The Effects of Universal Screening for Gestational Diabetes on Maternal and Child Health

Anton Barabasch* Kamila Cygan-Rehm† Anica Kramer‡

This version February 27, 2025

Preliminary results – please do not cite or circulate.

Abstract

This paper estimates the effects of universal screening for gestational diabetes mellitus (GDM) on maternal and infant health at birth. GDM is the leading cause of excessive fetal growth and can have adverse long-term consequences for both mother and child. We evaluate a policy that introduced a full reimbursement for an oral GDM test by the German Statutory Health Insurance in July 2013, which led to a sharp increase in screening rates among pregnant women by almost 25 percentage points. Applying a difference-in-discontinuities design to administrative data on all hospital births, we find no effects of universal GDM screening on neonatal health and maternal birth outcomes.

Keywords: Gestational diabetes, screening, prenatal care, child health, maternal health, birth outcomes
JEL Codes: I1, I13, I18

*Friedrich-Alexander University Erlangen-Nuremberg (FAU) and Dresden University of Technology (TUD); email: anton.barabasch@fau.de

†**Corresponding author.** Dresden University of Technology (TUD), CESifo, LIfBi Bamberg, and IZA; email: kamila.cygan-rehm@tu-dresden.de

‡Zentralinstitut für die kassenärztliche Versorgung (Zi Berlin), IZA, and RWI; email: AKramer@zi.de

1 Introduction

Gestational diabetes mellitus (GDM) is one of the most common medical complications of pregnancy and a leading cause of excessive fetal growth, i.e., fetal macrosomia (e.g., Lappe et al., 2023).[1] Although maternal blood glucose levels typically return to normal levels after delivery, GDM is associated with adverse perinatal outcomes such as cesarean section, obstructed labor due to shoulder dystocia, or birth injury (e.g., Metzger et al., 2008, McIntyre et al., 2019). In addition, extensive medical literature documents persistent associations between GDM and serious health problems for both mother and child later in life.[2]

In recent decades, GDM has become a growing public health concern due to its increasing prevalence in many parts of the world. While in many countries the prevalence of GDM was negligible (of about 1%) in the 1990s, current estimates suggest that nearly 17% of pregnant women worldwide are affected (e.g., Wang et al., 2022).[3] This upward trend is expected to continue as maternal age at birth and obesity rates increase, two major risk factors for GDM (Cleary-Goldman et al., 2005, Shah et al., 2021). As a result, many countries and health agencies around the world have adjusted their prenatal care guidelines by establishing screening and diagnostic standards. However, there is no consensus on the use of universal versus risk-based screening due to the lack of evidence on its potential impact on child development and maternal health.

This paper examines how universal access to free screening for GDM affects maternal and infant health, taking advantage of the introduction of full reimbursement for an oral GDM test by the German Statutory Health Insurance system in July 2013 (see e.g., Tamayo et al., 2016). We estimate the intention-to-treat effects of this policy by combining high-quality administrative data on all hospital births between 2008 and 2013 with a difference-in-discontinuities design that exploits quasi-random variation in eligibility for free GDM screening among expectant mothers in 2013. Specifically, we compare the birth outcomes of mothers who had just benefited from eligibility for free GDM screening to those who were not yet eligible, with cohorts who gave birth in exactly the same months but before 2013. We use a similar design to estimate the effects on newborn outcomes. Using complementary data from process-generated claims from a

[1] Fetal macrosomia refers to infants with excessive birth weight, typically defined as 4,000g or 4,500g and above (Gaudet et al., 2014).

[2] For example, affected mothers have a significantly higher lifetime risk of developing type 2 diabetes and cardiovascular disease (e.g., Bellamy et al., 2009, Farrar et al., 2016, Kramer et al., 2019). Their offspring has an increased risk of sudden infant death syndrome, overweight, and obesity with concomitant cardiometabolic disorders across the lifespan (e.g., Zhang et al., 2008, Zhu et al., 2016, Yu et al., 2019).

[3] In general, cross-country and over-time comparisons of GDM prevalence are challenging due to differences in screening approaches and changes in diagnostic criteria (Wang et al., 2022, Eades et al., 2024).

large health insurer, we also provide evidence of the immediate effect on actual GDM screening utilization among pregnant women.

We find that the introduction of a free GDM screening led to a sharp increase in screening rates among pregnant women by almost 25 percentage points (or 34%-38% compared to sample mean). However, despite the large effects on prenatal GDM screening rates, we find only a moderate (albeit statistically significant) increase of about 4% in the number of newborns affected by maternal GDM. The estimated effects on other neonatal health measures and maternal birth outcomes, such as birth weight, macrosomia, cesarean section, and shoulder dystocia, are virtually zero. As a result, we find no meaningful changes in length of stay and hospital reimbursement claims, which we interpret as proxies for economic costs to statutory health insurers.

Our estimates of zero effects on birth outcomes are very precise and robust. The point estimates and statistical significance are not affected by the choice of model specification or sample definitions. We also show that neither potential imbalances in the composition of the treatment and control groups nor endogenous sorting of mothers across the policy threshold explain our findings. In terms of mechanisms, our results suggest that even before the introduction of universal GDM screening in 2013, physicians were able to identify mothers most at risk for adverse GDM consequences at birth using a risk-based assessment.

In the future, we plan to examine whether the zero effects on immediate birth outcomes persist using process-generated claims data from all statutory health insurers in Germany, which include information on all drug dispensations, outpatient diagnoses, and services up to ten years after treatment. This will allow us to better understand whether the screening itself induced any potential changes in maternal behavior (e.g., changes in diet and physical activity) that may have long-term effects on the health of children and mothers. In addition, these data include information on maternal socioeconomic background and pre-pregnancy health, which will allow us to conduct important heterogeneity analyses. Previous research suggests that extended prenatal care benefits the most disadvantaged groups of mothers (e.g., Corman et al., 2019).

This paper contributes to several strands of the literature. First, our work is closely related to a scarce body of research that examines the effects of GDM screening on birth outcomes using quasi-experimental designs. For Finland, Riukula (2023) uses a regression discontinuity design (RDD) that exploits the discontinuous increase in screening rates for young mothers at the overweight threshold.[4] However, due to the small sample size, the estimated effects on birth outcomes are very imprecise and the confidence

[4]In Finland, first-time mothers under the age of 25 are screened only if their body mass index (BMI) exceeds the overweight threshold of 25 (Riukula, 2023).

intervals can only rule out large effects of about 10-20%. In ongoing work, Conti and Rodriguez-Lesmes (2021) apply RDD to a detailed but relatively small dataset from the Born-in-Bradford cohort study.[5] Exploiting a discontinuity along the blood glucose concentration at a specific GDM diagnostic threshold, they find a significant reduction in the odds of macrosomia among infants above the GDM threshold. However, the results for postnatal infant health and development are mixed. The authors point to an urgent need for studies of the medium- and long-term effects of GDM screening.

A distinct feature of our work is that we exploit policy-induced variation in universal eligibility for free GDM screening, which allows us to move away from estimating a local average treatment effect of a marginal GDM diagnosis at a given BMI or blood test threshold. Instead, we focus on the intention-to-treat effect of universal GDM screening, thereby adding rigorous empirical evidence to the debate on the effectiveness and consequences of universal versus risk-based GDM detection (e.g., Farrar et al., 2016), which has broad policy implications.[6] Specifically, our difference-in-differences design, applied to large data including nearly 2 million births, yields precisely estimated zero effects on maternal and infant health outcomes at birth. In the future, we will also examine the long-term consequences for child development and maternal health.

Second, we build on the broader literature that examines the short- and long-run consequences of prenatal care.[7] Most studies within this literature exploit sources of exogenous variation in access to prenatal care (e.g., Joyce, 1999, Currie and Grogger, 2002, Evans and Lien, 2005) and find modest reductions in poor birth outcomes such as preterm birth, low birth weight, and infant mortality. Much less attention has been paid to the role of monetary incentives, and so far, the results are inconclusive (e.g., Cygan-Rehm and Karbownik, 2022, Di Giacomo et al., 2022).[8] In general, the literature focuses on prenatal care as a means of preventing poor fetal growth and, consequently, low birth weight. We complement this research by examining the effects of a policy that removed financial barriers to prenatal care targeted at the upper tail of the birth weight distribution, i.e., preventing and mitigating excessive fetal growth and its later

[5]The data cover all pregnancies at the main hospital in this English city between 2007 and 2011, and include rich information on blood tests, in-utero growth, and anthropometrics for the first five years of life. Depending on the exact outcome, the samples include roughly between 300 and 2500 individuals.

[6]Compared with selective screening, universal screening requires more resources but may detect more cases of GDM. There is also no consensus regarding the optimal testing strategy such as one-step versus two-step approach (Farrar et al., 2017).

[7]For reviews, see, e.g., Almond and Currie (2011), Currie and Rossin-Slater (2015), Corman et al. (2019).

[8]For example, Cygan-Rehm and Karbownik (2022) find that a Polish reform that incentivized early initiation of antenatal care had modest positive effects on neonatal health due to improved maternal health knowledge and behavior during pregnancy. However, Di Giacomo et al. (2022) show that an Italian policy that eliminated co-payments for noninvasive screening tests in Italy did not affect newborn health, despite increased screening participation and positive effects on maternal health behaviors.

consequences.

Finally, our work connects to the literature evaluating the effectiveness of universal screening for other common diseases such as hypertension, type 2 diabetes, and breast and cervical cancer (e.g., Sabik and Bradley, 2016, Guthmuller et al., 2023, Alalouf et al., 2024). Extensive research supports a widespread use of medical screening an an effective tool promoting the use of preventive health care, which can save lives and induce favorable health behaviors at relatively low cost (e.g., Maciosek et al., 2010). However, universal screening recommendations remain a highly controversial issue because of the cost of treating marginal patients for whom the expected benefit is negligible and the potential harm from unnecessary procedures and psychological distress (e.g., Einav et al., 2020).[9]

We add to this literature by providing evidence of no effect of universal GDM screening on birth outcomes despite slightly increased detection rates. In this respect, our results support the view that in a high-quality health care system, risk-based assessment is sufficient to identify mothers most at risk for adverse GDM outcomes, and the benefits for marginally diagnosed mothers and their offspring are limited. However, we acknowledge that these conclusions are currently based only on the immediate effects on birth outcomes and that a longer-term perspective is necessary.

