The Marginal Value of a Diagnosis: Evidence from Gestational Diabetes*

Jaehyun Choi†

November 5, 2026

[Job Market Paper]

Abstract

Questions about whether the benefits of additional medical spending exceed their costs have long been studied in the health economics literature, but the marginal returns to care during pregnancy have received little attention. I study gestational diabetes mellitus (GDM), an increasingly common pregnancy complication, and estimate the marginal returns to diagnosis. I use a regression discontinuity design and compare patients just above and just below the diagnostic threshold. Using MarketScan claims and laboratory data, I find that diagnosis at the margin increases a woman's out-of-pocket spending on diabetes and nutrition services by approximately $28 per additional diagnosis, while producing no detectable increase in total pregnancy spending. For this modest increase in spending on these services, crossing the threshold yields substantial infant health benefits: women diagnosed at the margin have 1.36 percentage points lower rates of infant macrosomia and 1.33 percentage points lower rates of NICU admission. Relative to the rates just below the cutoff, these declines are roughly 60 and 55 percent.

Keywords: regression discontinuity, diagnostic thresholds, gestational diabetes, health insurance claims, diabetes self-management training.

JEL codes: I11, I13, I18.

*I am especially grateful to Coady Wing, Kosali Simon, Alberto Ortega, Seth Freedman, and Justin Ross. This work benefited from helpful comments from Andrew Butters, Denvil Duncan, Rick Harbaugh, Chandler Inman, Jin Yeub Kim, Costas Lambros, Jennifer Mangano, Marilyn Pease, Parker Rogers, Nilesh Shinde, Nicole Siegal, Anna Strauss, Jia Xiang, and Maranna Yoder, as well as seminar participants in the Health Policy Workshop, Public Finance Working Group, and Research Workshop at Indiana University's O'Neill School of Public and Environmental Affairs and the Business Economics and Public Policy seminar at its Kelley School of Business. Funding: No external funding supported this research. Conflicts of interest: The author declares no conflicts of interest. Data availability: MarketScan claims and laboratory data are restricted, individual-level data used under a data-use agreement and cannot be redistributed by the author. Code and aggregate outputs supporting the tables and figures are maintained by the author.

†O'Neill School of Public and Environmental Affairs, Indiana University Bloomington, IN, USA. E-mail: jc234@iu.edu

1 Introduction

The question of whether the additional care induced by diagnosis among marginally sick patients generates health benefits that justify its costs has motivated a growing literature on the marginal returns to medical care (Alalouf, Miller and Wherry, 2024; Almond et al., 2010; Gaggero et al., 2022, 2023; Mullainathan and Obermeyer, 2022). This question is particularly important in pregnancy, where decisions about medical intervention are often made under substantial uncertainty (Oster, 2014).[1]

One area of uncertainty in pregnancy care is the diagnosis and treatment of gestational diabetes mellitus (GDM), an increasingly common pregnancy complication in the United States. The share of GDM among women giving birth increased from 6.0% in 2016 to 8.3% in 2021. The prevalence is substantially higher among older mothers, reaching 15.6% among mothers aged 40 and older in 2021 (Centers for Disease Control and Prevention, 2023). Despite the prevalence of GDM, considerable uncertainty remains over which women should be diagnosed and treated. The medical community remains divided over where to draw the diagnostic boundaries for GDM (International Association of Diabetes and Pregnancy Study Groups Consensus Panel, 2010; National Institutes of Health Consensus Development Conference Panel, 2013; Szmuilowicz, Josefson and Metzger, 2019). Different diagnostic strategies consequently use different glucose thresholds and classify different numbers of women as having GDM (Farrar, Duley, Dowswell and Lawlor, 2016; Hillier et al., 2021). This disagreement raises a question: what is the marginal benefit of GDM diagnosis for women whose glucose levels are just above a diagnostic threshold?

