n=565,000)	(=)
	Sotiriadis et al (2019)	SRMA of 5 ARRIVE trials (n=7,261)	(−) -81 g
	Grobman et al (2018)	ARRIVE trial (n=6,106)	(−) lower median birth weight
	Gans et al (2008)	NatEx: new Baby Bonus (n=1,040)	(−) -75 g
	Jacobson et al (2020)	Reduc.Form holiday effect (n=4,599)	(−) -2 g
	Hussain et al (2011)	SRMA of 14 ARRIVE trials (n=6,597)	(+) fewer births ≥ 4000 g [at 41 weeks]
Respiratory issues	Buckles et al (2017)	IV elective delivery policy (n=410,459)	(−) 4 x more likely [< 39 weeks]
	Lynch et al (2019)	RDD Baby Bonus (n=1,862)	(−) not postponing birth, normal breathing later
	Sotiriadis et al (2019)	SRMA of 5 ARRIVE trials (n=7,261)	(+)
	Jacobson et al (2020)	Reduc.Form holiday effect (n=4,599)	(=)
Mortality	Hussain et al (2011)	SRMA of 14 ARRIVE trials (n=6,597)	(+/ =) fewer deaths / equal stillbirths [at 41 weeks]
	Wennerholm et al (2009)	SRMA of 11 ARRIVE trials (n=5,920)	(+) fewer deaths [at 41 weeks]
	Saccone et al (2015)	SRMA of 5 ARRIVE trials (n=844)	(=) deaths
	Sanchez et al (2003)	SRMA of 16 ARRIVE trials (n=6,588)	(=) [at 41 weeks]
	Middleton et al (2018)	SRMA of 20 ARRIVE trials (n=9,960)	(+) deaths (2:16)/ stillbirths (1:10) [≥ 39 weeks]
	Gülmezoglu et al (2012)	SRMA of 17 ARRIVE trials (n=7,407)	(+) deaths (1:13) [≥ 39 weeks]
	Jürges (2018)	DID parental leave policy (n=565,000)	(=) death in first 7 days
Hospital/ intensive care visits	Lynch et al (2019)	RDD Baby Bonus (n=1,862)	(−) more visits for respiration
	Saccone et al (2015)	SRMA of 5 ARRIVE trials (n=844)	(=) #intensive care visits
	Sanchez et al (2003)	SRMA of 16 ARRIVE trials (n=6,588)	(=) #intensive care visits
	Middleton et al (2018)	SRMA of 13 ARRIVE trials (n=8,531)	(+) fewer intensive care visits [≥39 weeks]
	Gülmezoglu et al (2012)	SRMA of 10 ARRIVE trials (n=6,161)	(=) #intensive care visits [≥39 weeks]
	Dare et al (2018)	SRMA of 12 trials (n=6,814)	(+) fewer intensive care visits [w membrane rupture]
	Jürges (2018)	DID parental leave policy (n=565,000)	(=) #hospital visits
	Jacobson et al (2020)	Reduc.Form holiday effect (n=4,599)	(=) #intensive care visits

Notes: SRMA = systematic review & meta-analysis. Defaults: ARRIVE trials (randomizing induction vs. awaiting labor onset); natural experiments (postponing scheduled interventions) on singleton 39 weeks gestations. Deviations: marked, e.g., [<39 weeks] for preterm induction.

Table A.2: Overview Variable Specification

Variable	Specification	Function
Maternal characteristics		
region of origin	Germany=0; Middle/Northern Europe, North America=1; Mediterranean Countries=2; Eastern Europe=3; Middle East (incl. North Africa); 5=Asian (excl. 4); 9=other	core fixed effects
residence (state-level)	[1,16]	merge w holiday data
residence (3-digit zip code level)		clustering
single status	Binary indicator =1 if mother is single; 0 else	core controls
socio-economic status	Housewife=1; apprenticeship/college enrolment=2; un-/semiskilled workers=3; lower civil servants, employees w executing responsibilities, self-employed w small business=5; (at least) intermediate civil servants, employees w (at least) extensive responsibilities, self-employed w (at least) medium business, master, site foreman, overseer=6; unknown=9	core fixed effects
socio-economic status low	Binary indicator = 1 if socio-economic status=4; 0 if status=6	strata
employed	Binary indicator = 1 if mother is employed: 0 else	core controls
age	Age in years [18,35]	core fixed effects
older age	Binary indicator = 1 if age >25; 0 else	strata
weight^3	Pre-pregnancy weight (kg) as cubic	core controls
height^3	Height (cm) as cubic	core controls
bmi	Pre-pregnancy weight / (height/100)^2	core controls
bmi $\geqslant$ 90%ile	Binary indicator = 1 if BMI in 90%ile (all births); 0 else	strata
gestational age	#Days	miscellaneous controls
prenatal care	#Doctor visits	miscellaneous controls
prenatal care begin >12th week	Binary indicator = 1 if 1st prenatal visit >12th week of pregnancy	miscellaneous controls, placebo outcome
met obstetrician during pregnancy	Binary indicator = 1 if mother met obstetrician earlier in pregnancy	strata
hospital stay during pregnancy	Binary indicator = 1 if hospital stay earlier during pregnancy; 0 else	miscellaneous controls, placebo outcome
admitted after transfer	Binary indicator = 1 if transfer and receiving hospital id; 0 else	strata
year of completed delivery	Factor variable [2004;2019]	core fixed effects
month of completed delivery	Factor variable	additional fixed effects
weekday of completed delivery	Factor variable	additional fixed effects
hour of completed delivery	Factor variable	additional fixed effects
eclampsia	Binary variable = 1 if a mother has eclampsia; 0 else	strata
zero-precondition birth	Binary variable = 1 if non-risky pregnancy (gynecologist's label), single fetus, correct presentation, $\geqslant$ 37 gestation weeks, no prior uterine scar, no eclampsia, no growth restriction, age 18-35, BMI <90%tile, and <20 prenatal visits ; 0 else	strata
Neonatal characteristics		
birth order	Computed as the #previous (live + still) births +1	strata
birth weight	Measured in (g)	miscellaneous controls
body measures low	Binary indicator = 1 if weight <2.5 kg, length <45, or head circumf. <32 cm	
Hospital characteristics		
hospital id	Hospital identifier	additional fixed effects
emergency c-section time >20 min	Binary indicator = 1 if condition holds; 0 else	hospital controls
emergency c-section time <3 min	Binary indicator = 1 if condition holds; 0 else	hospital controls
hospital quality low	Binary indicator = 1 if emergency c-section time >20 min | <3 min	
hospital small	Binary indicator = 1 if hospital-year specific #obstetricians <median #obstetricians p.a.; 0 else	
hospital w/o in-patient midwives	Binary indicator = 1 if #deliveries w in-patient midwives >0; 0 else	strata
Health outcomes		
emergency c-section	Binary indicator = 1 if mother needs an emergency c-section; 0 else	dependent variable
perineal tearing (III/IV)	Binary indicator = 1 if mother suffers high-level perineal damage; 0 else	dependent variable
APGAR score (5 min. postbirth)	ordinal [0-10] neonatal fitness measure (10 being top score)	dependent variable
Hospital staff capacity outcomes		
labor duration	#Hours	dependent variable
pushing contractions	#Minutes	descriptives
maternal postnatal hospital stay	#Days from completed delivery till discharge	dependent variable
neonatal postnatal hospital stay	#Days from completed delivery till discharge	dependent variable

Continued on the next page.

Variable	Specification	Function
Treatments		
induced labor	Binary indicator = 1 if induction by membrane sweep, medication, or other procedures (excl. cervical ripening); 0 else	main explanatory variable
non-emergency c-section	Binary indicator if un/planned (excl. emergency) c-section; 0 else	explanatory variable
vaginally operative procedures	Binary indicator = 1 if forceps, spatula, vacuum, episiotomy; 0 else	explanatory variable
Staff capacity instruments		
predicted due date a non-working day	Binary indicator = 1 if predicted due date a Saturday/ Sunday/ public holiday; 0 else (incl. due dates updated by hospital)	instrument
predicted due date not informative	Binary indicator = 1 if hospital discards due date as invalid; 0 else	descriptives
pre-labor membrane rupture	Binary indicator = 1 if condition holds; 0 else	strata
pre-labor membrane rupture at night	Binary indicator = 1 if pre-labor membrane break 8pm-4am; 0 else	instrument
midwife shortage upon admission	0 if no current deliveries; else hospital-minute-wise ratio of #current deliveries w/o midwife/ #all current deliveries	instrument
Obstetrician preferences instruments		
preference induced labor	Obstetrician' s #prior inductions / #all prior deliveries	instrument
preference non-emergency c-section	Obstetrician's #prior non-emergency c-section/ #all prior deliveries	instrument
preference vaginally operative procedures	Obstetrician's #prior vaginally operative procedures/ #all prior deliveries	instrument

Notes: Annual Geburtshilfe datasets provided by the IQTIG institute constitute the main data source supplemented by calendar data to construct the non-working day instrument. A factor variable enters the regression as a set of binary indicators.

Table A.3: Overview Sample Specification

<table>
<tr><td rowspan="15">all births</td><td rowspan="14">non-missing for central variables</td><td rowspan="11">1st births</td><td rowspan="9">MAIN ANALYSIS SAMPLE
zero preconditions</td><td>w pre-labor membrane rupture</td></tr>
<tr><td>at hospitals w/o in-patient midwives</td></tr>
<tr><td>obstetrician unknown pre-admission</td></tr>
<tr><td>mothers admitted after transfer</td></tr>
<tr><td>mothers aged >26</td></tr>
<tr><td>single mothers</td></tr>
<tr><td>mothers w low socioeconomic status</td></tr>
<tr><td>at small hospitals</td></tr>
<tr><td>at low-quality hospitals</td></tr>
<tr><td colspan="2">delivered pre-arrival to hospital</td></tr>
<tr><td colspan="2">others</td></tr>
<tr><td rowspan="2">2nd births</td><td colspan="2">zero preconditions</td></tr>
<tr><td colspan="2">others</td></tr>
<tr><td colspan="3">higher birth orders</td></tr>
<tr><td colspan="4">else</td></tr>
</table>

Table A.4: Characteristics of Births Across Subsamples Dedicated to Heterogeneity Checks

zero-precondition 1st births	**all**	**strata**				
		maternal			hospital	
		age >26	single	low ses	small	low quality
	(1)	(2)	(3)	(4)	(5)	(6)
maternal characteristics						
german (yes/no)	0.81	0.85	0.90	0.80	0.83	0.78
single (yes/no)	0.11	0.10	1.00	0.11	0.12	0.10
low socio-economic status (yes/no)	0.80	0.78	0.77	1.00	0.83	0.76
age	28	30	27	28	28	28
bmi	24	24	24	24	24	24
pre-pregnancy weight (kg)	67	67	68	67	68	67
gestational age (#days)	280	280	280	280	280	280
prenatal care (#doctor visits)	12	12	11	12	12	11
prenatal care starting >12th week (yes/no)	0.077	0.055	0.10	0.081	0.079	0.088
neonatal characteristics						
birth weight (g)	3415	3424	3407	3411	3418	3417
hospital characteristics						
emergency c-section time >20 min (yes/no)	0.015	0.014	0.017	0.015	0.026	0.016
health outcomes						
emergency c-section (yes/no)	0.01	0.01	0.02	0.01	0.01	0.01
perineal tearing (III/IV) (yes/no)	0.02	0.03	0.02	0.02	0.02	0.03
APGAR score (5 min.)	9.7	9.7	9.7	9.7	9.7	9.7
hospital staff capacity outcomes						
labor duration (#hours)	6.8	6.8	6.8	6.7	6.6	6.8
maternal postnatal hospital stay (#days)	3.4	3.4	3.4	3.4	3.5	3.3
neonatal postnatal hospital stay (#days)	3.2	3.2	3.2	3.2	3.4	3.1
treatments						
induced labor (yes/no)	0.28	0.27	0.23	0.28	0.27	0.27
non-emergency c-section (yes/no)	0.21	0.27	0.27	0.26	0.27	0.26
vaginally operative procedues (yes/no)	0.32	0.32	0.32	0.32	0.32	0.32
instruments staff capacity						
predicted due date a non-working day (yes/no)	0.33	0.33	0.34	0.33	0.33	0.34
pre-labor membrane rupture 8pm-4am (yes/no)	0.15	0.16	0.14	0.15	0.14	0.14
midwife shortage upon admission [0,1]	0.58	0.57	0.60	0.58	0.73	0.50
instruments obstetricians' preferences						
preference induced labor [0,1]	0.23	0.23	0.23	0.23	0.23	0.24
preference non-emergency c-section [0,1]	0.34	0.35	0.34	0.35	0.33	0.30
preference vaginally operative procedues [0,1]	0.20	0.20	0.20	0.21	0.21	0.20
N	177,215	119,041	19,986	141,605	91,936	19,914
N obstetricians' preferences	66,916	44,313	2,462	54,194	39,235	10,488

Notes: IQTIG birth records for Germany 2015-2016. Means for the central analysis variables based on zero-precondition 1st births and subsamples stratified by maternal and hospital chracteristics. See Table A.2 and Table A.3 for details on sample and variable construction.