This paper proceeds as follows. Section 2 provides institutional details. Section 3 describes the empirical strategy and Section 4 the data. Section 5 presents our main results and robustness tests. Section 6 concludes.

2 Institutional background

In Germany, health insurance is mandatory and is provided through both statutory and private insurance systems. Nearly 90% of the population is covered by statutory health insurance (SHI), while individuals whose income exceeds a certain threshold or who belong to a certain occupational groups (e.g. self-employed or civil servants) must or may choose to enroll in private health insurance (PHI) for substitutive full coverage. SHI covers a wide range of medical services that go well beyond basic health care, and the services are the same for all enrollees with a given SHI provider. Individuals covered by substitutive PHI typically enjoy coverage that is equal to or even better than within the SHI (for more details, see e.g., OECD, 2023).

Prenatal care is a central pillar of health care in Germany, and is regulated by na-

[9] Recent technological developments and the increasing use of artificial intelligence models for individualized risk assessment may help minimize over-diagnosis and unnecessary, invasive, and costly medical procedures (Eisemann et al., 2025).

tional maternity guidelines. These guidelines entitle all pregnant women covered by the SHI to access health counseling, support, and preventive measures free of charge (Vetter and Goeckenjan, 2013). However, cost-coverage is limited to medical services with proven effectiveness, as determined by the Federal Joint Committee (G-BA).[10] Services that lack sufficient evidence are excluded from SHI coverage but remain available as individual health services.[11] These services are typically provided after consultation with a physician and require patients to cover the costs on a self-pay basis (Schnell-Inderst et al., 2011). Historically, GDM screening was an individual health service. Thus, the decision to screen depended largely on a physician's assessment of underlying risk factors for GDM[12] and the patient's willingness to pay out of pocket.

In the early 2010s, the cost of GDM screening ranged from 10 to 25 euros (Beyerlein et al., 2016), and approximately 40% of pregnant women covered by SHI were screened for GDM (Lappe et al., 2023). To mitigate the potentially harmful effects of undiagnosed GDM, the G-BA decided to move from a risk-based to a universal GDM screening recommendation and add it to the list of services covered by SHI on March 3, 2012.[13] However, due to bureaucratic delays, the billing code for physician reimbursement was not issued until July 1, 2013. Until then, pregnant women were legally entitled to free GDM screening, but their physicians could not claim reimbursement. Not surprisingly, the initial inclusion of GDM screening in the maternity guidelines in March 2012 did not result in a noticeable increase in screening rates (see Figure 1). In contrast, the change in reimbursement procedure was eventually followed by a discontinuous jump in screening rates, which stabilized at a constant level of about 93% in the following years (Lappe et al., 2023).

In addition to coverage by SHI, the inclusion of GDM screening in the maternity guidelines also standardized the screening procedure. According to the recommendations, all pregnant women should be routinely tested for GDM between gestational weeks 24 and 28. The screening follows a two-step process, starting with a non-fasted glucose challenge test (GCT). If blood glucose levels exceed a specific threshold, an oral

[10] The G-BA is a decision-making body in the German health care system that determines which medical treatments, diagnostic procedures, and medications are reimbursed by the SHI.

[11] A particular medical service may also remain excluded from the list of services covered by the SHI if the costs disproportionately outweigh the benefits.

[12] Overweight, GDM, macrosomia, or complications in previous pregnancies (e.g. miscarriages or congenital malformations), and diagnosed diabetes mellitus in parents or siblings were considered risk factors for GDM (Haschka et al., 2022).

[13] Initial discussions about offering free GDM screening began in 2002, but were paused due to insufficient evidence of its benefits. Discussions resumed after a growing body of medical evidence demonstrated the effectiveness of screening and treatment for GDM. The decision to add free GDM screening to the list of services covered by the SHI was ultimately based on an expert report from the Institute for Quality and Efficiency in Health Care (IQWiG).

glucose tolerance test (oGTT) is performed the following day under fasting conditions (Tamayo et al., 2016). Women with a positive oGTT result are officially diagnosed with GDM, which is recorded in their maternity records.[14] Women diagnosed with GDM are advised to make dietary changes and increase physical activity while closely monitoring their blood glucose levels. If blood glucose levels cannot be adequately controlled by lifestyle changes, pharmacologic interventions (e.g., metformin or insulin) may be used as complementary treatment options (Reitzle et al., 2021).[15]

3 Empirical strategy

The introduction of the reimbursement code for universal GDM testing in July 2013 created a natural experiment, allowing us to compare birth outcomes of the first mothers eligible for free screening with those who were not yet eligible for the test. We define eligibility based on the month of birth and the recommendation that screening should be performed between 24 and 28 weeks of gestation. Thus, mothers with a due date in October 2013 were the first birth cohort eligible for free GDM screening.[16]

In the main analysis, we focus on births that occurred up to three months before and after the October 1, 2013 threshold. To eliminate possible seasonal effects, we additionally use births that occurred in previous years as a control group. For this purpose, we include five pre-treatment cohorts (i.e., from 2008 to 2012). This empirical strategy combines a discontinuity design with a difference-in-difference (DiD) approach.[17] Specifically, we estimate a linear model of the form:

$$Y_i = \alpha\, post_i + year_i'\, \beta + \delta\, (post_i \times treat_i) + X_i'\gamma + \varepsilon_i \tag{1}$$

where Y_i is an outcome of an individual Y_i (mother or child). We analyze several

[14] Glucose levels are measured three times: in the fasting state and one and two hours after the oGTT. A woman is diagnosed with GDM if one of these measurements exceeds a certain threshold (Tamayo et al., 2016).

[15] In Germany, approximately 20 to 30 percent of pregnant women with GDM require insulin treatment (Schäfer-Graf et al., 2020).

[16] Given the recommended interval between weeks 24 and 28, there is some fuzziness in treatment assignment using birth month, as some of the mothers who gave birth in the last weeks of September may have already been treated, and some of the mothers who gave birth in the first weeks of October 2013 may not have been treated yet. In Section 5.2, we address this issue by showing that our results are robust to a donut-hole type of regression specification that excludes mothers who gave birth close to October 1.

[17] This design is common in studies evaluating the effects of policies that create a cutoff date for eligibility (e.g., Dustmann and Schönberg, 2012, Schönberg and Ludsteck, 2014, Avdic and Karimi, 2018, Felfe et al., 2020, Borra et al., 2021, Cygan-Rehm and Karbownik, 2022). The main rationale for including earlier years as a control group is the evidence of significant seasonality patterns in births and their correlation with parental socioeconomic characteristics (e.g., Bobak and Gjonca, 2001, Buckles and Hungerman, 2013, Currie and Schwandt, 2013, Clarke et al., 2019).

outcomes that describe the birth outcomes of mothers and the neonatal health of children (see Section 4). The variable $post_i$ is an indicator that takes the value 1 for births that occurred in the calendar months October to December and 0 for the months July to September. It captures any seasonal differences in the outcomes of summer and fall births, as long as the seasonal effects are constant across years. The vector $Year$ contains five year of birth dummies for the years 2009 to 2013 (with 2008 as the reference year). The year fixed effects flexibly capture any year-specific differences in birth outcomes (e.g. due to institutional or economic factors). The binary variable $treat$ indicates 2013 births (i.e., the treatment cohort). The vector of covariates X_i contains individual socio-demographic characteristics (e.g., maternal age at birth, a child's gender, region of residence) and regional contextual variables (e.g., aggregated indices of health care quality and economic situation, such as unemployment rate or gross domestic product). The terms α, β, δ, and γ are the coefficients to be estimated and ε_i is an error term.

The estimate of the interaction term δ is of primary interest as it identifies the intention-to-treat (ITT) effect of eligibility for free GDM screening. The key assumption for an internally valid DiD estimate of δ is the parallel trends assumption, which here implies that the potential seasonal patterns are common across years. In other words, we assume that the potential seasonal differences in the outcomes of summer and fall births would have remained the same if free GDM screening had not been introduced. While this is inherently untestable, we provide various empirical exercises to support the plausibility of this assumption. For example, we show graphically that the seasonal patterns were very similar in the pre-treatment years. Second, we estimate the effects of a placebo "treatment" assuming the policy occurred one year later. Finally, we also show that our main results are robust to alternative choices of control years, supporting the argument that any potential seasonality appears to be constant over time.[18]

Another important assumption is that the treatment did not induce a different sorting of births beyond the threshold of October 1, 2013. Although universal GDM screening had already been included in the maternity guidelines in March 2012, the final introduction of the reimbursement code in July 2013 was rather unexpected. Moreover, given the low out-of-pocket costs for GDM even before the policy change, it is unlikely that parents would have postponed conception because of the change. However, our

[18] Using a similar DiD design, Cygan-Rehm (2016) argues that the assumption of parallel trends implies that a deliberate choice of control years is important. The immediately preceding cohorts are natural candidates for this role, but the question remains how many of them should be considered. On the one hand, including many pre-treatment cohorts increases the estimation sample and may imply efficiency gains. On the other hand, a small number of control years reduces the risk that the underlying seasonal effects have changed over time or that other policy changes may contaminate the control groups.

identification strategy would also be threatened if the free screening significantly affected gestational age and thus, changed the composition of mothers giving birth before and after October 1, 2013. To mitigate such concerns, we perform balancing tests for the predetermined characteristics and test for a different mass of births due to the treatment. Finally, we estimate a donut-hole type regression specification that excludes mothers who gave birth near the October 1, 2013 cutoff, which aims to eliminate any confounding effects of potential endogenous sorting across the cutoff.

4 Data and samples

For the main analysis, we use administrative data from hospital discharge records based on the Diagnosis-Related Groups Statistic (DRG) reimbursement claims.[19] The data cover all inpatient hospital stays in Germany, and our estimation samples include years between 2008 and 2014. In addition to the exact date and reason of admission and discharge, the data include comprehensive information on all treatments and diagnoses during a given hospital stay. Clinical procedures are coded according to the German classification of medical operations and procedures (OPS). Diagnoses are coded according to the International Statistical Classification of Diseases and Related Health Problems (ICD-10). With respect to patient-level information on sociodemographic characteristics, we observe a patient's age, gender, and place of residence (FDZ, 2019).