This paper examines the effects of GDM diagnosis among women near the diagnostic margin on medical care utilization, health care spending, and maternal and infant health. I exploit the 95 mg/dL fasting glucose threshold in the Carpenter-Coustan diagnostic criteria for GDM. Gestational diabetes mellitus refers to hyperglycemia (or high blood sugar) first diagnosed during pregnancy among women without known preexisting diabetes.[2] GDM provides a useful setting for studying the consequences of diagnosis for two reasons. First, diagnosis is based in part on a continuous measure of blood glucose and a clearly defined clinical threshold. Crossing this threshold raises the probability of a recorded GDM diagnosis. The design assumes that underlying health and potential outcomes change smoothly through the cutoff. Second, fasting glucose is measured in a laboratory, making precise manipulation

[1] Pregnant women have historically been underrepresented in clinical research. A few reasons include concerns about fetal safety, regulatory and liability risks, and difficulties in recruiting participants for pregnancy research (Shankar et al., 2026; National Academies of Sciences, Engineering, and Medicine, 2023, 2024).

[2] Women with type 1 or type 2 diabetes diagnosed before pregnancy are classified as having pregestational diabetes rather than gestational diabetes mellitus.

around the diagnostic threshold difficult. I find no evidence of manipulation at the cutoff using the test for discrete running variables developed by Frandsen (2017).

I use claims and laboratory data from a large national database of commercially insured patients. I compare women with fasting glucose values just above and below the cutoff. I then follow them through the remainder of pregnancy and for up to 60 months after delivery. I examine GDM-specific care, health care spending, delivery and infant outcomes, postpartum follow-up, and longer-term maternal and child health.

I find that patients just above the threshold are 11 percentage points more likely to receive a recorded GDM diagnosis. Using the optimal-bandwidth estimates, receiving a diagnosis at the cutoff raises the probability of a Diabetes Self-Management Training (DSMT) claim in days 0 to 29 by 29.1 percentage points and of a Medical Nutrition Therapy (MNT) claim by 15.5 points among women whose diagnosis is induced by crossing the cutoff, and out-of-pocket spending on these services rises by approximately $28 per newly diagnosed woman. It also raises the probability of any metformin fill within 90 days by 13.8 percentage points and of any biophysical profile claim in days 90 to 179 by 9.5 points. Despite these increases in GDM-related care, I find no statistically significant change in total pregnancy spending at the optimal bandwidth.

I find that crossing the threshold lowers the probability of infant macrosomia by 1.36 percentage points and of NICU admission by 1.33 percentage points. These correspond to relative declines of approximately 60 and 55 percent, respectively, compared with the rates just below the cutoff. Among women whose GDM diagnosis is induced by crossing the threshold, I find that receiving a diagnosis reduces the probability of infant macrosomia by 10.8 percentage points and of NICU admission[3] by 11.9 percentage points. I find no detectable changes in cesarean delivery, obstetric anesthesia, or delivery payments, which provides no evidence that the improvement in infant health simply reflects more intensive care at delivery.

This paper contributes to the economics literature on how diagnostic classifications affect medical care and patient outcomes. An early study by Almond et al. (2010) examines the 1,500 gram cutoff used to classify very low birth weight infants. They find that infants just below the cutoff receive more care because they are classified as very low birth weight. These infants have lower mortality than those just above the 1,500 gram cutoff.[4] Other studies

[3]The neonatal intensive care unit (NICU) is the hospital unit that provides specialized care to premature or seriously ill newborns.

[4]Barreca et al. (2011) subsequently show that the estimated mortality effect is sensitive to the inclusion of observations recorded at or near the birth weight cutoff. Related studies document improvements in later educational outcomes following additional interventions for newborns near the same threshold (Bharadwaj, Løken and Neilson, 2013; Chyn, Gold and Hastings, 2021).

examine behavioral and psychological responses to diagnosis. Gaggero et al. (2022) find that a diabetes diagnosis reduces body mass index, with less evidence of changes in smoking or alcohol consumption. Gaggero et al. (2023) find that a diabetes diagnosis increases the likelihood of a subsequent depression diagnosis, particularly among women.