Table A.5: Characteristics of Births Across Subsamples Dedicated to Endogeneity Checks

	zero-precondition 1st births					delivery pre-arrival only precondition 1st births
	all	w pre-labor membrane rupture	at hospitals w/o in-patient midwives	unknown to obstetrician pre-admission	admitted after transfer	
	(1)	(2)	(3)	(4)	(5)	(6)
maternal characteristics						
german (yes/no)	0.81	0.81	0.80	0.81	0.78	0.83
single (yes/no)	0.11	0.11	0.14	0.16	0.13	0.12
low socio-economic status (yes/no)	0.80	0.80	0.81	0.79	0.75	0.80
age	28	28	28	28	29	28
bmi	24	24	24	24	24	24
pre-pregnancy weight (kg)	67	67	67	67	67	66
gestational age (#days)	280	278	280	279	280	279
prenatal care (#doctor visits)	12	11	11	11	11	11
prenatal care starting >12th week (yes/no)	0.077	0.072	0.075	0.085	0.087	0.976
neonatal characteristics						
birth weight (g)	3415	3391	3414	3403	3378	3382
hospital characteristics						
emergency c-section time >20 min (yes/no)	0.015	0.013	0.01	0.018	0.018	0.012
health outcomes						
emergency c-section (yes/no)	0.01	0.01	0.02	0.02	0.04	0.02
perineal tearing (III/IV) (yes/no)	0.02	0.02	0.03	0.02	0.02	0.03
APGAR score (5 min.)	9.7	9.7	9.7	9.7	9.1	9.7
hospital staff capacity outcomes						
labor duration (#hours)	6.8	6.7	6.8	6.7	5.7	7.1
maternal postnatal hospital stay (#days)	3.4	3.4	3.4	3.3	3.1	3.2
neonatal postnatal hospital stay (#days)	3.2	3.2	3.2	3.2	2.1	3.1
treatments						
induced labor (yes/no)	0.28	0.35	0.28	0.23	0.33	0.09
non-emergency c-section (yes/no)	0.21	0.23	0.25	0.24	0.34	0.18
vaginally operative procedues (yes/no)	0.32	0.33	0.32	0.33	0.33	0.35
instruments staff capacity						
predicted due date a non-working day (yes/no)	0.33	0.33	0.33	0.33	0.33	0.33
pre-labor membrane rupture 8pm-4am (yes/no)	0.15	0.50	0.16	0.16	0.15	0.26
midwife shortage upon admission [0,1]	0.58	0.57	0.57	0.59	0.65	0.56
instruments obstetricians' preferences						
preference induced labor [0,1]	0.23	0.24	0.25	0.23	0.25	0.23
preference non-emergency c-section [0,1]	0.34	0.34	0.39	0.32	0.35	0.34
preference vaginally operative procedues [0,1]	0.20	0.20	0.21	0.20	0.21	0.20
N	177,215	52,815	64,926	54,198	3,233	4,171
N obstetricians' preferences	66,916	18,885	17,205	17,776	917	1,397

Notes: IQTIG birth records for Germany 2015-2016. Means for the central analysis variables based on zero-precondition 1st births, subsamples, and the sample of mothers without pregnancy or birth risks suffering eclampsia. See Table A.2 and Table A.3 for details on sample and variable construction.

The original OLS model reads

$$\underset{}{Y_m} = \underset{1\times t}{\mathbf{t}_m'} \underset{t\times 1}{\beta} + \underset{1\times k}{\mathbf{x}_m'} \underset{k\times 1}{\delta} + \underset{1\times s}{\lambda} \underset{s\times 1}{\mathbf{1}} + v_m \quad (3)$$

where

treatments $\mathbf{t}_m \equiv \begin{bmatrix} InducedLabor\ (IL_m) \\ CSection\ (CS_m) \\ VaginalOperations\ (VO_m) \\ IL_m * CS_m \\ IL_m * VO_m \\ CS_m * VO_m \\ IL_m * CS_m * VO_m \end{bmatrix}$, covariates $\mathbf{x}_m \equiv \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_{k-1} \end{bmatrix}$, sets of fixed effects $\lambda \equiv \begin{bmatrix} \lambda_1 & \dots & \lambda_s \end{bmatrix}$

- Y_m: outcome of mother (or her neonate) m
- $IL_m,\ CS_m,\ \ VO_m \in \{0,1\}$: =1 if mother m has an induction (and maybe other interventions), a non-emergency c-section (dto.), or vaginal operations (dto.) respectively; 0 else
- details on pre-determined controls, fixed effects, and cluster-robust standard errors given below Table A.11

Then, the corresponding IV model can be written as

$$\underset{t\times 1}{\mathbf{t}_m} = \underset{t\times z}{\mathbf{\Gamma}} \underset{z\times 1}{\mathbf{z}_m} + \underset{t\times k}{\mathbf{\Phi}} \underset{k\times 1}{\mathbf{x}_m} + \underset{t\times s}{\mathbf{\Lambda}} \underset{s\times 1}{\mathbf{1}} + \underset{t\times 1}{\epsilon_m} \quad (4)$$

Notation builds on Equation 3. There are $1,...,t$ treatments, $1,...,k-1$ covariates, and $1,...,s$ sets of fixed effects observed for mother m, while $1,...,z$ instruments are defined as

either $\mathbf{z}_m \equiv \begin{bmatrix} DueDateNoWorkday\ (DN_m) \\ MembraneRuptureNight\ (RN_m) \\ MidwifeShortage\ (MS_m) \\ DN_m * RN_m \\ DN_m * MS_m \\ RN_m * MS_m \\ DN_m * RN_m * MS_m \end{bmatrix}$, or $\mathbf{z}_m \equiv \begin{bmatrix} InducedLaborPref\ (ILP_m) \\ CSectionPref\ (CSP_m) \\ VaginalOperPref\ (VOP_m) \\ ILP_m * CSP_m \\ ILP_m * VOP_m \\ CSP_m * VOP_m \\ ILP_m * CSP_m * VOP_m \end{bmatrix}$

- $DN_m \in \{0,1\}$: =1 if the due date is a weekend day or public holiday, 0 else
- $RN_m \in \{0,1\}$: =1 if mother m has a pre-labor membrane rupture between 8 pm to 4 am, 0 else
- $MS_m \in [0,1] = \begin{cases} 0 \text{ if \#current deliveries at that hospital} = 0 \\ \frac{\text{\#current deliveries at that hospital without a midwife}}{\text{\#current deliveries at that hospital}} \text{ else} \end{cases}$
- $ILP_m,\ CSP_m,\ VOP_m \in [0,1]$: mean prior rate of inductions, non-emergency c-sections, and vaginal operations of obstetrician treating mother m

Table A.6: Multi-Treatment Model Underidentification Tests

	Dependent variable:								
	induced labor	vaginal operations	non-emergency c-section	induction + surgery	non-emergency c-section x vaginal operations	non-emergency c-section x induced labor	vaginal operations x induced labor	non-emergency c-section x vaginal operations x induced labor	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
obstetricians' intervention preferences									66,916
	0.0000	0.0000	0.0000	0.0000					
	0.2800	0.6400	0.9200		0.7500	0.4100	0.5000	1.0000	
hospital staffing constraints									177,215
	0.8800	0.8000	0.7700	0.8000					
	0.9900	0.9500	0.7400		0.9500	1.0000	0.6900	1.0000	
Mean (dep. var.)	0.2770	0.3161	0.2558		0.0003	0.0896	0.0820	0.0001	

Notes: *p<0.1; **p<0.05; ***p<0.01. P-values reported for Sanderson and Windmeijer (2016)'s cluster-robust underidentification test (H_0 : *There is underidentification.*) for multiple endogenous treatments derived from running the Sargan-Hansen J-Test for overidentification (Cameron and Miller, 2015) in auxiliary regressions. The 1st set of instruments is derived from obstretricians' preferences to perform induction, c-section, or vaginal oerations. The 2nd set, based on hospital staffing constraints, involves *Midwife shortages upon arrival*, *Prelabor membrane rupture during night shift*, and *Due date a non-working day.* To identify a four-treatment model (Equation 1), the instruments' triple interaction is added to each set; for seven-treatment models (Equation 3), all instruments' interactions are added. The auxiliary regressions incl. core controls and are based on zero-precondition first-births (the 1st instrument set restricts further to non-missing obstetrician ids , see Table A.2 and Table A.3 for details on sample and variable construction). Robust standard errors clustered by 3-digit zip codes of maternal residence. Means are only available for the full sample of zero-precondition first births.

Table A.7: First-Stage Effects Based on Hospital Staff Capacity Constraints

	Dependent variable:						
	induced labor	vaginally operative procedures	non-emergency c-section	non-emergency c-section x vaginally operative procedures	non-emergency c-section x induced labor	vaginally operative procedures x induced labor	non-emergency c-section x vaginally operative procedures x induced labor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Midwife shortage upon admission	*insign.*	*sign.*	*sign.*	*insign.*	*sign.*	*insign.*	*insign.*
Due date non-working day	*insign.*	*insign.*	*insign.*	*insign.*	*insign.*	*insign.*	*insign.*
Pre-labor membrane rupture at night	*sign.*	*sign.*	*sign.*	*insign.*	*insign.*	*sign.*	*sign.*
Midwife shortage upon admission x due date non-working day	*sign.*	*insign.*	*insign.*	*insign.*	*insign.*	*insign.*	*insign.*
Midwife shortage upon admission x pre-labor membrane rupture at night	*sign.*	*insign.*	*insign.*	*insign.*	*insign.*	*insign.*	*insign.*
Due date non-working day x pre-labor membrane rupture at night	*insign.*	*insign.*	*insign.*	*insign.*	*insign.*	*sign.*	*insign.*
Midwife shortage upon admission x due date non-working day x pre-labor membrane rupture at night	*insign.*	*insign.*	*insign.*	*insign.*	*insign.*	*sign.*	*insign.*
Mean (dependent variable)	0.2770	0.3161	0.2558	0.0003	0.0896	0.0820	0.0001
Underidentification (p-value)	0.9900	0.9500	0.7400	0.9500	1.0000	0.6900	1.0000

Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births (N=177,215). Sample and variable creation detailed in Table A.2 and Table A.3. *For some remote execution issue, coefficents on interacted dependent variables are not released, only an indicator for at least 10%-level significance "sign." or less "insign."*. Underlying regressions follow model Equation 3 and use robust standard errors clustered by 3-digit zip codes of maternal residence. Underidentification is tested (see Table A.6 for details). Means are available for the main sample.

Table A.8: First-Stage Effects Based on Obstetricians' Preferences for Interventions

	Dependent variable:						
	induced labor	vaginally operative procedures	non-emergency c-section	non-emergency c-section x vaginally operative procedures	non-emergency c-section x induced labor	vaginally operative procedures x induced labor	non-emergency c-section x vaginally operative procedures x induced labor
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Induced labor preference	*sign.*	*sign.*	*insign.*	*sign.*	*insign.*	*sign.*	*sign.*
Non-emergency c-section preference	*sign.*	*insign.*	*sign.*	*insign.*	*sign.*	*sign.*	*sign.*
Vaginally operative procedures' preference	*sign.*	*sign.*	*sign.*	*insign.*	*sign.*	*sign.*	*sign.*
Induced labor preference x Non-emergency c-section preference	*insign.*	*insign.*	*insign.*	*insign.*	*sign.*	*insign.*	*sign.*
Induced labor preference x Vaginally operative procedures' preference	*sign.*	*insign.*	*insign.*	*insign.*	*sign.*	*insign.*	*sign.*
Non-emergency c-section preference x Vaginally operative procedures' preference	*sign.*	*sign.*	*insign.*	*insign.*	*sign.*	*insign.*	*sign.*
Induced labor preference x Non-emergency c-section preference x Vaginally operative procedures' preference	*sign.*	*insign.*	*insign.*	*sign.*	*insign.*	*sign.*	*insign.*
Mean (dependent variable)	0.2770	0.3161	0.2558	0.0003	0.0896	0.0820	0.0001
Underidentification (p-value)	0.2800	0.6400	0.9200	0.7500	0.4100	0.5000	1.0000

Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births (N=177,215). Instrumenting by intervention preferences creates a subsample of births with non-missing obstetrician id (N=66,916). Sample and variable creation detailed in Table A.2 and Table A.3. *For some remote execution issue, coefficents on interacted dependent variables are not released, only an indicator for at least 10%-level significance "sign." or less "insign.".* Underlying regressions follow model Equation 3 and use robust standard errors clustered by 3-digit zip codes of maternal residence. Underidentification is tested (see Table A.6 for details). Means are available for the main sample.