To identify deliveries, we use "birth" as the registered cause of admission. As the DRG data are structured by hospital case, it is not possible to directly link mothers to their children. We therefore construct two separate samples. For mothers, we use the date of delivery to restrict the sample and assign the treatment status. In the child sample, we use the child's date of birth, which is directly recorded in the data. By doing this, we are able to identify about 90-93% of newborn children and 96-97% of mothers when we compare the numbers in the DRG data with the corresponding figures from the official natality statistics (see Appendix Table A.1). For our DiD analysis, we restrict the samples to births that took place in the second half (i.e. July to December) of the years 2008 to 2013, which yields nearly two million births.

As for birth outcomes, we focus on a child's birth weight both as a continuous measure (in grams) and as indicators of fetal macrosomia (i.e. using the common thresholds of either 4,000 or 4,500 grams). We also include an indicator of whether a

[19] The data are collected by the Institute for the Hospital Remuneration System (InEK) from all virtually hospitals in Germany for accounting purposes, which is required by law. Not included are prison, police, and psychiatric hospitals. The InEK transmits a legally defined subset of variables to the Federal Statistical Office, which makes the data available for research purposes via on-site use at its Research Data Centers (FDZ). For more details on the data, see FDZ (2019).

newborn suffers from the syndrome of the mother with gestational diabetes (ICD-10 P70.0), which is typically used to identify newborns affected by maternal GDM for reimbursement purposes. We also use an indicator for low birth weight (i.e., less than 2500 grams) as a placebo outcome because we do not expect effects at the lower end of the birth weight distribution. In the maternal sample, we focus on cesarean delivery (OPS-Codes: 5-740, 5-741, 5-742, 5-749) and an indicator of obstructed labor due to shoulder dystocia (ICD-10 O66.0), which are the two most common birth complications associated with GDM, and pregnancy duration. Finally, as a proxy for the economic costs of childbirth, we also consider the length of stay and total claims submitted by hospitals to a patient's health insurer, which we observe for both infants and mothers.

Table 1 shows the sample means in our estimation samples. On average, 3.6% of newborns were identified at birth as affected by maternal GDM. This is lower than the average incidence rates and suggests a significant underestimation. The average birth weight is about 3.3 kg, and about 10% of newborns are affected by macrosomia when defined by the 4,000 g threshold (1.5% when defined by the 4,500 g threshold). On average, more than 30% of mothers deliver by cesarean section. Obstructed labor due to shoulder dystocia occurs in 1.5% of deliveries. Almost 10% of mothers give birth post-term and almost 7% pre-term. On average, infants stay in hospital slightly longer than mothers (4.8 days vs. 4.4 days), but generate lower costs to health insurers (€1,715 vs. €1,744).

Because DRG data do not include retrospective information on medical care and services during pregnancy, we use aggregated data from AOK Plus for complementary analyses of the actual utilization of GDM screening. The AOK is one of the largest statutory health insurers in Germany, providing comprehensive medical insurance. We obtained aggregate time series on GDM screening rates for one federal state from AOK PLUS Saxony, which is the largest insurer in Saxony, covering 3.5 million people, giving it a market share of nearly 60 percent. The monthly screening rates were calculated using mothers who gave a hospital birth between January 2010 and December 2014.[20] The data are restricted to mothers who were insured with AOK Plus for at least 267 days (i.e., since the expected start of pregnancy) and who delivered at a gestational

[20]The time series (see Figure 1) was calculated at the Center for Evidence-Based Healthcare (ZEGV) at the Faculty of Medicine of the Dresden University of Technology (TUD). The ZEGV has access to individual-level data from the AOK Plus Saxony for research on Covid-19 and dementia. Unfortunately, the individual-level data cannot currently be used for other research purposes. We have received special permission to use the aggregated time series directly from the responsible business unit of AOK Plus. For details, see Acknowledgments. We obtained the data starting in January 2008, but do not use the first two years due to a significant break in the series in October 2009 as a result of a change in relevant reimbursement codes.

age greater than 26 weeks.[21] This results in approximately 1,300 deliveries per month used for the calculations. Although the AOK Plus data only cover Saxony, the average screening rates for GDM appear to be very similar to those reported in other studies for Germany as a whole. For example, the average screening rate in our data was 39.7% in 2010 and 92.2% in 2014, compared with 40.2% and 90.8%, respectively in Lappe et al. (2023).

5 Results

5.1 Main results

We begin by providing descriptive evidence on the effect of the policy on the actual uptake of GDM screening using aggregate data from AOK Plus. Figure 1 shows the evolution of GDM screening rates over time. In general, we observe an increasing trend over time, probably due to the increasing number of older and overweight pregnant women - two major risk factors for developing gestational diabetes (Cleary-Goldman et al., 2005). There is no substantial increase in the screening rates immediately following the introduction of the GDM screening into maternity guidelines in 2012, but the trend appears to slightly level off thereafter. From mid-2012 to summer 2013, the proportion of mothers screened for GDM during pregnancy increased steadily from about 50% to 65%. In the fall of 2013, screening rates jumped to almost 90% immediately after the introduction of screening reimbursement by statutory health insurers, and reached almost 95% by the end of 2014. The figure is consistent with evidence suggesting that pure information treatments do not significantly increase prenatal care utilization until coupled with financial incentives (e.g., Cygan-Rehm and Karbownik, 2022).

We confirm the graphical evidence in a regression framework by estimating a similar model specification as in Equation (1) to the aggregate time series (see Appendix Table A.2). For the estimations, we use only the months from July to December in years 2010 to 2013 and weight the regressions by the number of deliveries in each year $\times$ month cell. We confirm that the introduction of the free GDM screening increased the uptake of the GDM screening by more than 22 percentage points. The estimate increases slightly to 24 percentage points in a donut-hole specification that excludes deliveries in September and October, where (in)eligibility for testing may not perfectly match the birth month. The conclusions do not change when we exclude 2012 from the control group, suggesting that the earlier introduction of GDM screening into maternity guidelines does not bias the estimate. The point estimates translate into a relative

[21] The sample restrictions used for the calculations follow Lappe et al. (2023), who report annual GDM screening rates between 2010 and 2020 in Germany based on data from another insurer (BARMER).

increase of between 34% and 38% when compared with average screening rates immediately prior to the introduction of the cost reimbursement.[22]

Next, we test whether the increase in GDM screening rates translates into an improvement in birth outcomes. To do this, we estimate linear regressions of our main model specification as in Equation (1) on the individual-level data from the DRG hospital records. Each regression includes a full set of year of birth fixed effects and a *post* dummy. This allows us to flexibly capture possible year-specific and seasonal differences in birth outcomes due to reasons other than the policy change under study.

Figure 2 illustrates the rationale for our DiD design by showing substantial seasonality in birth weight. Each connected line of two dots shows the sample means for children born in the third and fourth quarters of a given calendar year. We observe that being born in late summer months is associated with higher birth weight and higher incidence of macrosomia compared to being born in the winter in all included years. This pattern is consistent with previous evidence on the seasonality of birth outcomes (e.g., Buckles and Hungerman, 2013, Currie and Schwandt, 2013). The remarkable stability of the pattern across years supports the parallel trends assumption, which ensures that the *post* dummy captures the seasonality effect. The different levels of each pair of connected dots across years are captured by the year fixed effects. The corresponding figures for other outcomes are included in Appendix Figure A.1 and Figure A.2.

Table 2 shows the reduced-form estimates of free GDM screening on newborns outcomes. Column 1 confirms that reimbursement for GDM screening increased the number of newborns affected by maternal GDM. The point coefficient translates into a relative increase of 4.3% compared to the sample mean. This result is consistent with higher screening rates. However, the relatively small effect size suggests that prior to the introduction of universal GDM screening in 2013, physicians were able to identify most affected mothers using a risk assessment based on pre-existing conditions.

Despite the slight increase in GDM diagnoses among newborns, we do not find significant effects on birth weight either when measured continuously (in grams, column 2) or as an indicator of fetal macrosomia (columns 4 and 5). The coefficients are consistently close to zero and relatively precisely estimated. For example, the 95% confidence intervals in column 2 allow us to exclude birth weight decreases greater than 3.72 grams and birth weight increases greater than 4.87 grams (i.e., $-/+0.1$% relative to the sample mean). The insignificant effect on low birth weight (column 5) suggests that there are also no effects at the lower end of the birth weight distribution. This was

[22] In the last column of Appendix Table A.2, we perform a placebo test by assuming that the free GDM screening was introduced in 2014 (instead of 2013). This specification yields zero effect on screening rates, supporting the validity of our DiD design.

expected and supports the internal validity of our main estimates. Neither the effect on length of stay (column 5) nor on reimbursement claim (column 6) are statistically significant, although the last point estimate suggests a slight cost increase for health insurers of 1.4%.

Similarly, in Table 3, we find virtually no effects on maternal outcomes. The point estimates in the first two columns are close to zero and the 95% confidence intervals allow us to exclude any meaningful effect sizes on the the probability of cesarean delivery or obstructed labor due to shoulder dystocia. Column 3 suggests a slight decrease in late-term deliveries of about 2% relative to the sample mean. However, the effect is only marginally statistically significant at the 10% level, which precludes strong conclusions. The alleged reduction in late-term births is likely a statistical artifact due to sample variability, as there is no parallel increase in full-term births (column 4), but rather a shift toward pre-term births (column 5). Both estimates are statistically insignificant, but the latter is larger in magnitude and less plausible. Finally, consistent with the results for the newborns, we find no meaningful increase in maternal length of stay (column 5) or in costs to health insurers (column 6). Although more precisely estimated, the latter effect (of 0.3%) is much smaller than in the newborn sample.

In Section 5.2 we show that the results are robust to alternative model specifications and sample restrictions. We also show that they are not driven by a differential sorting of mothers into hospital deliveries due to the introduction of the free GDM screening.

5.2 Validity and robustness checks

While the graphical evidence in Section 5.1 for the remarkable stability of seasonal patterns in birth outcomes before 2013 strongly supports the parallel trends assumption, our main estimates could still be biased if free GDM screening endogenously changed the composition of mothers giving birth around October 1, 2013. In this section, we perform various empirical exercises to mitigate such concerns.