Within this literature, my analysis is most closely related to Alalouf, Miller and Wherry (2024). Using claims and laboratory data from a large national U.S. insurer, they exploit the 6.5% HbA1c threshold to estimate the effects of a type 2 diabetes diagnosis among marginal patients. They find that diagnosis increases spending on recommended diabetes care by approximately $1,097 in the first year, with higher spending persisting over six years. Blood sugar improves initially, but improvements in subsequent years are not statistically significant. The estimated effects on cholesterol and mortality are also statistically insignificant, although the confidence intervals do not rule out clinically meaningful benefits. Their sample excludes patients with pregnancy related diagnoses in order to separate standard type 2 diabetes diagnoses from gestational diabetes. My analysis extends this diagnostic margin framework to pregnancy and infant health outcomes.

Second, my paper contributes to the economics literature on infant health and prenatal care. A central idea in this literature is that investments made early in life can have high returns. The early investments shape later health and human capital (Heckman, 2008; Bütikofer, Løken and Salvanes, 2019). A large literature documents the lasting consequences of conditions during pregnancy and at birth for education, earnings, and other adult outcomes (Black, Devereux and Salvanes, 2007; Almond, 2006). Other studies examine how policies and medical care during pregnancy affect infant health at birth, including expansions in Medicaid eligibility for pregnant women (Currie and Gruber, 1996), nutritional assistance through WIC (Hoynes, Page and Stevens, 2011), and prenatal care utilization (Yan, 2020). Much of this work focuses on low birth weight and related birth outcomes. I study a different margin: whether a maternal diagnosis changes prenatal care among women who have already undergone glucose testing. In GDM, a major concern is excessive fetal growth rather than low birth weight (HAPO Study Cooperative Research Group, 2008).

Beyond infant outcomes, I also examine whether a marginal GDM diagnosis changes maternal care after delivery. Women with a history of GDM face substantially higher risks of developing type 2 diabetes later in life (Bellamy et al., 2009; Vounzoulaki et al., 2020). However, recommended postpartum glucose testing is often missed (Blatt, Nakamoto and Kaufman, 2011; Smirnakis et al., 2005). Existing studies document these risks and gaps in postpartum screening, but provide less evidence on whether diagnosis at the clinical margin changes subsequent care.

Third, my paper contributes to the clinical literature on GDM screening, diagnosis, and

treatment. While several randomized trials show that dietary advice, glucose monitoring, and insulin can reduce pregnancy complications for women with GDM (Crowther et al., 2005; Landon et al., 2009), the evidence is less clear on whether expanding the number of women diagnosed with GDM improves maternal and infant health at the population level. Hillier et al. (2021) randomly assigned pregnant women to two recommended GDM screening strategies. One strategy nearly doubled the rate of GDM diagnosis but did not produce statistically significant improvements in maternal and perinatal outcomes. Similarly, Crowther et al. (2022) find that using lower diagnostic thresholds increases the number of women diagnosed with GDM but does not significantly reduce large-for-gestational-age births overall.[5] Studies of screening policy show a similar pattern. Riukula (2023) exploits a body mass index cutoff in screening guidelines in Finland and finds that crossing the cutoff induces a 1.5 percentage point increase in GDM diagnosis, while estimates for insulin treatment and adverse birth outcomes are less precise. Nethery et al. (2025) study a change in screening policy in British Columbia and find an increase in GDM managed with lifestyle measures, but no detectable changes in newborn birth weight outcomes. However, some studies find evidence of improvements in birth outcomes. Tennant et al. (2022) use a fuzzy regression discontinuity design at the fasting glucose cutoff of 5.6 mmol/L in the United Kingdom. Using routine maternity records for 7,062 screened pregnancies, they find that diagnosis reduces standardized birth weight and lowers the odds that an infant is large-for-gestational-age. Song, Hurley and Lia (2020) also use a regression discontinuity framework to compare marginally diagnosed and undiagnosed groups in Australia. They find that GDM diagnosis reduces large-for-gestational-age births but increases labor induction and neonatal hypoglycemia. I extend this literature by studying the Carpenter–Coustan fasting threshold among commercially insured women. Using linked claims and laboratory data, I estimate the effects of GDM diagnosis on health care utilization, patient out-of-pocket payments, total pregnancy spending, maternal health, and infant health.