Table A.9: Reduced Form Effects Based on Hospital Staff Capacity Constraints

	Dependent variable:				
	maternal health		staff capacity		
	emergency c-section	perineal tearing (III/IV)	labor duration (#hours)	hospital stay	
				mother	neonate
	(1)	(2)	(3)	(4)	(5)
Midwife shortage upon admission	0,0024***	-0,0027***	-1,0266***	0,2289***	0,2439***
	(0,0009)	(0,0010)	(0,0597)	(0,0159)	(0,0184)
Due date non-working day	-0.0009	0.0006	0.0622	0.0157	0.0406**
	(0.0009)	(0.0013)	(0.0489)	(0.0130)	(0.0180)
Pre-labor membrane rupture at night	-0.0007	-0.0011	-0.1124	-0.0392**	-0.0317
	(0.0013)	(0.0020)	(0.0682)	(0.0179)	(0.0204)
Midwife shortage upon admission x Due date non-working day	-0.0002	-0.0003	-0.0556	-0.0271	-0.0410
	(0.0012)	(0.0017)	(0.0629)	(0.0174)	(0.0250)
Midwife shortage upon admission x Pre-labor membrane rupture at night	-0.0030*	0.0016	0.1553*	-0.0034	-0.0240
	(0.0017)	(0.0024)	(0.0878)	(0.0249)	(0.0266)
Due date non-working day x Pre-labor membrane rupture at night	0.0007	-0.0034	0.0108	-0.0116	-0.0161
	(0.0023)	(0.0033)	(0.1128)	(0.0282)	(0.0348)
Midwife shortage upon admission x Due date non-working day x Pre-labor membrane rupture at night	0.0046	0.0022	-0.1053	0.0465	0.0761
	0.0032	(0.0042)	(0.1547)	(0.0406)	(0.0567)
Mean (dependent variable)	0.01	0.02	6.80	3.40	3.20
Adjusted R^2	0.0011	0.0017	0.0147	0.0162	0.0079

Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births (N=177,215). Sample and variable creation detailed in Table A.2 and Table A.3. Reported coefficients stem from reduced forms following model Equation 3 (detailed below Table A.11). Robust standard errors clustered by 3-digit zip codes of maternal residence. Adjusted R^2 reported for regressions including core controls.

Table A.10: Reduced Form Effects Based on Obstetricians' Preferences for Interventions

	Dependent variable:				
	maternal health		staff capacity		
	emergency c-section	perineal tearing (III/IV)	labor duration (#hours)	hospital stay	
				mother	neonate
	(1)	(2)	(3)	(4)	(5)
Induced labor preference	-0.0004	0.0040	-1.0944***	-0.2074*	-0.2102**
	(0.0039)	(0.0095)	(0.3743)	(0.1065)	(0.1058)
Non-emergency c-section preference	0.0401***	0.0070	-5.0095***	0.9834***	0.7371***
	(0.0057)	(0.0048)	(0.2763)	(0.0562)	(0.0801)
Vaginally operative procedures' preference	0.0128**	0.0026	-1.7420***	0.6928***	0.6602***
	(0.0056)	(0.0084)	(0.3457)	(0.0989)	(0.1156)
Induced labor preference x Non-emergency c-section preference	0.0432**	0.0004	1.6491**	0.1917	0.1332
	(0.0206)	(0.0161)	(0.7761)	(0.1742)	(0.2101)
Induced labor preference x Vaginally operative procedures' preference	0.0017	-0.0111	2.1683**	-0.0396	0.0654
	(0.0123)	(0.0226)	(0.8737)	(0.2014)	(0.2644)
Non-emergency c-section preference x Vaginally operative procedures' preference	-0.0638**	0.0507	6.3833***	0.8410**	1.1898**
	(0.0287)	(0.0314)	(1.4891)	(0.3739)	(0.4624)
Induced labor preference x Non-emergency c-section preference x Vaginally operative procedures' preference	0.1527	0.0993	-8.7465**	-1.6280	-2.7907**
	(0.0982)	(0.0991)	(4.4016)	(1.1039)	(1.2280)
Mean (dependent variable)	0.01	0.02	6.80	3.40	3.20
Adjusted R^2	0.0092	0.0025	0.0402	0.0433	0.0157

Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births. Instrumenting by intervention preferences creates a subsample of births with non-missing obstetrician id (N=66,916). Sample and variable creation detailed in Table A.2 and Table A.3. Reported coefficients stem from reduced forms following model Equation 3 (detailed below Table A.11). Robust standard errors clustered by 3-digit zip codes of maternal residence. Adjusted R^2 reported for regressions including core controls. Means are available for the main sample.

Table A.11: Maternal Health Effects of Non-Medically Indicated Induced Labor

	OLS		IV			
Instruments based on			staff capacity		obstetricians' preferences	
Dependent variable	emergency c-section	perineal tearing (III/IV)	emergency c-section	perineal tearing (III/IV)	emergency c-section	perineal tearing (III/IV)
	(1)	(2)	(3)	(4)	(5)	(6)
No controls	0.0205***	0.0029**	0.2594	-0.1425	-0.2074	0.0867
	(0.0017)	(0.0013)	(6.5993)	(0.3221)	(0.3652)	(0.5341)
Core controls	0.0201***	0.0032**	-0.7842	-0.1189	-0.1710	0.1017
	(0.0017)	(0.0013)	(53.4963)	(1.3234)	(0.3484)	(0.5327)
Add month, weekday, hour FE	0.0197***	0.0032**	0.3764	-0.1283	-0.2067	0.1176
	(0.0018)	(0.0013)	(4.9598)	(0.2761)	(0.2255)	(0.3652)
Add hospital controls	0.0199***	0.0025*	0.4959	-0.1255	-0.1829	0.0523
	(0.0018)	(0.0013)	(8.6641)	(0.4045)	(0.2214)	(0.3540)
Add hospital FE	0.0195***	0.0025*	0.2141	-0.1555	-0.4820	0.4643
	(0.0018)	(0.0013)	(0.2688)	(0.2232)	(0.9310)	(0.6601)
Core controls & labor	0.0224***	0.0033**	0.1021	-0.1333	-0.3167	0.0420
	(0.0017)	(0.0013)	(0.1990)	(0.1303)	(0.3714)	(0.6500)
Miscelleanous controls	0.0200***	0.0030**	0.1776	-0.1466	-0.1081	-0.0255
	(0.0017)	(0.0013)	(0.8141)	(0.1343)	(0.3033)	(0.4903)
Main effects only (core controls)	0.0081***	0.0029***	0.0996	-0.0563	0.0730***	0.0223
	(0.0007)	(0.0009)	(0.1968)	(0.1920)	(0.0196)	(0.0207)
Mean (dependent variable)	0.01	0.02	0.01	0.02		
Adjusted R^2	0.0227	0.0157				
N	177,215	177,215	177,215	177,215	66,916	66,916

Notes: *p<0.1; **p<0.05; ***p<0.01. Birth records from all German hospitals covering 2015-2016, provided by the IQTIG institute. Linear probability models based on the main analysis sample of zero-precondition first births (neither pregnancy nor birth risks known antepartum), for which all central regression inputs are non-missing. Sample and variable creation detailed in Table A.2 and Table A.3. Reported coefficients for induced labor stem from separate regressions following model Equation 3. Intervention treatments are represented by binary indicators for induced labor, non-emergency c-sections, and vaginally operative procedures. Staff capacity-based instruments are binary indicators for a mother's due date on a non-working day, a pre-labor rupture of membranes between 8 pm and 4 am, and a minute-wise measure $\in [0, 1]$ of midwife shortages upon maternal admission. The instruments based on obstetricians' preferences are computed for each of the three main interventions as the mean intervention rate across an obstetrician's past deliveries. Instrumenting by intervention preferences creates a subsample of births with non-missing obstetrician id. Treatments and instruments enter as main effects and interactions. Core controls include the year of delivery, a mother's age, her region of origin (7 categories), her socio-economic status (6 categories), and her single status (yes/no), where categorical variables enter as sets of binary indicators. Moreover, continuous measures are created for maternal height (as cubic), maternal weight at the beginning of the pregnancy (as cubic), and maternal BMI. Binary hospital (stay) controls indicate whether 1) the mother brings her own midwife, 2) she has been introduced to her obstetrician during pregnancy, and 3) documentation of her delivery seem to be done in a haste. Miscellaneous controls are binary indicators for maternal alcohol consumption, psychological or social problems, minor diseases or pregnancy risks , as well the count of doctor visits. Finally, there is a dummy for maternal employment status. Main effects only (...) refers to the core specification w/o interactions of treatments or instruments. Robust standard errors clustered by 3-digit zip code of maternal residence. Adjusted R^2 reported for regressions including core controls. Means are available for the main sample.

Table A.12: Sample-specific Health Effects of Non-Medically Indicated Induced Labor

Instruments based on	staff capacity		obstetricians' preferences			
Dependent variable	emergency c-section	perineal tearing (III/IV)	emergency c-section	perineal tearing (III/IV)	$N_{(1)-(2)}$	$N_{(3)}$
	(1)	(2)	(3)	(4)		
zero-precondition 1st births	-0.7842 (53.4963)	-0.1189 (1.3234)	-0.1710 (0.3484)	0.1017 (0.5327)	177,215	66
w pre-labor membrane rupture	0.0613 (7.6489)	0.1401 (8.0396)	0.3216 (0.3645)	0.1087 (0.1642)	52,815	18
at hospitals w/o in-patient midwives	-0.9998 (11.0254)	0.3518 (2.3171)	0.0147 (0.1188)	0.0610 (0.1712)	64,926	17
unknown to obstetrician pre-admission	-0.1817 (1.2580)	-0.2840 (1.2355)	0.0485 (0.6325)	-0.1975 (0.4618)	54,198	17
admitted after transfer	0.3296 (0.5707)	-0.0541 (0.3147)	0.6262 (1.2853)	0.2765 (1.4799)	3,233	
to mothers aged >26	0.3738 (1.2845)	-0.1431 (0.7605)	0.8961 (2.5987)	-1.9978 (5.5746)	119,041	44
to single mothers	-0.0081 (0.2146)	-0.3490 (0.1988)	0.1557 (0.2375)	0.1207 (0.2403)	19,986	2
to mothers w low socio-economic status	0.0383 (0.8211)	-0.1341 (0.3339)	0.0140 (0.3694)	-0.4802 (0.9299)	141,605	54
at small hospitals	0.0771 (0.3858)	-0.0767 (0.2856)	-0.9108 (1.2815)	0.2442 (0.6612)	91.936	39
at low quality hospitals	-0.2210 (0.3938)	-0.0073 (0.3432)	-0.1292 (0.1752)	-0.0496 (0.1527)	19,914	10
delivery pre-arrival	2.0004 (83.8723)	-14.9676 (572.1791)	0.0752 (0.8968)	-0.3553 (0.4475)	4,171	1
zero-precondition 2nd births	-0.0542 (0.1918)	0.0044 (0.1239)	0.0391 (0.0414)	-0.0310 (0.0337)	81,896	27

Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births. Instru
ing by intervention preferences creates a subsample of births with non-missing obstetrician id. Sample and variable creation detailed in Tab
and Table A.3. Reported coefficients for induced labor stem from separate regressions following model Equation 3 (detailed below Table A.1
run for alternative samples (see Table A.3). Robust standard errors clustered by 3-digit zip codes of maternal residence.

Table A.13: Hospital Staff Capacity Effects of Non-Medically Indicated Induced Labor

	OLS			IV					
Instruments based on				staff capacity			obstetricians' preferences		
Dependent variable	labor duration (#hours)	postnatal hospital stay (#days)		labor duration (#hours)	postnatal hospital stay (#days)		labor duration (#hours)	postnatal hospital stay (#days)	
		mother	neonate		mother	neonate		mother	neonate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
No controls	-1.0353***	0.1415***	0.0989***	-13.5044	1.1960	2.8748	-29.7075	3.7541	2.7828
	(0.0498)	(0.0110)	(0.0135)	(373.9101)	(14.0921)	(14.3149)	(47.3025)	(5.6907)	(5.9303)
Core controls	-1.0117***	0.1432***	0.1068***	44.4728	-0.8431	0.6261	-31.8266	3.8466	2.9638
	(0.0500)	(0.0109)	(0.0134)	(2986.7851)	(111.1334)	(128.8600)	(51.4751)	(5.9850)	(6.2079)
Add month, weekday, hour FE	-1.0231***	0.0883***	0.0509***	-29.7799	2.1956	4.1533	-20.5187	2.1190	1.6719
	(0.0511)	(0.0107)	(0.0141)	(464.6725)	(26.0612)	(21.4955)	(32.0754)	(3.3148)	(3.8279)
Add hospital controls	-1.0371***	0.0911***	0.0554***	-41.0236	2.6662	4.4217	-28.3986	2.0995	0.7391
	(0.0513)	(0.0107)	(0.0141)	(809.5115)	(40.8639)	(30.3282)	(33.1837)	(3.6160)	(3.8385)
Add hospital FE	-0.9030***	0.1168***	0.0920***	-8.9386	0.1634	1.8641	-12.6147	4.2166	-0.3783
	(0.0474)	(0.0105)	(0.0135)	(7.7969)	(2.8216)	(5.9000)	(33.3624)	(5.9394)	(5.4054)
Core controls & labor	-1.2397***	0.1451***	0.1074***	-6.6251	1.0142	2.8021	-26.5807	3.5885	2.9673
	(0.0507)	(0.0109)	(0.0134)	(8.7821)	(1.5816)	(2.7709)	(43.2773)	(5.2765)	(5.5831)
Miscelleanous controls	-1.0476***	0.1471***	0.1169***	-9.8327	1.0803	2.8048	-39.1492	3.6921	2.0941
	(0.0514)	(0.0108)	(0.0135)	(53.4313)	(2.3680)	(3.7456)	(50.1474)	(5.8272)	(5.4258)
Main effects only (core controls)	-0.7063***	0.0897***	0.0452***	-7.4338	-0.5037	-5.0500	-1.6727	-1.1620***	-1.3736***
	(0.0401)	(0.0080)	(0.0115)	(11.2756)	(3.6683)	(9.5341)	(1.0939)	(0.2981)	(0.3351)
Mean (dependent variable)	6.80	3.40	3.20	6.80	3.40	3.20			
Adjusted R^2	0.1459	0.1236	0.0568						
N	177,215	177,215	177,215	177,215	177,215	177,215	177,215	66,916	66,916

Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births. Instrumenting by intervention preferences creates a subsample of births with non-missing obstetrician id. Sample and variable creation detailed in Table A.2 and Table A.3. Reported coefficients for induced labor stem from separate regressions following model Equation 3 (detailed below Table A.11). Robust standard errors clustered by 3-digit zip codes of maternal residence. Adjusted R^2 reported for regressions including core controls. Means are available for the main sample.