First, we formally test whether the predetermined characteristics are balanced across the October 1 threshold. We do this by estimating difference-in-difference regressions similar to those in Equation (1) but replacing the outcome variable with a particular characteristic of the newborns or mothers. We find that the characteristics of the treatment and control groups are balanced on almost all covariates (see Appendix Table A.3). The only marginally significant difference is that, there are slightly fewer newborns (or mothers) from rural areas born (or giving birth) after October 1, 2013. However, this imbalance is small in magnitude and becomes insignificant when we adjust the p-values for multiple hypothesis testing. The results of the balancing tests

argue against the concern that our results are driven by differential composition of the treatment and the control group.

Second, we test for discontinuous changes in the number of hospital births due to the treatment. The rationale for this is that the free GDM screening may have led to a differential sorting of mothers into hospital births in 2013 compared to the control years. In general, large shifts are unlikely because out-of-hospital births are extremely rare in Germany (e.g., Kreyenfeld et al., 2010).[23] Nevertheless, to mitigate concerns that our estimates are biased by endogenous sample selection, we aggregate the data into cells by year × month × municipality × age group and count the number of deliveries in each cell.[24] We then regress the number of births on our main model specification (see Equation (1)) using the aggregated data. We do not find a significantly different mass of births across the October 1 treashold in 2013 (see Appendix Table A.4), but the estimates are relatively imprecise to allow for strong conclusions.

Finally, we test the robustness of our results to extended model specifications and sample restrictions (see Appendix Table A.5) and Table A.6). For example, we show that the estimates are remarkably robust to controlling for predetermined background characteristics, confirming that observable characteristics are balanced across treatment and control groups. We also estimate a doughnut-hole regression specification that excludes mothers who gave birth close to the October 1, 2013 cutoff. This specification eliminates any confounding effects of potential endogenous sorting across the cutoff. It also addresses some fuzziness in treatment assignment using birth month. All alternative results support our main conclusions.

6 Conclusions

This paper examines the effects of a universal screening for gestational diabetes mellitus (GDM) on maternal and infant health. GDM is one of the most common medical complications of pregnancy and is associated with adverse perinatal outcomes and serious lifelong health problems for both mother and a child (e.g., McIntyre et al., 2019, Yu et al., 2019, Shah et al., 2021). We take advantage of the quasi-experimental variation in free GDM screening eligibility induced by the introduction of full reimbursement for an oral GDM test by the German Statutory Health Insurance system in July 2013 (see e.g., Tamayo et al., 2016). Specifically, we estimate the intention-to-treat effects of this policy by combining high-quality administrative data on all hospital births

[23] The proportion of out-of-hospital births in Germany is less than 2%. The most common alternatives are birth in a certified center or home birth with a midwife.

[24] In the newborns sample, we use gender instead of age group for this data aggregation.

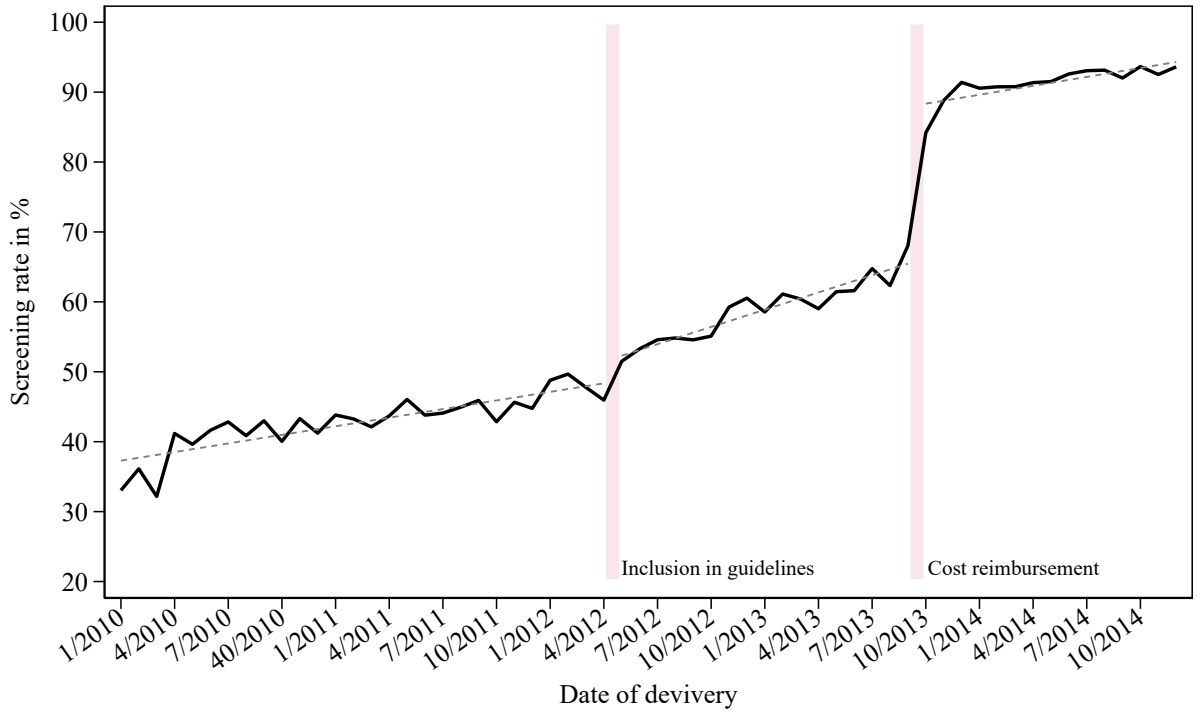
with a difference-in-discontinuities design. Using complementary data from process-generated claims from a large health insurer, we also provide evidence of the immediate effect on actual GDM screening utilization among pregnant women.

We find that the introduction of a free GDM screening led to a sharp increase in screening rates by almost 25 percentage points (or 34%-38% compared to the sample means). However, despite the large effects on prenatal GDM screening rates, we find only a moderate (albeit statistically significant) increase of about 4% in the number of newborns affected by maternal GDM. The estimated effects on other neonatal health measures and maternal birth outcomes, such as birth weight, macrosomia, cesarean section, and shoulder dystocia, are virtually zero. Consequently, we find no meaningful changes in length of stay and hospital reimbursement claims, which we interpret as proxies for economic costs of birth to statutory health insurers.

Our results suggest that even before the introduction of universal GDM screening in 2013, German physicians were able to identify mothers most at risk for adverse GDM consequences at birth using a risk-based assessment. In the future, we plan to examine whether the zero effects on immediate birth outcomes persist using process-generated claims data from all statutory health insurers in Germany, which include information on all drug dispensations, outpatient diagnoses, and services up to ten years after treatment. This will allow us to better understand whether the screening itself induced any potential changes in maternal behavior (e.g., changes in diet and physical activity) that may have long-term effects on the health of children and mothers.