2 Background

Gestational diabetes mellitus (GDM) is hyperglycemia diagnosed during pregnancy in women without known preexisting diabetes. GDM screening generally occurs at 24 to 28 weeks of gestation. Insulin resistance increases during pregnancy, particularly during the second and third trimesters. When pancreatic insulin secretion is insufficient to compensate for this

[5]Large-for-gestational-age generally refers to a birth weight above the 90th percentile for gestational age. At 40 weeks of gestation, this corresponds to approximately 4,057 grams, or about 8 pounds 15 ounces (Duryea et al., 2014).

increase, maternal blood glucose rises (American Diabetes Association Professional Practice Committee, 2026). Elevated maternal glucose is associated with greater fetal growth and a range of adverse maternal and neonatal outcomes, including infant macrosomia and other delivery complications (HAPO Study Cooperative Research Group, 2008). These risks vary continuously with glucose levels. Clinical diagnosis requires a discrete rule for classifying women as having GDM.

Several diagnostic criteria are used, but testing generally follows one of two broad approaches: the one step approach and the two step approach. The one step approach uses a fasting 75 gram oral glucose tolerance test (OGTT) without a preliminary screening test. The two step approach, supported by the American College of Obstetricians and Gynecologists (ACOG), consists of a preliminary screening test followed, when indicated, by a diagnostic test. The preliminary screening test is a nonfasting 50 gram glucose challenge test, with blood drawn one hour after the glucose drink. Clinicians apply a threshold to this screening result to determine whether further testing is needed, but there is no single required cutoff. Commonly used screening thresholds are 130, 135, and 140 mg/dL, and ACOG allows obstetric practices to select a threshold among these values (American College of Obstetricians and Gynecologists, 2018; American Diabetes Association, 2019*b*). If the screening value is below the selected threshold, testing generally stops at this stage.[6] If the screening value meets or exceeds the selected threshold, women are referred for a diagnostic test. This test is a fasting 100 gram OGTT, which measures glucose while fasting and at one, two, and three hours after the glucose drink (American Diabetes Association, 2019*b*).

Two main sets of criteria are used to interpret the diagnostic OGTT: the Carpenter and Coustan criteria and the National Diabetes Data Group (NDDG) criteria. In 2000, the American Diabetes Association (ADA) proposed adopting the lower Carpenter and Coustan thresholds in place of the previously recommended NDDG thresholds (Ferrara et al., 2002). ACOG permits either set of criteria, while the ADA notes advantages to using the lower Carpenter and Coustan thresholds when the two step approach is used (American College of Obstetricians and Gynecologists, 2018; American Diabetes Association, 2019*b*). Under the conventional Carpenter and Coustan criteria, GDM is diagnosed when at least two of the four glucose measurements meet or exceed 95, 180, 155, and 140 mg/dL for the fasting, one hour, two hour, and three hour measurements, respectively (Carpenter and Coustan, 1982; American Diabetes Association, 2019*b*).

Table 1 summarizes the two step screening and diagnostic protocol. Thus, the prelim-

[6]Screening thresholds may vary across practices. Because there is no single cutoff, the initial screening stage is not the primary focus of this study. Nevertheless, using MarketScan data, I estimate the effects of crossing the 130, 135, and 140 mg/dL screening thresholds on the probabilities of proceeding to subsequent diagnostic testing and ultimately receiving a GDM diagnosis. See Appendix A.

inary screening test determines which women are referred for diagnostic testing, while the diagnostic OGTT provides the glucose measurements used to establish a GDM diagnosis. The 95 mg/dL fasting threshold studied in this paper is one of four components of the Carpenter–Coustan diagnostic rule. It is not a standalone diagnostic cutoff. Under the Carpenter–Coustan criteria, GDM is diagnosed when at least two of the four measurements from the 100 gram OGTT are abnormal. The fasting value therefore becomes decisive when exactly one of the other three measurements is also abnormal. A fasting value at or above 95 mg/dL is not sufficient for diagnosis if the other three measurements are normal. Conversely, a woman with fasting glucose below 95 mg/dL can certainly be diagnosed if at least two of the three post-load measurements are abnormal. ACOG also notes that clinicians may diagnose GDM on the basis of a single measurement (American College of Obstetricians and Gynecologists, 2018; American Diabetes Association, 2019*b*). For this reason, crossing the 95 mg/dL threshold raises the probability of receiving a GDM diagnosis but does not determine diagnosis with certainty.