Table A.14: Relative Effects of Induction vs. Surgical Intervention

	Dependent variable: hospital stay (# days)					
	OLS		IV			
			staff capacity		obstetricians' preferences	
	mother	neonate	mother	neonate	mother	neonate
	(1)	(2)	(3)	(4)	(5)	(6)
Induced labor	0,1432***	0.1068***	-0.8431	0.6261	3.8466	2.9638
	(0.0109)	(0.0134)	(111.1334)	(128.8600)	(5.9850)	(6.2079)
Non-emergency c-section	1.3202***	1.1537***	-9.8964	-11.4455	2.3109**	2.0986*
	(0.0175)	(0.0198)	(752.6946)	(868.8066)	(1.1644)	(1.1971)
Vaginal operations	0.1828***	0.1493***	-29.3680	-31.5018	4.4629	3.9694
	(0.0114)	(0.0139)	(1942.0808)	(2240.6533)	(3.8845)	(4.1319)
Induced labor x Non-emergency c-section	-0.1191***	-0.1239***	69.0230	76.0869	-3.3350	-3.5243
	(0.0192)	(0.0292)	(3863.9451)	(4459.4136)	(4.8581)	(4.9821)
Induced labor x Vaginal operations	-0.0681***	-0.0943***	9.3691	6.6078	-12.8115	-10.1293
	(0.0174)	(0.0220)	(759.7923)	(875.8903)	(14.8890)	(15.6703)
Non-emergency c-section x Vaginal operations	-0.2544	-0.2741	-107.0354	-240.5277	-130.1122	-281.2032
	(0.1984)	(0.2528)	(2575.7215)	(2882.4232)	(631.5509)	(670.7902)
Induced labor x Non-emerg. c-section x Vag. oper.	0.0907	0.2971	2686.9309	3597.2043		
	(0.3656)	(0.4860)	(154104.5424)	(177408.8657)		
Mean (dependent variable)	3.40	3.20	3.40	3.20		
Adjusted R^2	0.1236	0.0568				
N	177,2150	177,2150	177,2150	177,2150	66,916	66,916

Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births. Instrumenting by intervention preferences creates a subsample of births with non-missing obstetrician id. Sample and variable creation detailed in Table A.2 and Table A.3. Reported coefficients for intervention follow model Equation 3 (detailed below Table A.11). Robust standard errors clustered by 3-digit zip codes of maternal residence. Adjusted R^2 reported for regressions including core controls. Means are available for the main sample.

Table A.15: Sample-specific Staff Capacity Effects of Non-Medically Indicated Induced Labor

Instruments based on	staff capacity			obstetricians' preferences				
Dependent variable	labor duration (#hours)	postnatal hospital stay (#days)		labor duration (#hours)	postnatal hospital stay (#days)			
		mother	neonate		mother	neonate	$N_{(1)-(3)}$	$N_{(4)-(6)}$
	(1)	(2)	(3)	(4)	(5)	(6)		
zero-precondition 1st births	44.4728	-0.8431	0.6261	-31.8266	3.8466	2.9638	177,215	66,916
	(2986.7851)	(111.1334)	(128.8600)	(51.4751)	(5.9850)	(6.2079)		
with pre-labor membrane rupture	13.2418	31.4363	53.7219	-7.1686	-1.1150	-0.1051	52,815	18,885
	(407.1023)	(604.7513)	(1065.8878)	(8.9018)	(2.5213)	(3.2961)		
at hospitals w/o in-patient midwives	73.7779	-22.1140	-12.9251	4.5229	0.2519	0.9101	64,926	17,205
	(626.3206)	(174.6608)	(77.2291)	(6.6268)	(2.1979)	(2.2216)		
unknown to obstetrician pre-admission	-6.3800	-8.8533	-5.5982	33.4803	-3.2455	-2.2927	54,198	17,776
	(25.7600)	(30.5830)	(27.8711)	(27.9019)	(5.2948)	(5.9328)		
admitted after transfer	6.2873	1.6719	-1.3041	22.8765	7.2392	11.8903	3,233	917
	(21.1890)	(5.1644)	(2.5281)	(42.6121)	(13.5778)	(16.6935)		
to mothers aged >26	-15.7227	0.4123	2.9843	21.5082	-12.4561	-12.0063	119,041	44,313
	(80.5652)	(5.2204)	(11.1776)	(167.0019)	(30.7977)	(29.4835)		
to single mothers	5.9564	-0.9699	-0.6513	4.9491	3.8176	3.2246	19,986	2,462
	(17.8985)	(4.0973)	(3.3721)	(12.6889)	(3.5791)	(4.0644)		
to mothers w low socio-economic status	2.1151	0.5567	1.6434	-70.7439	4.9493	0.2901	141,605	54,194
	(41.6917)	(2.9570)	(3.5705)	(100.3852)	(8.9630)	(5.5070)		
at small hospitals	2.5659	-2.8901	-0.2278	-27.0442	9.1423	2.8971	91,936	39,235
	(36.5997)	(4.2561)	(12.4331)	(39.6827)	(18.8312)	(13.7552)		
at low-quality hospitals	-32.0840	6.1798	3.4350	14.2136	1.7202	4.1716	19,914	10,488
	(27.4157)	(8.8278)	(9.6531)	(14.8235)	(1.7062)	(3.6793)		
delivery pre-arrival	-590.1633	-121.3088	-88.6764	24.5799	-9.3640	-6.5620	4,171	1,395
	(26621.1444)	(4140.5791)	(2906.5923)	(30.8458)	(7.9748)	(7.1380)		
zero-precondition 2nd births	-2.1172	-3.6416	-0.2685	-5.1756	0.5379	-0.1112	81,896	27,558
	(12.0511)	(5.1512)	(4.6309)	(1.7290)	(0.7955)	(0.7917)		

Table A.16: Relative Effects of Induction vs. Surgical Intervention

	Dependent variable: hospital stay (# days)					
	OLS		IV			
			staff capacity		obstetricians' preferences	
	mother	neonate	mother	neonate	mother	neonate
	(1)	(2)	(3)	(4)	(5)	(6)
Induced labor	0,1432***	0.1068***	-0.8431	0.6261	3.8466	2.9638
	(0.0109)	(0.0134)	(111.1334)	(128.8600)	(5.9850)	(6.2079)
Non-emergency c-section	1.3202***	1.1537***	-9.8964	-11.4455	2.3109**	2.0986*
	(0.0175)	(0.0198)	(752.6946)	(868.8066)	(1.1644)	(1.1971)
Vaginal operations	0.1828***	0.1493***	-29.3680	-31.5018	4.4629	3.9694
	(0.0114)	(0.0139)	(1942.0808)	(2240.6533)	(3.8845)	(4.1319)
Induced labor x Non-emergency c-section	-0.1191***	-0.1239***	69.0230	76.0869	-3.3350	-3.5243
	(0.0192)	(0.0292)	(3863.9451)	(4459.4136)	(4.8581)	(4.9821)
Induced labor x Vaginal operations	-0.0681***	-0.0943***	9.3691	6.6078	-12.8115	-10.1293
	(0.0174)	(0.0220)	(759.7923)	(875.8903)	(14.8890)	(15.6703)
Non-emergency c-section x Vaginal operations	-0.2544	-0.2741	-107.0354	-240.5277	-130.1122	-281.2032
	(0.1984)	(0.2528)	(2575.7215)	(2882.4232)	(631.5509)	(670.7902)
Induced labor x Non-emerg. c-section x Vag. oper.	0.0907	0.2971	2686.9309	3597.2043		
	(0.3656)	(0.4860)	(154104.5424)	(177408.8657)		
Mean (dependent variable)	3.40	3.20	3.40	3.20		
Adjusted R^2	0.1236	0.0568				
N	177,2150	177,2150	177,2150	177,2150	66,916	66,916

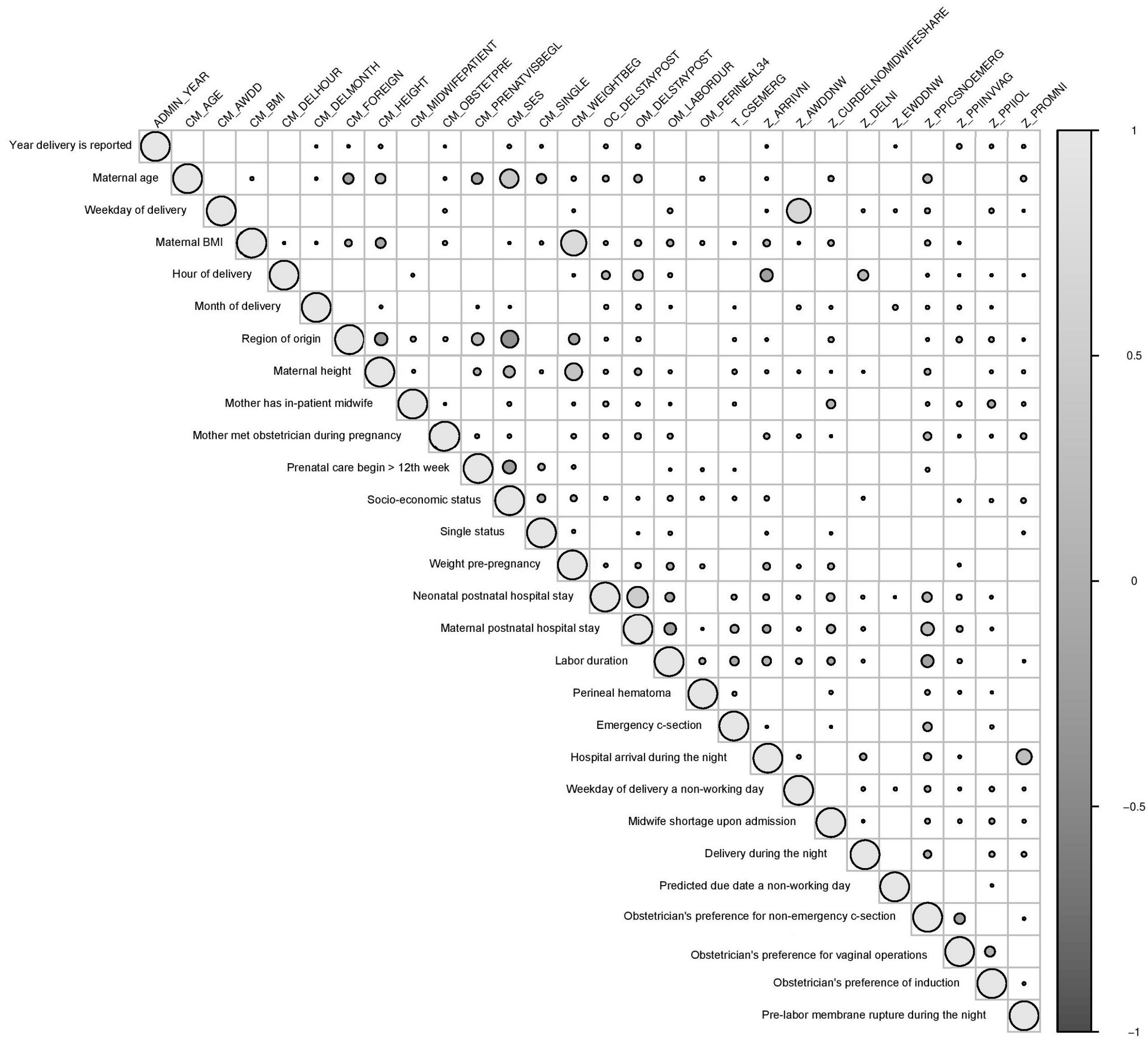
Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births. Instrumenting by intervention preferences creates a subsample of births with non-missing obstetrician id. Sample and variable creation detailed in Table A.2 and Table A.3. Reported coefficients for intervention follow model Equation 3 (detailed below Table A.11). Robust standard errors clustered by 3-digit zip codes of maternal residence. Adjusted R^2 reported for regressions including core controls. Means are available for the main sample.