Figures and Tables

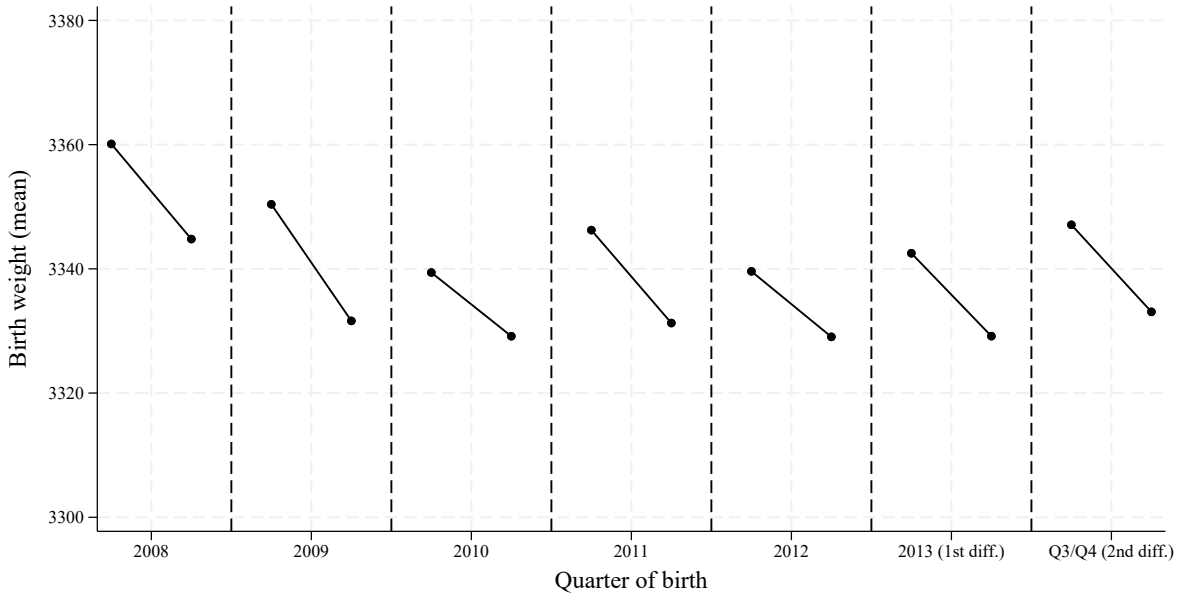
Figure 1: GDM screening rates over time



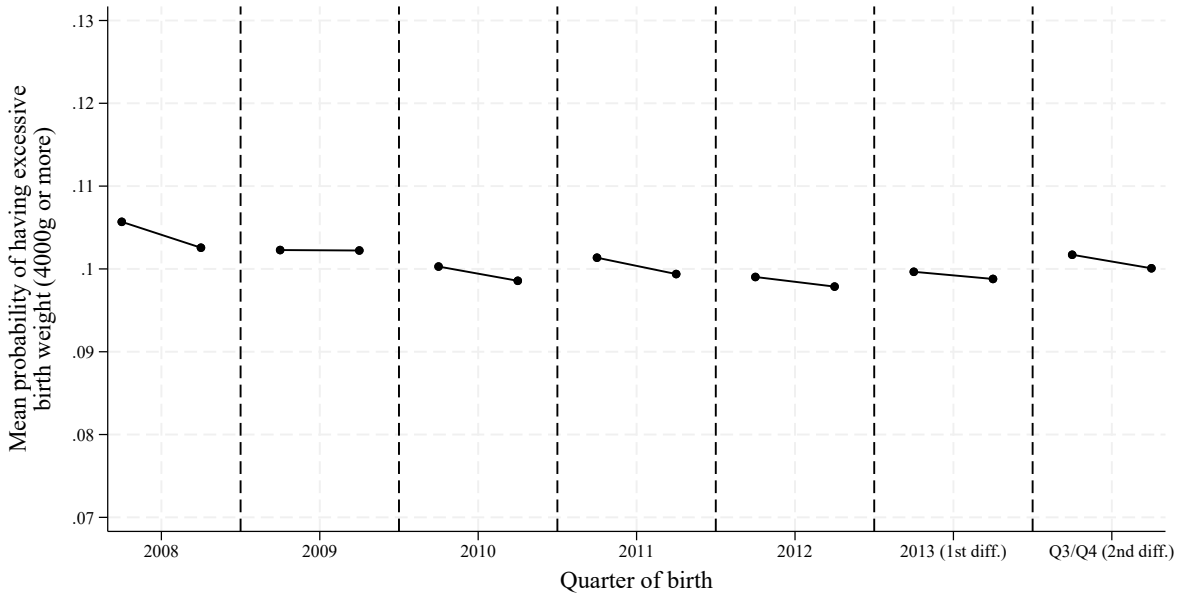
Note: The solid line shows the proportion of mothers tested for GDM during pregnancy by the date of delivery. The dashed lines represent trends that fitted to the data separately for three time intervals. The data are restricted to mothers who were insured with AOK Plus for at least 267 days (i.e., since the expected start of pregnancy) and who delivered at a gestational age greater than 26 weeks.
Source: AOK Plus Sachsen.

Figure 2: Year-to-year seasonality in birth weight

(a) Birth weight (in gram)



(b) Macrosomia (>4,000 gram)



Note: Sample restricted to births from July to December in the years 2008 to 2013. Each set of connected dots compares the sample means for children born in the 3rd and 4th quarters of a given calendar year. The 2nd difference comparison is between the means for the 3rd and 4th quarters aggregated over the pre-reform years 2008-2012.
Source: Diagnosis-Related Groups Statistic (DRG).

Table 1: Sample means

Estimation sample	Newborns	Mothers
A: Outcomes		
Affected by maternal GDM	0.036	
Birth weight (in gram)	3,339.775	
Macrosomia $\geq$4,000g	0.101	
Macrosomia $\geq$4,500g	0.012	
Low birth weight ($<$2500g)	0.064	
C-section		0.308
Obstructed labor due to shoulder dystocia		0.014
Late-term birth ($>$ 41 weeks)		0.097
Full-term term (37-41 weeks)		0.822
Pre-term birth ($<$ 37 weeks)		0.067
Length of stay (in days)	4.768	4.415
Reimbursement claim to the insurer (in euros)	1,715.572	1,943.649
B: Individual characteristics		
Female	0.489	1.000
Age	0.000	29.961
Schleswig-Holstein	0.033	0.033
Hamburg	0.026	0.025
Lower Saxony	0.092	0.095
Bremen	0.008	0.008
North Rhine-Westphalia	0.212	0.217
Hesse	0.076	0.076
Rhineland-Palatinate	0.049	0.048
Baden-Wuerttemberg	0.138	0.136
Bavaria	0.158	0.157
Saarland	0.011	0.010
Berlin	0.049	0.048
Brandenburg	0.027	0.026
Mecklenburg-Western Pomerania	0.018	0.019
Saxony	0.052	0.051
Saxony-Anhalt	0.025	0.026
Thuringia	0.025	0.025
City	0.522	0.521
Urban	0.322	0.321
Rural	0.156	0.158
Obs.	1,932,448	1,957,643

Note: Samples restricted to births from July to December in years 2008 to 2013. The regional indicators refer to the place of residence.
Source: Diagnosis-Related Groups Statistic (DRG).

Table 2: Effcts of free GDM screening eligibility on newborns' outcomes

	(1) Affected by GDM	(2) Birth weight	(3) Macrosomia ($\geq$ 4,000g)	(4) Macrosomia ($\geq$ 4,500g)	(5) Low birth weight	(6) Length of stay	(7) Reimbursement claim
post $\times$ *treat*	0.002**	0.574	0.001	0.000	0.001	0.014	23.559
	(0.001)	(2.191)	(0.001)	(0.000)	(0.001)	(0.030)	(25.876)
	[0.000; 0.003]	[-3.720; 4.868]	[-0.002; 0.003]	[-0.001; 0.001]	[-0.001; 0.003]	[-0.044; 0.073]	[-27.157; 74.276]
Rel. to Y-mean	4.3%	0.0%	0.7%	1.8%	1.3%	0.3%	1.4%
Y-Mean	0.036	3339.775	0.101	0.012	0.064	4.786	1,715.572
Obs.	1,932,448	1,932,448	1,932,448	1,932,448	1,932,448	1,932,448	1,932,448

Note: Samples restricted to births from July to December in the years 2008 to 2013. Each cell is based on a separate linear regression of equation (1). All regressions include year of birth fixed effects and a *post* dummy. Robust standard errors are shown in parentheses and the corresponding 95% confidence intervals in brackets.
Source: Diagnosis-Related Groups Statistic (DRG).

Table 3: Effcts of free GDM screening eligibility on mothers' birth outcomes

	(1) C-section	(2) Shoulder dystocia	(3) Late-term (>41 weeks)	(4) Full-term (37-41 weeks)	(5) Pre-term (<37 weeks)	(6) Length of stay	(7) Reimbursement claim
post × *treat*	-0.001	0.000	-0.002*	0.000	0.001	0.008	6.105*
	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)	(0.012)	(3.680)
	[-0.004; 0.002]	[-0.001; 0.001]	[-0.004; 0.000]	[-0.003; 0.003]	[-0.001; 0.003]	[-0.016; 0.032]	[-1.108; 13.318]
Rel. to Y-mean	-0.3%	1.0%	-1.9%	0.0%	1.7%	0.2%	0.3%
Y-Mean	0.308	0.014	0.097	0.822	0.067	4.415	1943.649
Obs.	1,957,643	1,957,643	1,957,643	1,957,643	1,957,643	1,957,643	1,957,643

Note: Samples restricted to births from July to December in the years 2008 to 2013. Each cell is based on a separate linear regression of equation (1). All regressions include year of birth fixed effects and a *post* dummy. Robust standard errors are shown in parentheses and the corresponding 95% confidence intervals in brackets.
Source: Diagnosis-Related Groups Statistic (DRG).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used *ChatGPT 4o* in order to conflate ideas, get feedback on logical reasoning, and for table formatting. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data statement

This paper uses proprietary data from the Diagnosis-Related Groups Statistic (DRG) 2008-2014 that cannot be published. However, the data can be requested (e.g., for replication purposes) from the Research Data Center (FDZ) of the Federal Statistical Office (as EVAS 23141 statistics) in Wiesbaden (Germany). The authors are willing to assist.

Acknowledgements

We would like to thank Prof. Dr. med. Jochen Schmitt and Falko Tesch from the Center for Evidence-Based Healthcare (ZEGV) at the Faculty of Medicine of the Dresden University of Technology (TUD) for providing us with the time series for Figure 1 and Stefan Baßler (Business Unit Manager, Customer and Business Analytics) from the AOK PLUS for the special permission to use the AOK Plus Sachsen data for this figure. We acknowledge feedback from seminar participants at the University of Erlangen-Nuremberg (FAU) and the Dresden University of Technology (TUD).

References

Alalouf, Mattan, Sarah Miller, and Laura R Wherry, "What difference does a diagnosis make? Evidence from marginal patients," *American Journal of Health Economics*, 2024, *10* (1), 97–131.

Almond, D and J Currie, "Human capital development before age five. Handbook of Labor Economics, Vol. 4B, ed. OC Ashenfelter, D Card," 2011.

Avdic, Daniel and Arizo Karimi, "Modern family? Paternity leave and marital stability," *American Economic Journal: Applied Economics*, 2018, *10* (4), 283–307.

Bellamy, Leanne, Juan-Pablo Casas, Aroon D Hingorani, and David Williams, "Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis," *The Lancet*, 2009, *373* (9677), 1773–1779.

Beyerlein, Andreas, Daniela Koller, Anette-Gabriele Ziegler, Nicholas Lack, and Werner Maier, "Does charge-free screening improve detection of gestational diabetes in women from deprived areas: a cross-sectional study," *BMC pregnancy and childbirth*, 2016, *16*, 1–6.

Bobak, Martin and Arjan Gjonca, "The seasonality of live birth is strongly influenced by socio-demographic factors," *Human reproduction*, 2001, *16* (7), 1512–1517.

Borra, Cristina, Ana Costa-Ramón, Libertad González, and Almudena Sevilla-Sanz, "The causal effect of an income shock on children's human capital," *Barcelona GSE Working Paper Series*, 2021, (1272).

Buckles, Kasey S and Daniel M Hungerman, "Season of birth and later outcomes: Old questions, new answers," *Review of Economics and Statistics*, 2013, *95* (3), 711–724.

Clarke, Damian, Sonia Oreffice, and Climent Quintana-Domeque, "The demand for season of birth," *Journal of Applied Econometrics*, 2019, *34* (5), 707–723.

Cleary-Goldman, Jane, Fergal D Malone, John Vidaver, Robert H Ball, David A Nyberg, Christine H Comstock, George R Saade, Keith A Eddleman, Susan Klugman, Lorraine Dugoff et al., "Impact of maternal age on obstetric outcome," *Obstetrics & Gynecology*, 2005, *105* (5 Part 1), 983–990.

Conti, Gabriella and Paul Rodriguez-Lesmes, "Early Childhood Health Inequalities and In-Utero Health Interventions: Evidence from the Treatment of Gestational Diabetes," Presented at the NBER Health Economics Program Meeting, May 6-7, 2021 2021.

Corman, Hope, Dhaval Dave, and Nancy E. Reichman, "The Effects of Prenatal Care on Birth Outcomes: Reconciling a Messy Literature," 04 2019.