Table 1: Two step protocol for gestational diabetes screening and diagnosis

Step	Measurement	Threshold (mg/dL)
Step 1	1 hour, 50 g glucose challenge test	130 / 135 / 140
Step 2	Fasting, 100 g OGTT	95
	1 hour, 100 g OGTT	180
	2 hour, 100 g OGTT	155
	3 hour, 100 g OGTT	140
Diagnostic rule: Under the Carpenter–Coustan criteria, GDM is diagnosed when at least two of the four Step 2 measurements meet or exceed their corresponding thresholds.		

Notes: Step 1 does not require fasting. Women whose screening result meets or exceeds the selected threshold are referred for the Step 2 diagnostic OGTT, which requires fasting. All glucose thresholds are in mg/dL (American College of Obstetricians and Gynecologists, 2018; American Diabetes Association, 2019*b*; Carpenter and Coustan, 1982).

2.1 The choice of diagnostic thresholds

The Carpenter–Coustan criteria have their roots in the work of O'Sullivan and Mahan (1964). The original O'Sullivan–Mahan thresholds were developed to identify glucose levels during pregnancy that predicted a mother's later risk of diabetes. The NDDG and Carpenter–Coustan criteria later updated these thresholds using different conversions to account for

changes in glucose measurement methods (National Diabetes Data Group, 1979; Carpenter and Coustan, 1982; American Diabetes Association, 2019*b*). This led to different cutoffs. For example, the NDDG fasting threshold is 105 mg/dL, while the Carpenter–Coustan fasting threshold is 95 mg/dL. As a result, the same glucose profile can meet the Carpenter–Coustan criteria but not the NDDG criteria (Harper et al., 2016).

The HAPO (Hyperglycemia and Adverse Pregnancy Outcomes) Cooperative Research Group recruited women at 15 centers in nine countries and administered a 75 gram OGTT between 24 and 32 weeks of gestation. The study found that higher maternal glucose was associated with higher birth weight and other adverse pregnancy outcomes, without a clear threshold at which these risks suddenly increased (HAPO Study Cooperative Research Group, 2008; Farrar, Simmonds, Bryant, Sheldon, Tuffnell, Golder, Dunne and Lawlor, 2016). In 2010, the International Association of Diabetes and Pregnancy Study Groups (IADPSG) Consensus Panel used the HAPO findings to develop a one step diagnostic approach that differed from the NDDG and Carpenter–Coustan criteria. Under the IADPSG approach, women receive a fasting 75 gram OGTT as a one step test. GDM is diagnosed if any one of the fasting, 1 hour, or 2 hour values reaches its threshold (International Association of Diabetes and Pregnancy Study Groups Consensus Panel, 2010).

Professional organizations have adopted different diagnostic rules. In 2013, an NIH consensus panel recommended keeping the two step approach instead of one step approach (National Institutes of Health Consensus Development Conference Panel, 2013). ACOG has continued to support the two step approach, while the ADA's 2019 standards included both the one step and two step strategies (American College of Obstetricians and Gynecologists, 2018; American Diabetes Association, 2019*b*). Recommendations also differ across countries. In the United Kingdom, for example, NICE (National Institute for Health and Care Excellence) recommends a 75 gram OGTT for women with specified risk factors and uses its own fasting and two hour thresholds (National Institute for Health and Care Excellence, 2015). Table 2 summarizes these diagnostic approaches.