Table A.17: Placebo Effects of Non-Medically Indicated Induced Labor

	OLS	IV	
Instruments based on		staff capacity	obstetricians' preferences
Dependent variable	1st prenatal care >12th week		
	(1)	(2)	(3)
No controls	0.0071***	-0.0459	0.9548
	(0.0024)	(2.3953)	(1.0697)
Core controls	0.0006	0.6016	0.9232
	(0.0023)	(31.7039)	(1.1911)
Add month. weekday. hour FE	0.0016	-0.2098	0.5529
	(0.0024)	(4.3706)	(0.6172)
Add hospital controls	0.0024	-0.3335	0.5504
	(0.0024)	(8.0282)	(0.6471)
Add hospital FE	0.0035	0.0041	0.6069
	(0.0024)	(0.3108)	(1.0568)
Core controls & labor	0.0006	0.0358	0.7703
	(0.0023)	(0.3282)	(1.0084)
Miscellaneous controls	0.0081***	-0.0023	0.5192
	(0.0023)	(0.7882)	(0.7658)
Main effects only (core controls)	0.0025*	0.0520	-0.0397
	(0.0014)	(0.3512)	(0.0394)
Mean (dependent variable)	0.077	0.077	
Adjusted R^2	0.0636		
N	177,215	177,215	66,916

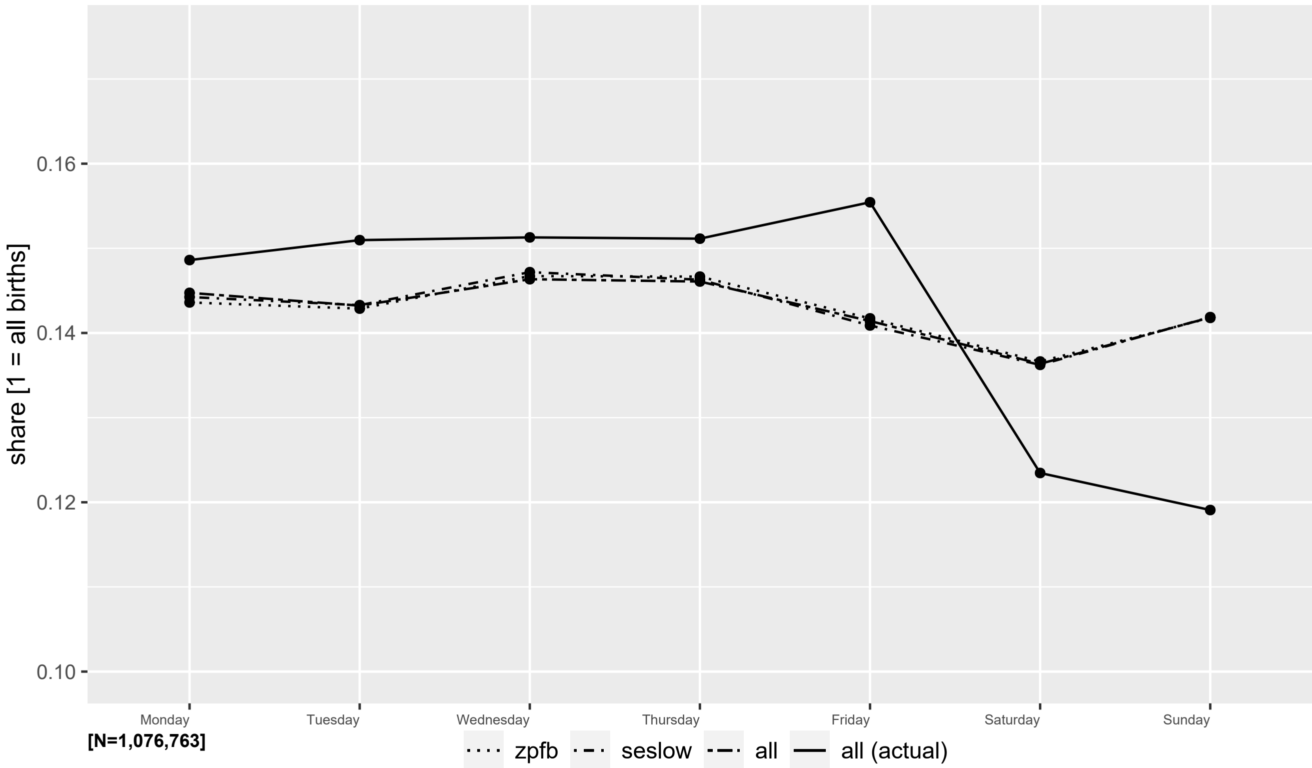
Notes: *p<0.1; **p<0.05; ***p<0.01. IQTIG birth records for Germany 2015-2016. The main sample are zero-precondition first-births. Instrumenting by intervention preferences creates a subsample of births with non-missing obstetrician id. Sample and variable creation detailed in Table A.2 and Table A.3. Reported coefficients for induced labor stem from separate regressions following model Equation 3 (detailed below Table A.11). Robust standard errors clustered by 3-digit zip codes of maternal residence. Adjusted R^2 reported for regressions including core controls. Means are available for the main sample.

Figure A.1: Heatmap of Unconditional Correlations of Central Variables



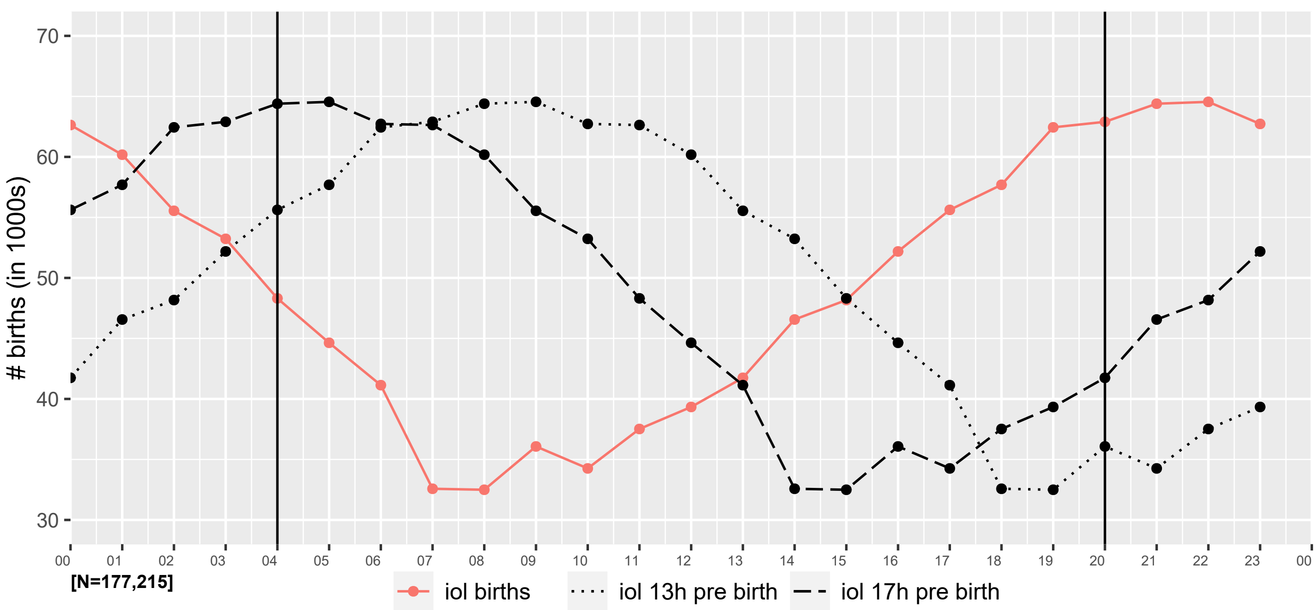
Source: IQTIG German hospital birth records for 2015-2016 restricted to a subsample of zero-precondition first births with information on obstetrician ids (N=66,916 out of 177,215). Sample and variable construction is detailed in the notes to Table A.2 and Table A.3. Variables not derived from obstetrician ids correlate similarly in the main sample. Correlation *values* are indicated by the color ramp, *significance* by the size of circles (no circle if insignificant at the 10% level). Variable names, not labels are shown on the horizontal line. Own calculations.

Figure A.2: Socioeconomic Status & Due Date Distribution Across Weekdays



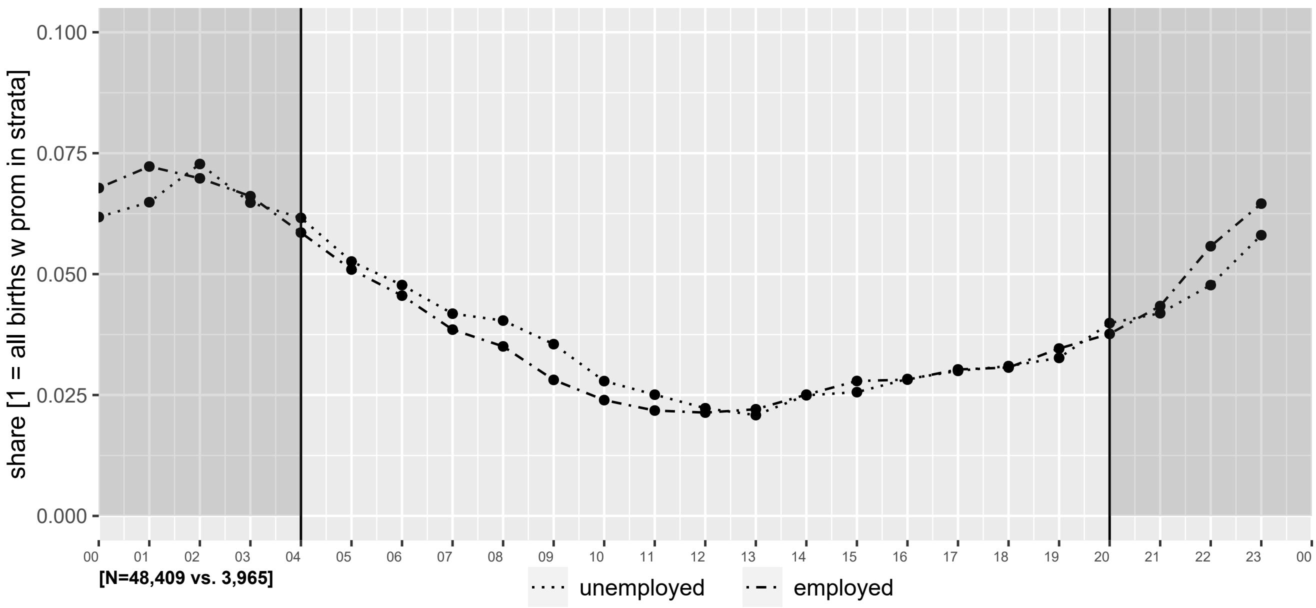
Source: IQTIG German hospital birth records for 2015-2016. *all* refers to the sample of all 1st and 2nd births. *zpfb* refers to the main analysis sample of zero-precondition first births (detailed in Table A.2 and Table A.3), *seslow* restricts this sample to mothers with lower socioeconomic status. Benchmarking fluctuations in predicted due dates, *actual* plots actual weekdays of delivery. Own calculations.

Figure A.3: Intervention & Delivery Timing of Induced Births Across Daily Hours



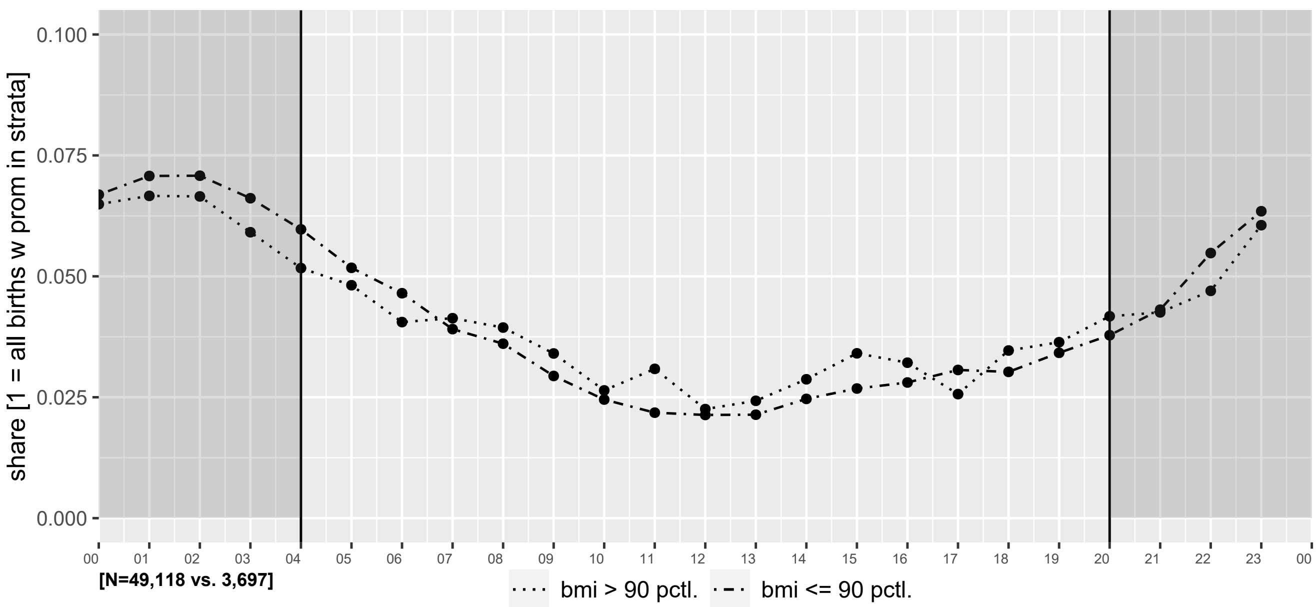
Source: IQTIG German hospital birth records for 2015-2016 restricted to the main analysis sample of zero-precondition first births (detailed in Table A.2 and Table A.3). Unobserved induction timing is proxied lagging birth timing by, e.g., 17 hours. Own calculations.