Currie, Janet and Hannes Schwandt, "Within-mother analysis of seasonal patterns in health at birth," *Proceedings of the National Academy of Sciences*, 2013, *110* (30), 12265–12270.

__ **and Jeffrey Grogger**, "Medicaid expansions and welfare contractions: offsetting effects on prenatal care and infant health?," *Journal of health economics*, 2002, *21* (2), 313–335.

__ **and Maya Rossin-Slater**, "Early-life origins of life-cycle well-being: Research and policy implications," *Journal of policy Analysis and management*, 2015, *34* (1), 208–242.

Cygan-Rehm, Kamila, "Parental leave benefit and differential fertility responses: Evidence from a German reform," *Journal of Population Economics*, 2016, *29*, 73–103.

__ **and Krzysztof Karbownik**, "The effects of incentivizing early prenatal care on infant health," *Journal of Health Economics*, 2022, *83*, 102612.

Dustmann, Christian and Uta Schönberg, "Expansions in maternity leave coverage and children's long-term outcomes," *American Economic Journal: Applied Economics*, 2012, *4* (3), 190–224.

Eades, Claire E, Katherine A Burrows, Roza Andreeva, Daniel R Stansfield, and Josie MM Evans, "Prevalence of gestational diabetes in the United States and Canada: a systematic review and meta-analysis," *BMC Pregnancy and Childbirth*, 2024, *24* (1), 204.

Einav, Liran, Amy Finkelstein, Tamar Oostrom, Abigail Ostriker, and Heidi Williams, "Screening and selection: The case of mammograms," *American Economic Review*, 2020, *110* (12), 3836–3870.

Eisemann, Nora, Stefan Bunk, Trasias Mukama, Hannah Baltus, Susanne A Elsner, Timo Gomille, Gerold Hecht, Sylvia Heywang-Köbrunner, Regine Rathmann, Katja Siegmann-Luz et al., "Nationwide real-world implementation of AI for cancer detection in population-based mammography screening," *Nature Medicine*, 2025, pp. 1–8.

Evans, William N and Diana S Lien, "The benefits of prenatal care: evidence from the PAT bus strike," *Journal of Econometrics*, 2005, *125* (1-2), 207–239.

Farrar, Diane, Lelia Duley, Therese Dowswell, and Debbie A Lawlor, "Different strategies for diagnosing gestational diabetes to improve maternal and infant health," *Cochrane Database of Systematic Reviews*, 2017, (8).

__ **, Mark Simmonds, Maria Bryant, Trevor A Sheldon, Derek Tuffnell, Su Golder, Fidelma Dunne, and Debbie A Lawlor**, "Hyperglycaemia and risk of adverse perinatal outcomes: systematic review and meta-analysis," *BMJ*, 2016, *354*.

FDZ, "Metadata report. Part I: General and methodological information on the Diagnosis-Related Groups Statistic (DRG) (EVAS number: 23141)," Version 2, Research Data Centres of the Statistical Offices ot the Federation and the Federal States (FDZ). Wiesbaden. DOI: 10.21242/23141.2013.00.00.1.2.0 2019.

Felfe, Christina, Helmut Rainer, and Judith Saurer, "Why birthright citizenship matters for immigrant children: Short-and long-run impacts on educational integration," *Journal of Labor Economics*, 2020, *38* (1), 143–182.

Gaudet, Laura, Zachary M Ferraro, Shi Wu Wen, and Mark Walker, "Maternal obesity and occurrence of fetal macrosomia: a systematic review and meta-analysis," *BioMed research international*, 2014, *2014* (1), 640291.

Giacomo, Marina Di, Massimiliano Piacenza, Luigi Siciliani, and Gilberto Turati, "The effect of co-payments on the take-up of prenatal tests," *Journal of Health Economics*, 2022, *81*, 102553.

Guthmuller, Sophie, Vincenzo Carrieri, and Ansgar Wübker, "Effects of organized screening programs on breast cancer screening, incidence, and mortality in Europe," *Journal of Health Economics*, 2023, *92*, 102803.

Haschka, Stefanie J, Christina Gar, Vanessa Sacco, Friederike Banning, Uta Ferrari, Ines Freibothe, Stefanie Kern-Matschilles, Anne L Potzel, Barbara Rauch, Louise U Fueessl et al., "Pre-diabetes, diabetes and fluctuations of glucose tolerance after gestational diabetes mellitus: 5-year follow-up of a contemporary, prospective study in Germany," *BMJ Open Diabetes Research and Care*, 2022, *10* (2), e002621.

Joyce, Theodore, "Impact of augmented prenatal care on birth outcomes of Medicaid recipients in New York City," *Journal of health economics*, 1999, *18* (1), 31–67.

Kramer, Caroline K, Sara Campbell, and Ravi Retnakaran, "Gestational diabetes and the risk of cardiovascular disease in women: a systematic review and meta-analysis," *Diabetologia*, 2019, *62*, 905–914.

Kreyenfeld, Michaela, Rembrandt Scholz, Frederik Peters, and Ines Wlosnewski, "Ordnungsspezifische Geburtenraten für Deutschland," *Zeitschrift für Bevölkerungswissenschaft*, 2010, *35*, 225–244.

Lappe, Veronika, Gregory Gordon Greiner, Ute Linnenkamp, Anja Viehmann, Heinke Adamczewski, Matthias Kaltheuner, Dietmar Weber, Ingrid Schubert, and Andrea Icks, "Gestational diabetes in Germany—prevalence, trend during the past decade and utilization of follow-up care: an observational study," *Scientific Reports*, 2023, *13* (1), 16157.

Maciosek, Michael V, Ashley B Coffield, Thomas J Flottemesch, Nichol M Edwards, and Leif I Solberg, "Greater use of preventive services in US health care could save lives at little or no cost," *Health Affairs*, 2010, *29* (9), 1656–1660.

McIntyre, H David, Patrick Catalano, Cuilin Zhang, Gernot Desoye, Elisabeth R Mathiesen, and Peter Damm, "Gestational diabetes mellitus," *Nature reviews Disease primers*, 2019, *5* (1), 47.

Metzger, Boyd E, Lynn P Lowe, Alan R Dyer, Elisabeth R Trimble, Udom Chaovarindr, Donald R Coustan, David R Hadden, David R McCance, Moshe Hod, Harold David McIntyre et al., "Hyperglycemia and adverse pregnancy outcomes," *The New England journal of medicine*, 2008, *358* (19), 1991–2002.

OECD, *Germany: Country Health Profile 2023, State of Health in the EU*, OECD Publishing, Paris/European Observatory on Health Systems and Policies, Brussels, 2023.

Reitzle, Lukas, Christian Schmidt, Christin Heidemann, Andrea Icks, Matthias Kaltheuner, Thomas Ziese, and Christa Scheidt-Nave, "Gestational diabetes in Germany: Development of screening participation and prevalence," *Journal of Health Monitoring*, 2021, *6* (2), 3.

Riukula, Krista, "The effects of screening for gestational diabetes," *Empirical Economics*, 2023, *65* (4), 1931–1964.

Sabik, Lindsay M and Cathy J Bradley, "The impact of near-universal insurance coverage on breast and cervical cancer screening: evidence from Massachusetts," *Health economics*, 2016, *25* (4), 391–407.

Schäfer-Graf, Ute, Katharina Laubner, Sandra Hummel, Ulrich Gembruch, Tanja Groten, Franz Kainer, Melita Grieshop, Dagmar Bancher-Todesca, Mila Cervar-Zivakovic, Irene Hösli et al., "Gestationsdiabetes mellitus (GDM), Diagnostik, Therapie und Nachsorge–Praxisempfehlung–Kurzfassung der S3-Leitlinie (AWMF-Registernummer: 057-008)," *Diabetologie und Stoffwechsel*, 2020, *15* (S 01), S101–S111.

Schnell-Inderst, Petra, Theresa Hunger, Katharina Hintringer, Ruth Schwarzer, Vanadin Regina Seifert-Klauss, Holger Gothe, Jürgen Wasem, and Uwe Siebert, "Individual health services," *GMS Health Technology Assessment*, 2011, *7*.

Schönberg, Uta and Johannes Ludsteck, "Expansions in maternity leave coverage and mothers' labor market outcomes after childbirth," *Journal of Labor Economics*, 2014, *32* (3), 469–505.

Shah, Nilay S, Michael C Wang, Priya M Freaney, Amanda M Perak, Mercedes R Carnethon, Namratha R Kandula, Erica P Gunderson, Kai McKeever Bullard, William A Grobman, Matthew J O'Brien et al., "Trends in gestational diabetes at first live birth by race and ethnicity in the US, 2011-2019," *Jama*, 2021, *326* (7), 660–669.

Tamayo, T, M Tamayo, W Rathmann, and P Potthoff, "Prevalence of gestational diabetes and risk of complications before and after initiation of a general systematic two-step screening strategy in Germany (2012–2014)," *Diabetes research and clinical practice*, 2016, *115*, 1–8.

Vetter, K and M Goeckenjan, "Prenatal care in Germany," *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 2013, *56* (12), 1679–1685.

Wang, Hui, Ninghua Li, Tawanda Chivese, Mahmoud Werfalli, Hong Sun, Lili Yuen, Cecilia Ambrosius Hoegfeldt, Camille Elise Powe, Jincy Immanuel, Suvi Karuranga et al., "IDF diabetes atlas: estimation of global and regional gestational diabetes mellitus prevalence for 2021 by International Association of Diabetes in Pregnancy Study Group's Criteria," *Diabetes research and clinical practice*, 2022, *183*, 109050.

Yu, Yongfu, Onyebuchi A Arah, Zeyan Liew, Sven Cnattingius, Jørn Olsen, Henrik Toft Sørensen, Guoyou Qin, and Jiong Li, "Maternal diabetes during pregnancy and early onset of cardiovascular disease in offspring: population based cohort study with 40 years of follow-up," *Bmj*, 2019, *367*.