Figure A.4: Employment Status & Births after Pre-Labor Membrane Ruptures Across Daily Hours

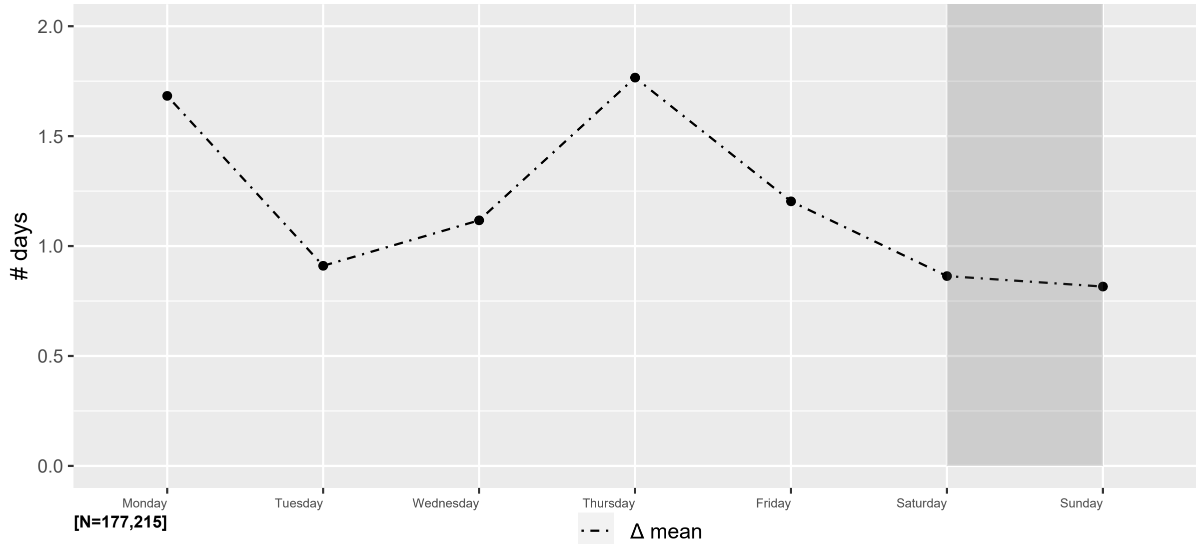


Source: IQTIG German hospital birth records for 2015-2016 restricted to the subsample of zero-precondition first-time mothers with pre-labor membrane ruptures (detailed in Table A.2 and Table A.3). Strata by maternal employment status. Own calculations.

Figure A.5: Maternal Fitness & Births after Pre-Labor Membrane Ruptures Across Daily Hours

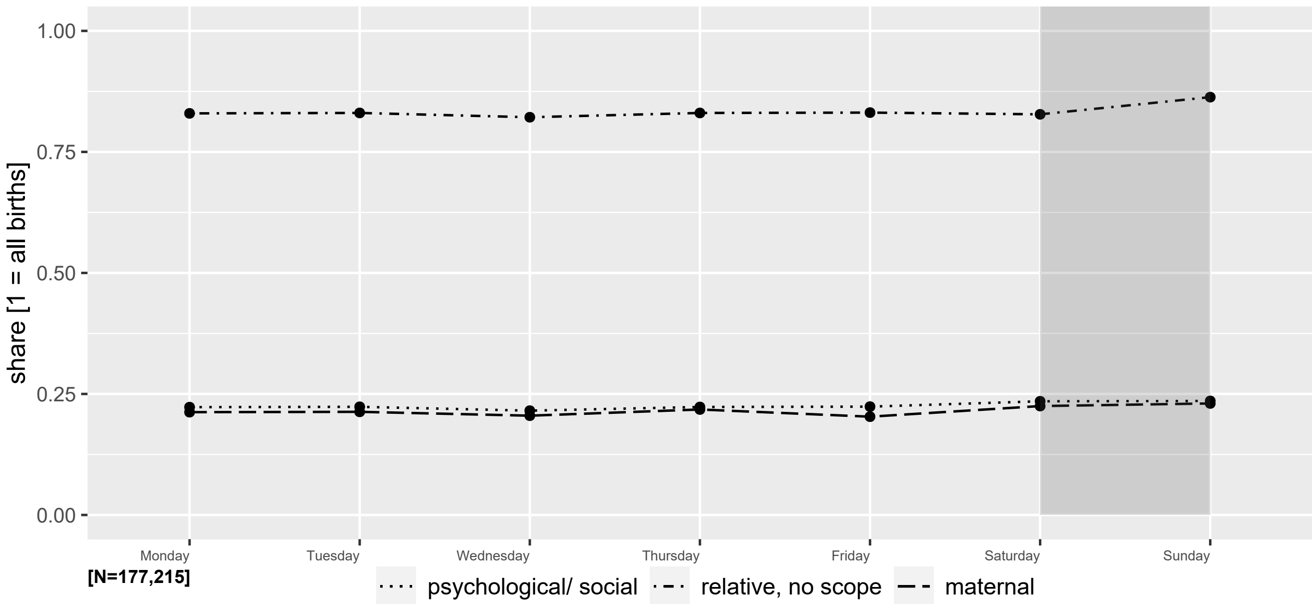


Source: IQTIG German hospital birth records for 2015-2016 restricted to the subsample of zero-precondition first-time mothers with pre-labor membrane ruptures (detailed in Table A.2 and Table A.3). Strata by maternal BMI. Own calculations.



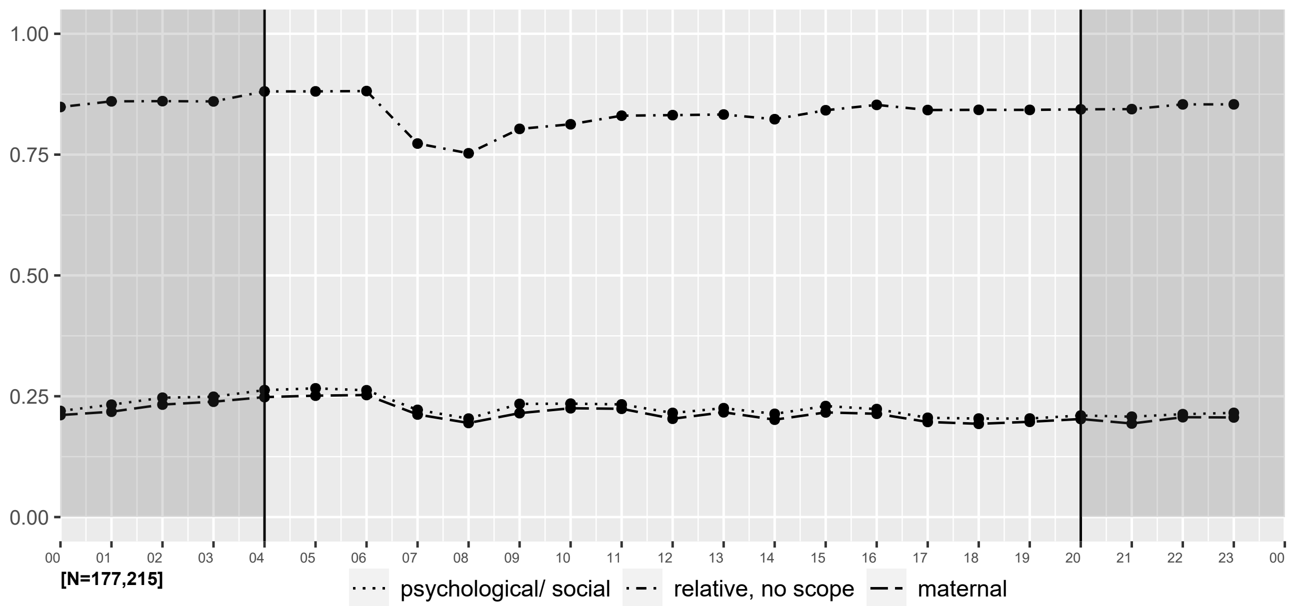
Source: IQTIG German hospital birth records for 2015-2016 restricted to the main analysis sample of zero-precondition first births (detailed in Table A.2 and Table A.3). $Birthdate - predicteddueddate = \Delta$. Own calculations.

Figure A.7: Distribution of Intervention Indications Across Weekdays

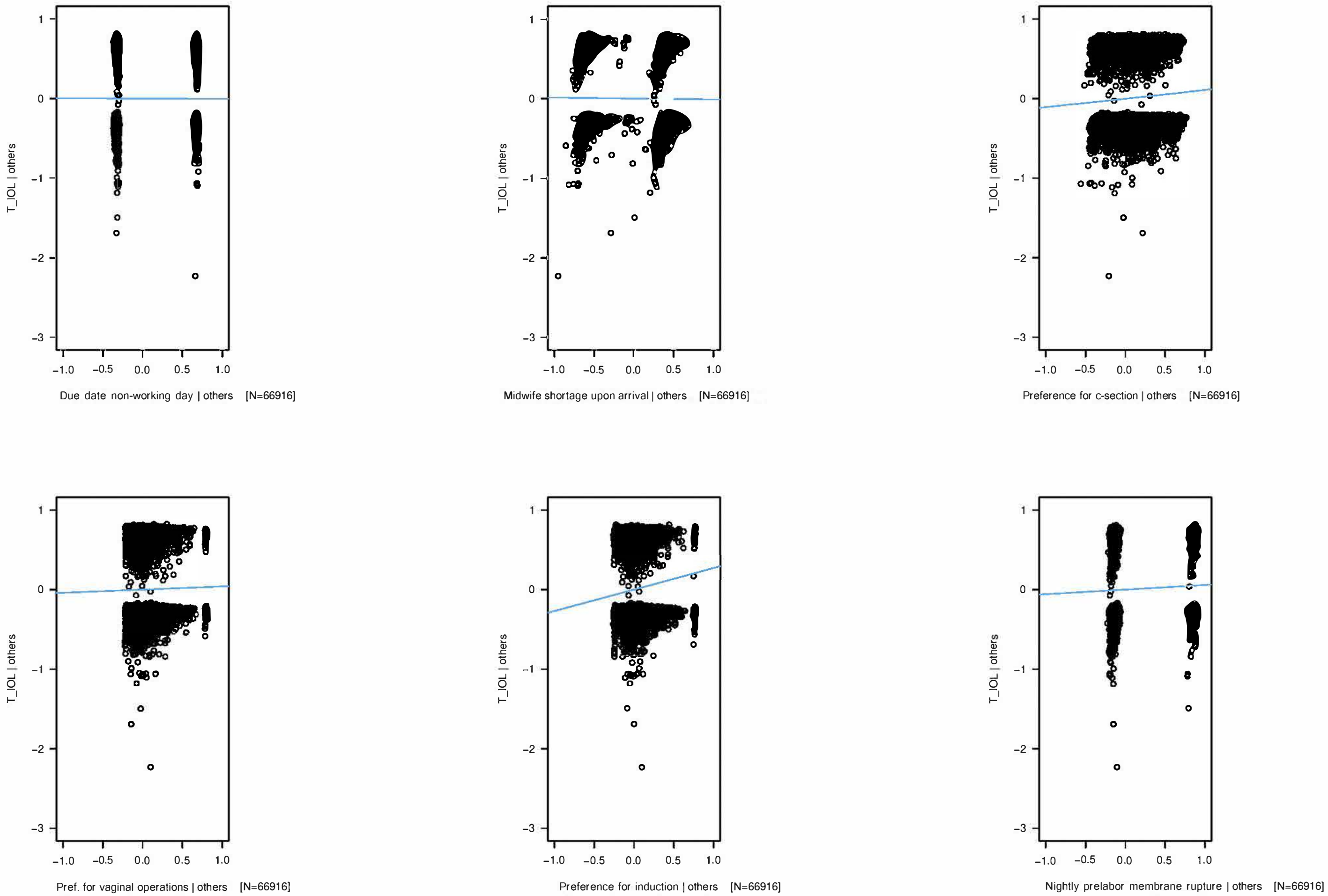


Source: IQTIG German hospital birth records for 2015-2016 restricted to the main analysis sample of zero-precondition first births (detailed in Table A.2 and Table A.3). Indications are grouped by implied medical decision scope for birth intervention, where *relative, no scope* comprise clearly stated medical conditions motivating (but not forcing) intervention. More vaguely defined are *psychological/social* conditions, and *maternal* refers to intervention on maternal request. Own calculations.