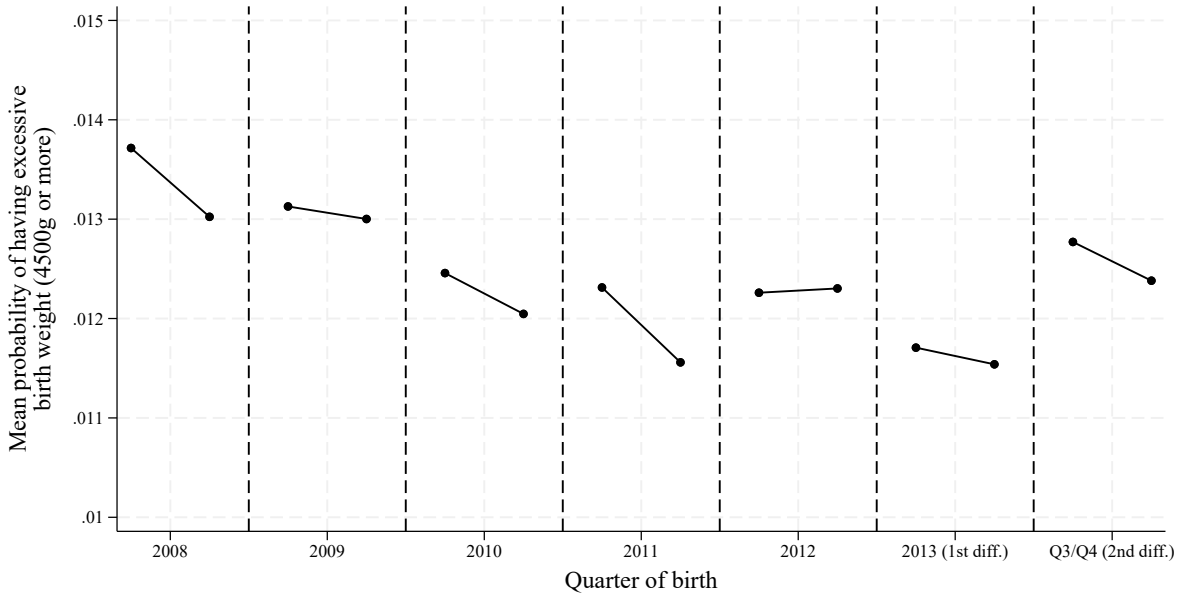
Zhang, Xun, Adriana Decker, Robert W Platt, and Michael S Kramer, "How big is too big? The perinatal consequences of fetal macrosomia," *American journal of obstetrics and gynecology*, 2008, *198* (5), 517–e1.

Zhu, Yeyi, Sjurdur F Olsen, Pauline Mendola, Edwina H Yeung, Allan Vaag, Katherine Bowers, Aiyi Liu, Wei Bao, Shanshan Li, Camilla Madsen et al., "Growth and obesity through the first 7 y of life in association with levels of maternal glycemia during pregnancy: a prospective cohort study," *The American Journal of Clinical Nutrition*, 2016, *103* (3), 794–800.

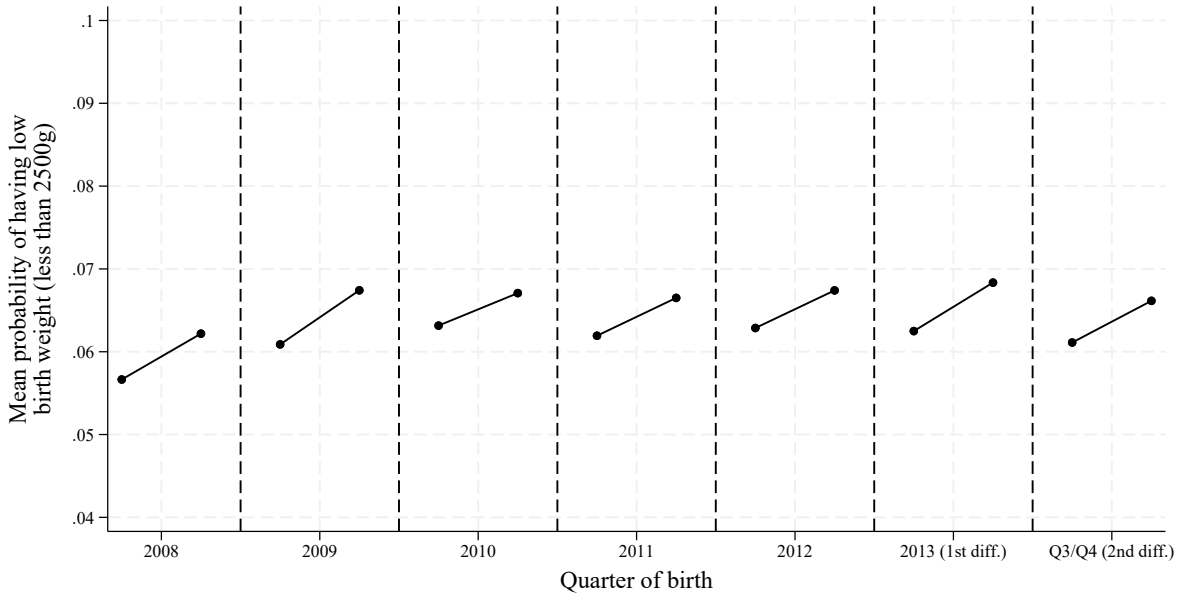
– Online Appendix –

Figure A.1: Year-to-year seasonality in extremely high and low birth weight

(a) Macrosomia (>4,500 gram)



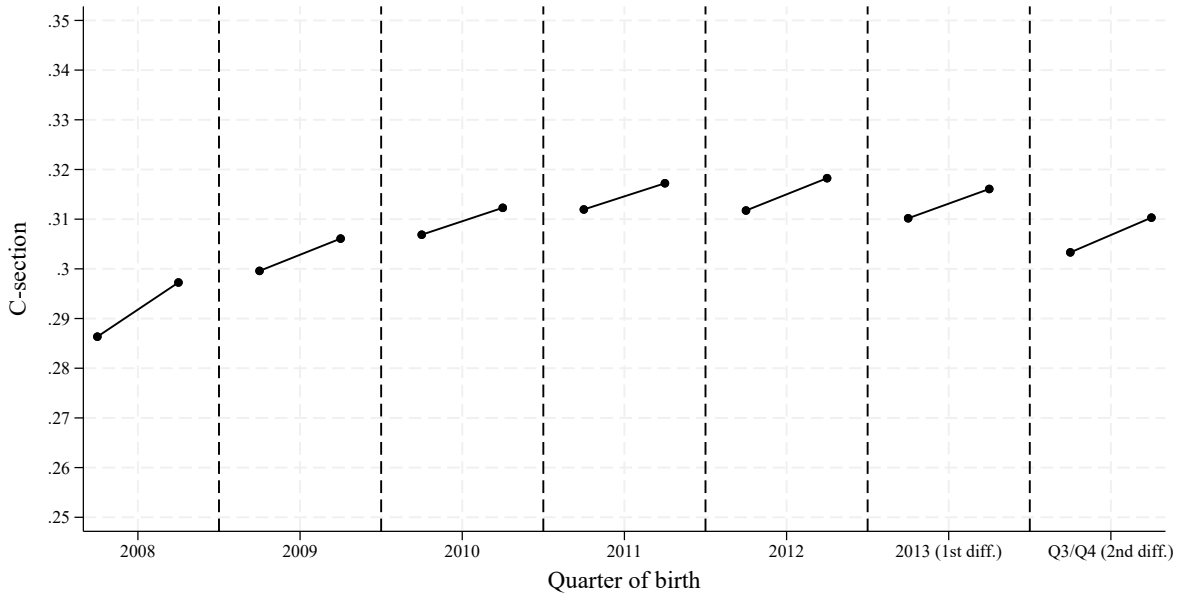
(b) Low birth weight(<2,500 gram)



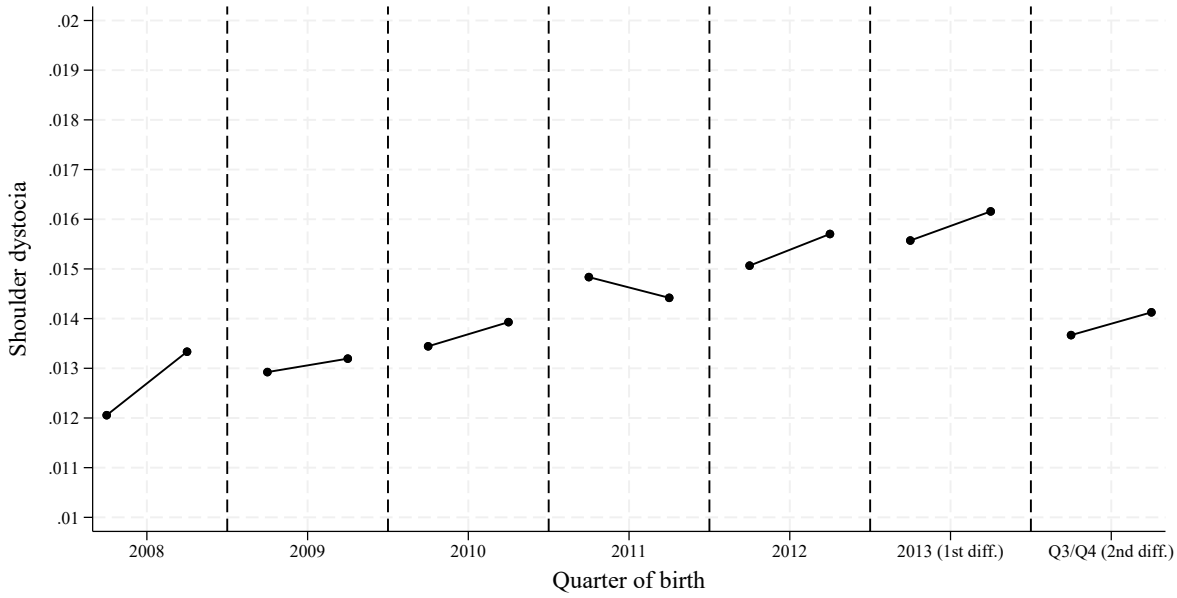
Note: Sample restricted to births from July to December in the years 2008 to 2013. Each set of connected dots compares the sample means for children born in the 3rd and 4th quarters of a given calendar year. The 2nd difference comparison is between the means for the 3rd and 4th quarters aggregated over the pre-reform years 2008-2012.
Source: Diagnosis-Related Groups Statistic (DRG).