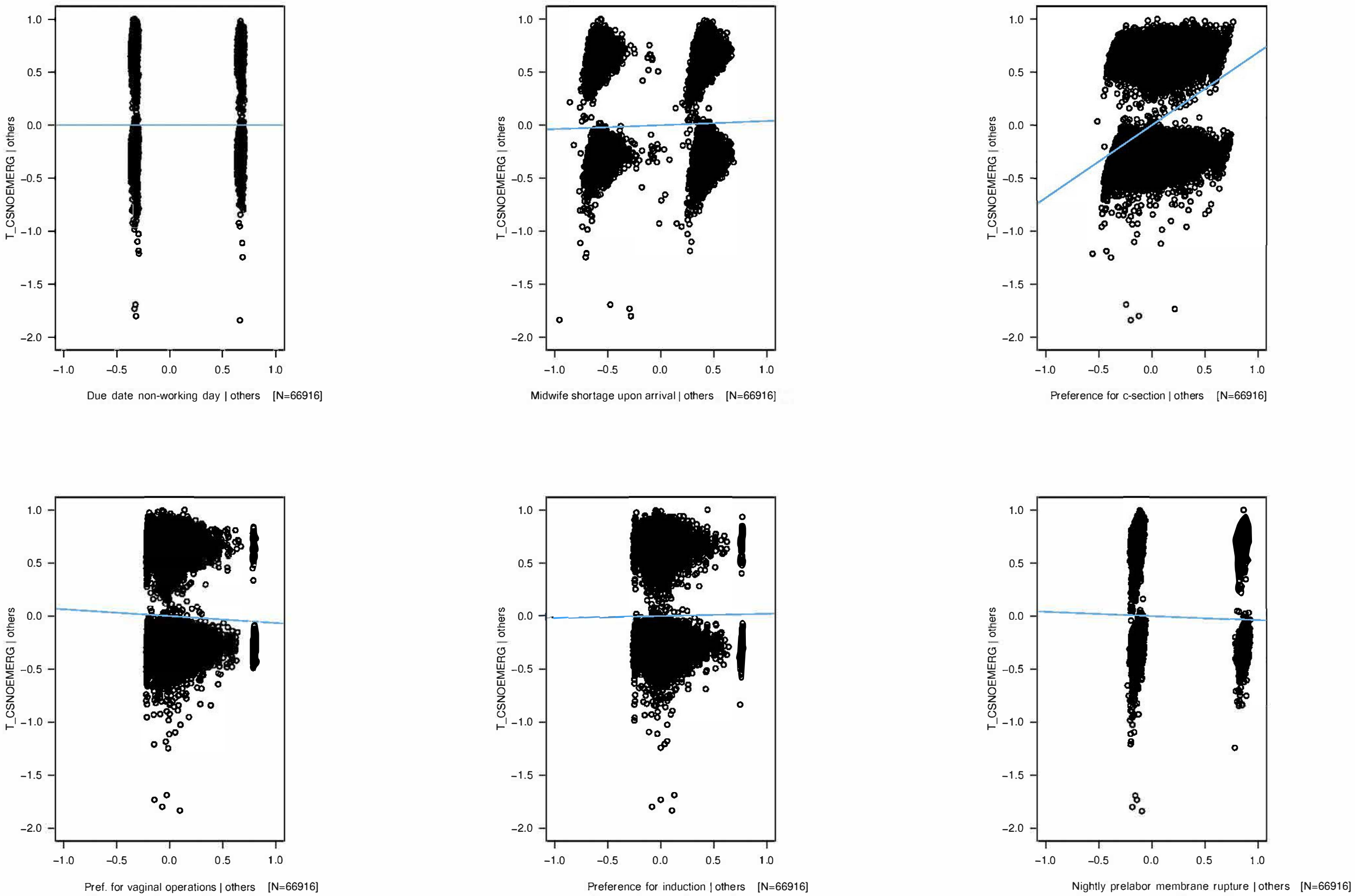
Figure A.8: Distribution of Intervention Indications Across Daily Hours



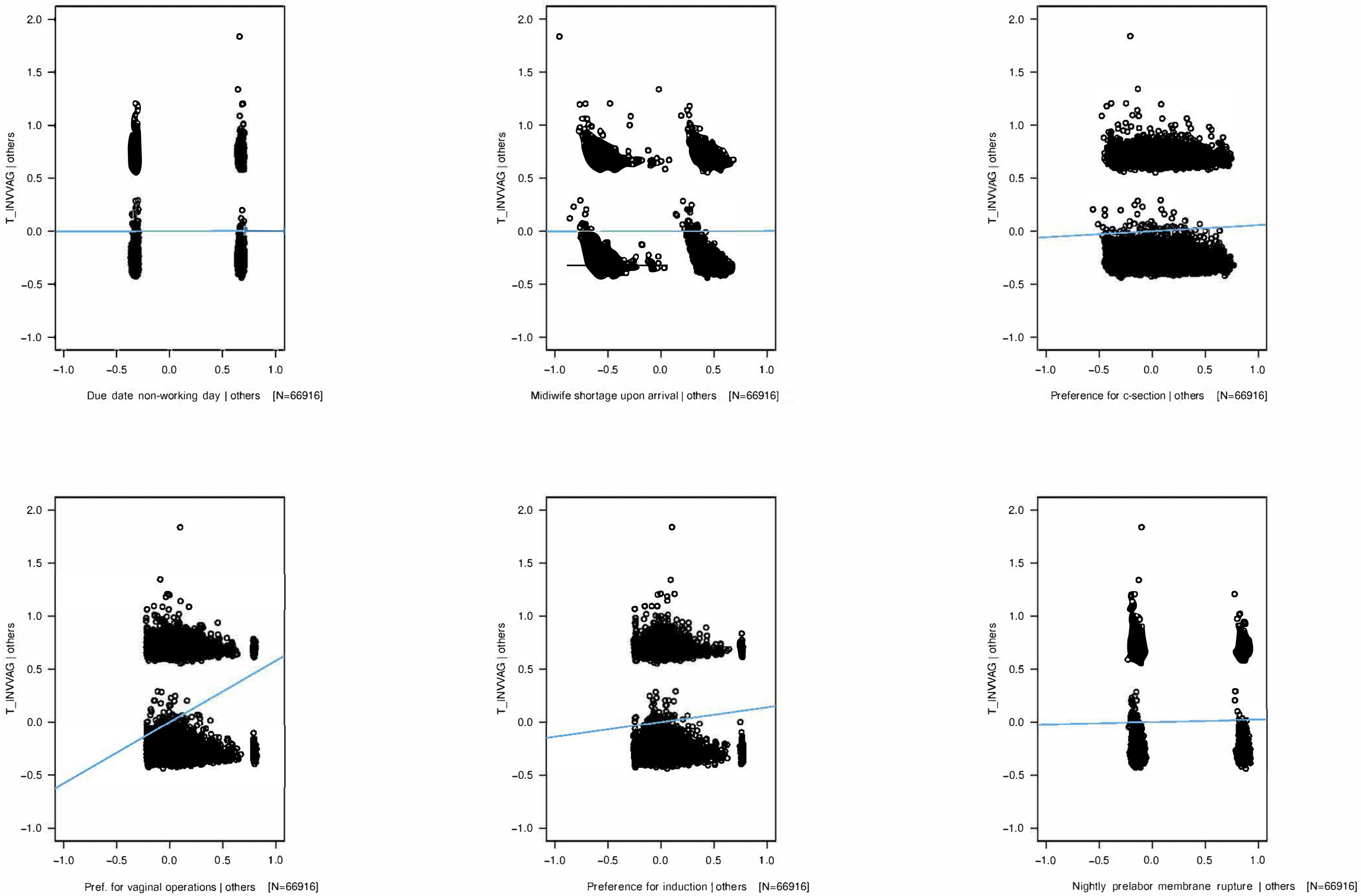
Source: IQTIG German hospital birth records for 2015-2016 restricted to the main analysis sample of zero-precondition first births (detailed in Table A.2 and Table A.3). Indications are grouped by implied medical decision scope for birth intervention, where *relative, no scope* comprise clearly stated medical conditions motivating (but not forcing) intervention. More vaguely defined are *psychological/social* conditions, and *maternal* refers to intervention on maternal request. Own calculations.



Source: IQTIG German hospital birth records, 2015-16, zero-precondition first births with information on obstetrician ids to compute obstetricians' preferences for intervention, see Table A.2 and Table A.3. Residual correlation conditional on core controls defined below Table 3. Own calculations.



Source: IQTIG German hospital birth records, 2015-16, zero-precondition first births with information on obstetrician ids to compute obstetricians' preferences for intervention, see Table A.2 and Table A.3. Residual correlation conditional on core controls defined below Table 3. Own calculations.



Source: IQTIG German hospital birth records, 2015-16, zero-precondition first births with information on obstetrician ids to compute obstetricians' preferences for intervention, see Table A.2 and Table A.3. Residual correlation conditional on core controls defined below Table 3. Own calculations.

References

Alberto Abadie and John F. Kennedy. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113(2):231–263, 2003.

Victoria M. Allen, Colleen M. O'Connell, and Thomas F. Baskett. Cumulative economic implications of initial method of delivery. *Obstetrics and Gynecology*, 108(3):549–555, 2006.

Douglas Almond, Kenneth Y. Chay, and David S. Lee. The Costs of Low Birth Weight. *The Quarterly Journal of Economics*, 120(3):1031–1083, 2005.

Isaiah Andrews and James H. Stock. Weak Instruments and What To Do About Them. Technical report, NBER Summer Institute, 2018.

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Number 8769. Princeton University Press, 1 edition, 2009.

R. Axt-Fliedner, U. Wiegank, M. Friedrich, and K. Diedrich. Elektive Einleitung gegenüber spontanem Geburtsbeginn am Termin. *Gynäkologe*, 37(4):346–351, 2004.

Manudeep Bhuller, Gordon B Dahl, Katrine V Løken, and Magne Mogstad. Incarceration, Recidivism, and Employment. *Journal of Political Economy*, 128(4), 2020.

Carrie F. Bonsack, Anthony Lathrop, and Mary Blackburn. Induction of labor: Update and review. *Journal of Midwifery and Women's Health*, 59(6):606–615, 2014.

Heather M. Bradford, Vicky Cárdenas, Katherine Camacho-Carr, and Mona T. Lydon-Rochelle. Accuracy of birth certificate and hospital discharge data: A certified nurse-midwife and physician comparison. *Maternal and Child Health Journal*, 11(6):540–548, 2007.

Alexandra Bruns. Das deutsche DRG-System: Die pauschale Geburt. *Deutsches Ärzteblatt*, 7 (3):2014, 2014.

Alexandra Bruns. Analyse zum Änderungsvorschlag „Kostenstelle Kreißsaal". Technical report, Geburt e.V., Kronshagen, 2017.

Kasey Buckles and Melanie Guldi. Worth the Wait? The Effect of Early Term Birth on Maternal and Infant Health. *Journal of Policy Analysis and Management*, 36(4):748–772, 2017.

BZgA. „Die Frauen können es, man lässt sie nur nicht!" Ein Gespräch mit Professor Alfred Rockenschaub über 50 Jahre Geburtshilfe und die wichtige Rolle der Frau. *FORUM Sexualaufklärung und Familienplanung*, 211(2):382–383, 2005.

A. Colin Cameron and Douglas L. Miller. A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–372, 2015.

David Card, David Silver, and Alessandra Fenizia David. The Health Effects of Cesarean Delivery for low-risk first births. *NBER Working Paper Series*, 106(11):1323–1330, 2018.

Suzan L. Carmichael and Jonathan M. Snowden. The ARRIVE Trial – Interpretation from an Epidemiologic Perspective. *Journal of Midwifery & Women's Health.*, 64(5):657–663, 2019.

Sarah Carter, Amos Channon, and Ann Berrington. Socioeconomic risk factors for labour induction in the United Kingdom. *BMC Pregnancy and Childbirth*, 20(146):284–300, 2020.

Victor Chernozhukov, Christian Hansen, and Michael Jansson. Admissible Invariant Similar Tests for Instrumental Variables Regression. *Econometric Theory*, 25(3):806–818, 2009.

Dominiek Coatesid, Angela Makris, Christine Catling, Amanda Henry, Vanessa Scarf, Nicole Watts, Deborah Fox, Purshaiyna Thirukumar, Vincent Wong, Hamish Russell, and Caroline Homer. A systematic scoping review of clinical indications for induction of labour. *Plos One*, 15(1):1–11, 2020.

Ana Costa-Ramón, Mika Kortelainen, Ana Rodríguez-González, and Lauri Sääksvuori. The Long-Run Effects of Cesarean Sections. *VATT Working Papers*, 125(1):1–48, 2019.

Ana María Costa-Ramón, Ana Rodríguez-González, Miquel Serra-Burriel, and Carlos Campillo-Artero. It's about time: Cesarean sections and neonatal health. *Journal of Health Economics*, 59(1):46–59, 2018.

DGGG. Empfehlungen zum Vorgehen beim vorzeitigen Blasensprung S1 Leitlinie. Technical report, Deutsche Gesellschaft für Gynäkologie und Geburtshilfe, Göttingen, 2006.

DGGG. Geburtseinleitung Leitlinienklasse S2k Stand Dezember 2020. Technical report, Deutsche Gesellschaft für Gynäkologie und Geburtshilfe, Göttingen, 2020a.

DGGG. Sectio caesarea Leitlinienklasse S3 Stand Juni 2020. Technical report, Deutsche Gesellschaft für Gynäkologie und Geburtshilfe, Göttingen, 2020b.

DGGG. Leitlinie zum Management von Dammrissen III. und IV. Grades nach vaginaler Geburt S2k Stand Dezember 2020. Technical report, Deutsche Gesellschaft für Gynäkologie und Geburtshilfe, Göttingen, 2020c.

DHV and DGGG. DHV und DGGG fordern Korrekturen am Gesetzentwurf für die 1:1-Betreuung in der klinischen Geburtshilfe. Technical report, Deutscher Hebammen Verband und Deutsche Gesellschaft für Gynäkologie und Geburtshilfe, Berlin, 2020.