Figure A.2: Year-to-year seasonality in birth complications

(a) C-section



(b) Obstructed labor due to shoulder dystocia



Note: Sample restricted to births from July to December in the years 2008 to 2013. Each set of connected dots compares the sample means for children born in the 3rd and 4th quarters of a given calendar year. The 2nd difference comparison is between the means for the 3rd and 4th quarters aggregated over the pre-reform years 2008-2012.
Source: Diagnosis-Related Groups Statistic (DRG).

Table A.1: Comparison of sample sizes with official statistics

	Number of newborns			**Number of deliveries**		
Year	DRG newborn sample	Official stats. (Destatis)	Share (in %)	DRG mothers' sample	Official stats. (Destatis)	Share (in %)
2008	616,908	682,514	90.39%	633,928	662,783	95.65%
2009	614,642	665,126	92.41%	621,703	644,274	96.50%
2010	630,582	677,947	93.01%	636,210	656,390	96.93%
2011	615,421	662,685	92.87%	625,067	642,791	97.24%
2012	628,807	673,544	93.36%	636,787	653,215	97.49%
2013	637,445	682,069	93.46%	644,655	661,138	97.51%
Total	3,743,805	4,043,885	92.58%	3,798,350	3,920,591	96.88%

Note: Share corresponds to the ratio of the year-specific number of observations in the DRG estimation samples to the number of newborns or deliveries from official statistics.
Source: Diagnosis-Related Groups Statistic (DRG); Federal Statistical Office (Destatis).

Table A.2: Effects of free GDM screening eligibility on screening rates

	(1)	(2)	(3)	(4)
	Baseline	Donut-hole	W/o 2012	Placebo
post × *treat*	0.222***	0.244***	0.236***	-0.002
	(0.027)	(0.019)	(0.027)	(0.010)
post	0.008	0.021**	-0.006	0.008
	(0.009)	(0.009)	(0.007)	(0.009)
y2010	ref.	ref.	ref.	ref.
y2011	0.028***	0.028**	0.028***	0.028***
	(0.009)	(0.010)	(0.007)	(0.009)
y2012	0.145***	0.151***	-	0.145***
	(0.012	(0.013)		(0.012)
y2013 = *treat*	0.236***	0.225***	0.229***	-
	(0.017)	(0.014)	(0.017)	
y2014 = *treat*	-	-	-	0.512***
				(0.008)
Const.	0.415***	0.411***	0.422***	0.415***
	(0.007)	(0.009)	(0.006)	(0.007)
Y-mean	0.651	0.636	0.651	0.927
Rel. to Y-mean	34.1%	38.4%	36.3%	-0.03%
No. of year × month cells	24	16	18	24
No. ob deliveries	33,451	22,242	25,073	33,843

Note: Unless otherwise noted, data are restricted to births from July to December in the years 2008 to 2013. In column 2, September and October births are excluded. In column 4, the treatment year 2013 is replaced by 2014. Each column is based on a separate linear regression of equation (1). All regressions include year of birth fixed effects and a *post* dummy. The regressions are based on data are aggregated into year × month cells and weighted by the number of deliveries in each cell. Robust standard errors are reported in parentheses. Y-mean denotes the sample mean in the last quarter before the introduction of free GDM screening.
Source: AOK Plus Sachsen.

Table A.3: Balancing tests

Estimation sample	Newborns			Mothers		
	Coeff.	St. Err.	p-value	Coeff.	St. Err.	p-value
Female	0.000	(0.002)	0.977	-	-	-
Age	-	-	-	-0.011	(0.002)	0.003
Schleswig-Holstein	0.001	(0.001)	0.380	0.000	(0.001)	0.663
Hamburg	0.001	(0.001)	0.097	0.001	(0.001)	0.184
Lower Saxony	-0.001	(0.001)	0.442	-0.002	(0.001)	0.163
Bremen	0.000	(0.000)	0.912	0.000	(0.000)	0.941
North Rhine-Westphalia	-0.001	(0.002)	0.590	0.000	(0.002)	0.983
Hesse	-0.001	(0.001)	0.390	-0.001	(0.001)	0.534
Rhineland-Palatinate	0.000	(0.001)	0.834	-0.001	(0.001)	0.495
Baden-Wuerttemberg	0.001	(0.001)	0.665	0.000	(0.001)	0.743
Bavaria	0.000	(0.001)	0.874	0.001	(0.001)	0.318
Saarland	0.000	(0.000)	0.434	0.000	(0.000)	0.573
Berlin	0.001	(0.001)	0.295	0.000	(0.001)	0.615
Brandenburg	0.000	(0.001)	0.993	0.000	(0.001)	0.717
Mecklenburg-Western Pomerania	0.000	(0.000)	0.576	-0.001	(0.001)	0.115
Saxony	-0.001	(0.001)	0.494	0.000	(0.001)	0.785
Saxony-Anhalt	0.000	(0.001)	0.961	0.000	(0.001)	0.752
Thuringia	0.000	(0.001)	0.533	0.000	(0.001)	0.686
City	0.002	(0.002)	0.176	0.002	(0.002)	0.223
Urban	-0.000	(0.001)	0.935	0.001	(0.002)	0.575
Rural	-0.002	(0.001)	0.098	-0.003	(0.002)	0.026
Obs.	1,932,448			1,957,643		

Note: Samples restricted to births from July to December in years 2008 to 2013. Each coefficient is based on a separate linear regression of a given characteristic on the interaction term between the *post* and *dummy* as specified in equation (1). All regressions include year of birth fixed effects and a *post* dummy. The federal state and regional indicators refer to the place of residence.
Source: Diagnosis-Related Groups Statistic (DRG).

Table A.4: Effects on the number of hospital births

	(1)	(2)	(3)	(4)
	Newborns		Mothers	
	Baseline	Incl. controls	Baseline	Incl. controls
post × *treat*	-6.332	-4.350	-1.253	-1.781
	(49.685)	(4.367)	(12.573)	(7.312)
Rel. to Y-mean	-4.3%	-3.0%	-2.4%	-3.4%
Municipality FE, gender	no	yes	no	no
Municipality FE, age group FE	no	no	no	yes
Y-mean	146.542	146.542	52.655	52.655
Obs.	13,187	13,187	37,179	37,179

Note: Samples restricted to births from July to December in years 2008 to 2013. Data in the newborn sample are aggregated into year × month × municipality × gender cells. Data in the mothers' sample are aggregated into year × month × municipality × age group cells. Municipality refers to the location of the hospital. The dependent variable is the number of births in each cell. Each column is based on a separate linear regression of equation (1). All regressions include year of birth fixed effects and a *post* dummy. FE = fixed effects.
Source: Diagnosis-Related Groups Statistic (DRG).

Table A.5: Sensitivity analysis: newborns' sample

	(1) Affected by GDM	(2) Birth weight	(3) Macrosomia ($\geq$ 4,000g)	(4) Macrosomia ($\geq$ 4,500g)	(5) Low birth weight	(6) Length of stay	(7) Reimbursement claim
A: Baseline	0.002**	0.574	0.001	0.000	0.001	0.014	23.559
(Obs. 1,932,448)	(0.001)	(2.191)	(0.001)	(0.000)	(0.001)	(0.030)	(25.876)
Rel. to Y-mean	4.3%	0.0%	0.7%	1.8%	1.3%	0.3%	1.4%
B: Incl. controls	0.002**	0.627	0.001	0.000	0.001	0.016	22.761
(Obs. 1,932,448)	(0.001)	(2.175)	(0.001)	(0.000)	(0.001)	(0.030)	(25.865)
Rel. to Y-mean	4.2%	0.0%	0.7%	1.8%	1.3%	0.3%	1.3%
C: Donut-hole	0.002**	0.627	0.001	0.000	0.001	0.016	22.761
(Obs. 1,623,807)	(0.001)	(2.175)	(0.001)	(0.000)	(0.001)	(0.030)	(25.865)
Rel. to Y-mean	4.8%	0.0%	1.2%	0.4%	1.2%	-0.2%	0.3%

Note: Samples restricted to births from July to December in the years 2008 to 2013. Each cell is based on a separate linear regression of equation (1). All regressions include year of birth fixed effects and a *post* dummy. Controls include an indicator for gender, state fixed effects, and indicators for urban and rural areas. Donut-hole specification excludes September and October births. Robust standard errors are shown in parentheses and the corresponding 95% confidence intervals in brackets.
Source: Diagnosis-Related Groups Statistic (DRG).

Table A.6: Sensitivity analysis: mothers' sample

	(1) C-section	(2) Shoulder dystocia	(3) Late-term (>41 weeks)	(4) Full-term (37-41 weeks)	(5) Pre-term (<37 weeks)	(6) Length of stay	(7) Reimbursement claim
A: Baseline	-0.001	0.000	-0.002*	0.000	0.001	0.008	6.105*
(Obs. 1,957,643)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)	(0.012)	(3.680)
Rel. to Y-mean	-0.3%	1.0%	-1.9%	0.0%	1.7%	0.2%	0.3%
B: Incl. controls	-0.001	0.000	-0.002*	0.000	0.001	0.011	6.546*
(Obs. 1,957,643)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)	(0.012)	(3.662)
Rel. to Y-mean	-0.2%	0.9%	-2.0%	0.0%	1.9%	0.2%	0.3%
C: Donut-hole	-0.002	0.000	-0.002	0.000	0.001	0.012	6.221
(Obs. 1,645,536)	(0.002)	(0.000)	(0.001)	(0.002)	(0.001)	(0.013)	(4.025)
Rel. to Y-mean	-0.7%	0.5%	-2.0%	-0.1%	2.1%	0.3%	0.3%

Note: Samples restricted to births from July to December in the years 2008 to 2013. Each cell is based on a separate linear regression of equation (1). All regressions include year of birth fixed effects and a *post* dummy. Controls include maternal age (linear and quadratic), state fixed effects, and indicators for urban and rural areas. Donut-hole specification excludes September and October births. Robust standard errors are shown in parentheses and the corresponding 95% confidence intervals in brackets.
Source: Diagnosis-Related Groups Statistic (DRG).