David Dranove and Paul Wehner. Physician-induced demand for childbirths. *Journal of Health Economics*, 13(1):61–73, 1994.

A. Feige. Budget und organisationsaspekte einer geburtshilflichen abteilung. *Gynäkologe*, 41(1): 42–48, 2008.

Joshua S. Gans and Andrew Leigh. Born on the first of July: An (un)natural experiment in birth timing. *Journal of Public Economics*, 1(1):1–18, 2008.

Susan Garthus-Niegel, Cecilie Knoph, Tilmann von Soest, Christopher S. Nielsen, and Malin Eberhard-Gran. The Role of Labor Pain and Overall Birth Experience in the Development of Posttraumatic Stress Symptoms: A Longitudinal Cohort Study. *Birth*, 41(1):108–115, 2014.

Ilka Gerhardts. The Economics of Labor & Patients' Health Outcomes: Evidence from Childbirth in Germany. *mimeo, University of Munich (LMU)*, 2024.

Salvatore Gizzo, Carlo Saccardi, Tito Silvio Patrelli, Stefania Di Gangi, Elisa Breda, Simone Fagherazzi, Marco Noventa, Donato D'antona, and Giovanni Battista Nardelli. Fertility rate and subsequent pregnancy outcomes after conservative surgical techniques in postpartum hemorrhage: 15 years of literature. *Fertility and Sterility*, 99(1):2097–2107, 2013.

GNPI. Betreuung von Neugeborenen in der Geburtsklinik - S2k Leitlinie. *AWMF online*, 024-005: 1–25, 2022.

Henci Goer. Parsing the ARRIVE Trial: Should First-Time Parents Be Routinely Induced at 39 Weeks? Technical report, Lamaze International, Washington DC, 2018.

Karin Gottvall and Ulla Waldenström. Does a traumatic birth experience have an impact on future reproduction? *BJOG: An International Journal of Obstetrics and Gynaecology*, 109(3): 254–260, 2002.

William A. Grobman, Madeline M. Rice, Uma M. Reddy, Alan T.N. Tita, Robert M. Silver, Gail Mallett, Kim Hill, Elizabeth A. Thom, Yasser Y. El-Sayed, Annette Perez-Delboy, Dwight J. Rouse, M.D. George R. Saade, Kim A. Boggess, Suneet P. Chauhan, Jay D. Iams, Edward K. Chien, Brian M. Casey, Ronald S. Gibbs, Sindhu K. Srinivas, Geeta K. Swamy, Hyagriv N. Simhan, and George A. Macones. Labor Induction versus Expectant Management in Low-Risk Nulliparous Women. *The NEW ENGLAND JOURNAL of MEDICINE*, 379(6):513–523, 2018.

İsmet Gün, Bülent Doğan, and Özkan Özdamar. Long- and short-term complications of episiotomy. *Turkish Journal of Obstetrics and Gynecology*, 13(2):144–148, 2016.

Martin Halla, Harald Mayr, Gerald J. Pruckner, and Pilar García-Gómez. Cutting fertility? Effects of cesarean deliveries on subsequent fertility and maternal labor supply. *Journal of Health Economics*, 72(1):102–125, 2020.

Lorie M. Harper, Aaron B. Caughey, Anthony O. Odibo, Kimberly A. Roehl, Qiuhong Zhao, and Alison G. Cahill. Normal progress of induced labor. *Obstetrics and Gynecology*, 321(3): 25–31, 2012.

InEK. aG-DRG-Version und Pflegeerlöskatalog 2021. Technical report, InEK Institut für das Entgeltsystem im Krankenhaus, Siegburg, 2021.

IQTIG. Qualitätsreport 2016. Technical report, Institut für Qualitätssicherung und Transparenz im Gesundheitswesen (i.A. Gemeinsamer Bundesausschuss), Berlin, 2017.

Christine Jachetta. *Topics in Health Economics*. PhD thesis, University of Illinois at Urbana-Champaign, 2016.

Anne Katz Jacobson. Are we losing the art of midwifery? *Journal of Nurse-Midwifery*, 38(3): 168–169, 1993.

Mireille Jacobson, Maria Kogelnik, and Heather Royer. Holiday, Just One Day Out of Life: Birth Timing and Post-Natal Outcomes. *NBER Working Paper Series*, 27236, 2020.

Wonjeong Jeong, Sung In Jang, Eun Cheol Park, and Jin Young Nam. The effect of socioeconomic status on all-cause maternal mortality: A nationwide population-based cohort study. *International Journal of Environmental Research and Public Health*, 17(12):1–13, 2020.

John Jolly, James Walker, and Kalvinder Bhabra. Subsequent obstetric performance related to primary mode of delivery. *BJOG: An International Journal of Obstetrics and Gynaecology*, 106(3):227–232, 1999.

Hendrik Jürges. Financial incentives, timing of births, and infant health: a closer look into the delivery room. *European Journal of Health Economics*, 18(1):195–208, 2017.

Hendrik Jürges and Juliane Köberlein. What explains DRG upcoding in neonatology? The roles of financial incentives and infant health. *Journal of Health Economics*, 43(2):13–26, 2015.

KBV. Einheitlicher Bewertungsmaßstab (EBM). 01:1–1779, 2020.

Petra Kolip, Hans-Dieter Nolting, and Karsten Zich. Kaiserschnittgeburten – Entwicklung und regionale Verteilung. Technical report, Bertelsmann Stiftung, Gütersloh, 2012.

Lisa D. Levine, Katheryne L. Downes, Michal A. Elovitz, Samuel Parry, Mary D. Sammel, and Sindhu K. Srinivas. Mechanical and Pharmacologic Methods of Labor Induction: A Randomized Controlled Trial. *Obstetrics & Gynecology*, 128(6):1357–1364, 2016.

Ulrike Lutz and Petra Kolip. Die GEK-Kaiserschnittstudie. In GEK-Gmünder-Ersatzkasse, editor, *Schriftenreihe zur Gesundheitsanalyse*. Asgard Verlag, Bremen/ Schwäbisch Gmünd, 2006.

Mona Lydon-Rochelle, Victoria L. Holt, Diane P. Martin, and Thomas R. Easterling. Association Between Method of Delivery and Maternal Rehospitalization. *JAMA*, 283(18):2411–2416, 2000.

John Lynch, Aurélie Meunier, Rhiannon Pilkington, and Stefanie Schurer. Baby Bonuses and Early-Life Health Outcomes: Using Regression Discontinuity to Evaluate the Causal Impact of an Unconditional Cash Transfer. *IZA Discussion Papers*, 12230(1):1–54, 2019.

Nicole Maestas, Kathleen J. Mullen, and Alexander Strand. Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI Receipt. *American Economic Review*, 103(5):1797–1829, 2013.

Ekaterina Mishanina, Ewelina Rogozinska, Tej Thatthi, Rehan Uddin-Khan, Khalid S. Khan, and Catherine Meads. Use of labour induction and risk of cesarean delivery: a systematic review and meta-analysis. *CMAJ*, 186(9):665–673, 2014.

Model Professional Code for Physicians in Germany. MBO-Ä 1997-**The Resolutions of the 121st German Medical Assembly 2018 in Erfurt as amended by a Resolution of the Executive Board of the German Medical Association on 14/12/2018, I. Principles Art.1 (1), 1997.

Ioannis Mylonas and Klaus Friese. The indications for and risks of elective cesarean section. *Deutsches Ärzteblatt International*, 112(29-30):489–495, 2015.

Karen Norberg and Juan Pantano. Cesarean sections and subsequent fertility. *Journal of Population Economics*, 29(1):5–37, 2016.

Charles O'Donovan and James O'Donovan. Why do women request an elective cesarean delivery for non-medical reasons? A systematic review of the qualitative literature. *Birth*, 45(2):109–119, 2018.

Vicky O'Dwyer, Sarah O'Kelly, Bernadette Monaghan, Ann Rowan, Nadine Farah, and Michael J. Turner. Maternal obesity and induction of labor. *Acta Obstetricia et Gynecologica Scandinavica*, 92(12):1414–1418, 2013.

Man Wah Pang, Tse Ngong Leung, Tze Kin Lau, and Tony Kwok Hang Chung. Impact of first childbirth on changes in women's preference for mode of delivery: Follow-up of a longitudinal observational study. *Birth*, 35(2):121–128, 2008.

Holly Priddis, Hannah G. Dahlen, Virginia Schmied, Annie Sneddon, Christine Kettle, Chris Brown, and Charlene Thornton. Risk of Recurrence, Subsequent Mode of Birth and Morbidity for Women who Experienced Severe Perineal Trauma in a First Birth in New South Wales between 2000–2008: A Population Based Data Linkage Study. *BMC Pregnancy and Childbirth*, 89(13):1471–1493, 2013.

Sandra Ilona Rummel. *Kosten und Erlöse bei der Abrechnung geburtshilflicher Leistungen nach dem System der Diagnosis-Related-Groups (DRG)*. PhD thesis, LMU Munich, 2007.

Elsa Lena Ryding, Mirjam Lukasse, Hildur Kristjansdottir, Thora Steingrimsdottir, Berit Schei, Ann Tabor, Helle Karro, An Sofie Van Parys, Anne Mette Schroll, Made Laanpere, and Anne Marie Wangel. Pregnant women's preference for cesarean section and subsequent mode of birth – a six-country cohort study. *Journal of Psychosomatic Obstetrics and Gynecology*, 37(3):75–83, 2016.

Eleanor Sanderson and Frank Windmeijer. A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2):212–221, 2016.

Anton Scharl, Frank Louwen, and Christian Albring. Offener Brief anlässlich der Öffentlichen Anhörung im Ausschuss für Gesundheit des Deutschen Bundestages am Mittwoch, 26. Juni 2019, zum Thema „Entwurf eines Gesetzes zur Reform der Hebammenausbildung". Technical report, German Board and College of Obstetrics and Gynecology, Berlin, 2019.

Lisa Schulkind and Teny Maghakian Shapiro. What a difference a day makes: Quantifying the effects of birth timing manipulation on infant health. *Journal of Health Economics*, 33(1): 139–158, 2014a.

Lisa Schulkind and Teny Maghakian Shapiro. What a difference a day makes: Quantifying the effects of birth timing manipulation on infant health. *Journal of Health Economics*, 33(1): 139–158, 2014b.

Christiane Schwarz, Rainhild Schäfers, Christine Loytved, Peter Heusser, Michael Abou-Dakn, Thomas König, and Bettina Berger. Temporal trends in fetal mortality at and beyond term and induction of labor in Germany 2005–2012: data from German routine perinatal monitoring. *Archives of Gynecology and Obstetrics*, 293(2):335–343, 2016.

Andrei Smarandache, Theresa H.M. Kim, Yvonne Bohr, and Hala Tamim. Predictors of a negative labour and birth experience based on a national survey of Canadian women. *BMC Pregnancy and Childbirth*, 16(1):903–920, 2016.

Hege Therese Størksen, Susan Garthus-Niegel, Samantha S. Adams, Siri Vangen, and Malin Eberhard-Gran. Fear of childbirth and elective caesarean section: A population-based study. *BMC Pregnancy and Childbirth*, 15(1):1–11, 2015.

A Tammaa, W Umek, and M. Wunderlich. Leitlinie zum Management von Dammrissen III. und IV. Grades nach vaginaler Geburt. *Speculum - Zeitschrift für Gynäkologie und Geburtshilfe*, 25(3):15–15, 2007.

Sibil Tschudin, Judith Alder, Stephanie Hendriksen, Johannes Bitzer, Karoline Aebi Popp, Rosanna Zanetti, Irene Hösli, Wolfgang Holzgreve, and Verena Geissbühler. Previous birth experience and birth anxiety: Predictors of caesarean section on demand? *Journal of Psychosomatic Obstetrics and Gynecology*, 30(3):175–180, 2009.

Peter Tyson. The Hippocratic Oath Today. *NOVA*, 2001.

Anjel Vahratian, Jun Zhang, James F. Troendle, Anthony C. Sciscione, and Matthew K. Hoffman. Labor progression and risk of cesarean delivery in electively induced nulliparas. *Obstetrics and Gynecology*, 105(4):698–704, 2005.

Francis P.J.M. Vrouenraets, Frans J.M.E. Roumen, Cary J.G. Dehing, Eline S.A. Van Den Akker, Maureen J.B. Aarts, and Esther J.T. Scheve. Bishop score and risk of cesarean delivery after induction of labor in nulliparous women. *Obstetrics and Gynecology*, 105(4):690–697, 2005.

WHO. WHO Statement on Caesarean Section Rates. Technical report, World Health Organisation, Geneva, 2015.

Katherine B. Wolfe, Rocco A. Rossi, and Carri R. Warshak. The effect of maternal obesity on the rate of failed induction of labor. *American Journal of Obstetrics and Gynecology*, 205(2): 128.e1–128.e7, 2011.

Andrea Mary Woolner, Dolapo Ayansina, Mairead Black, and Sohinee Bhattacharya. The Impact of Third- or Fourth-Degree Perineal Tears on the Second Pregnancy: A cohort study of 182,445 Scottish Women. *PLoS One*, 14(4):1–18, 2019.

C. M. Zahn and E. R. Yeomans. Postpartum hemorrhage: Placenta accreta, uterine inversion, and puerperal hematomas. *Clinical Obstetrics and Gynecology*, 33(3):422–431, 1990.