2.2 Who is screened

In the United States, guidelines recommend screening all pregnant women who are not known to have diabetes. The U.S. Preventive Services Task Force recommends screening asymptomatic pregnant women at 24 weeks of gestation or later, and ACOG recommends testing at 24 to 28 weeks. Neither exempts women without risk factors from testing (US Preventive Services Task Force et al., 2021; American College of Obstetricians and Gynecologists, 2018). In practice, however, screening is less complete. In a study of laboratory records from one

Table 2: Selected diagnostic approaches and thresholds for GDM

Criteria	Diagnostic OGTT	Fasting	One hour	Two hours	Three hours	Abnormal values required
Carpenter–Coustan	100 grams	95	180	155	140	At least two of four
NDDG	100 grams	105	190	165	145	At least two of four
IADPSG	75 grams	92	180	153	Not used	At least one of three
NICE	75 grams	≈ 101	Not used	≈ 140	Not used	At least one of two

Notes: Glucose values in the table are in mg/dL. A measurement at or above its threshold is abnormal. NICE uses thresholds of 5.6 mmol/L for fasting glucose and 7.8 mmol/L at two hours. I round their mg/dL equivalents for comparison with the other criteria. The 100 gram OGTT usually follows a 50 gram screening test. The table lists the usual number of abnormal measurements required for diagnosis. Clinicians may differ in how they respond to a single abnormal value. (American Diabetes Association, 2019*b*; National Institute for Health and Care Excellence, 2015).

national commercial laboratory, 68 percent of pregnant women aged 25 to 40 were screened, although that figure may understate screening because tests done at other laboratories are not observed (Blatt, Nakamoto and Kaufman, 2011).[7]

The United Kingdom has a different approach. First, NICE recommends a one step approach. It does not use separate screening and diagnosis tests. Second, testing is not universal. NICE recommends a 75 gram OGTT at 24 to 28 weeks only for women with at least one risk factor. These include a body mass index above 30, a previous baby weighing 4.5 kilograms or more, previous GDM, a first degree relative with diabetes, or an ethnic origin with a high prevalence of diabetes. NICE does not recommend using a glucose challenge test, fasting glucose, random glucose, or HbA1c to assess the risk of GDM (National Institute for Health and Care Excellence, 2015). Although the testing is selective based on these risk factors, a large share of pregnant women in the United Kingdom may be tested. One study found that 67.2 percent of women had at least one NICE risk factor (Farrar et al., 2017).

2.3 Care following diagnosis

Initial care typically involves diabetes self-management training (DSMT) and medical nutrition therapy (MNT) (American Diabetes Association Professional Practice Committee, 2026; American College of Obstetricians and Gynecologists, 2018). Diabetes self-management

[7]In MarketScan claims, 77.8 percent of delivered pregnancies among women without prior diabetes have a claim for CPT 82947, 82948, or 82950, and 90.3 percent have a claim under the broader glucose-related testing definition. These measures include tests used for purposes other than GDM screening and can miss tests outside the available claims. See Appendix D for details.

training (DSMT) is a general lifestyle education for managing glucose level in day-to-day lives. This includes learning how to monitor blood glucose with a glucose meter, make healthy food choices, stay physically active, and use medication when needed (Davis et al., 2022). Medical nutrition therapy (MNT) is similar to DSMT, but it focuses more on diet. MNT usually involves a dietitian providing nutrition counseling and meal planning (American Diabetes Association, 2019*a*). If glycemic targets cannot be reached through lifestyle management such as DSMT or MNT, clinicians may add pharmacologic treatment. Several options include insulin, metformin, and glyburide. Insulin is usually preferred first line treatment. The ADA and ACOG generally prefer insulin over metformin or glyburide because insulin does not cross the placenta. However, metformin and glyburide do. Evidence on the long term effects of fetal exposure to metformin and glyburide is limited (American Diabetes Association Professional Practice Committee, 2026; American College of Obstetricians and Gynecologists, 2018). In addition to medication and lifestyle management, treatments can include fetal monitoring, such as nonstress tests or biophysical profiles, to assess fetal well being.

A GDM diagnosis can also affect care after delivery. Women with a history of GDM have a much higher risk of developing type 2 diabetes later in life (Bellamy et al., 2009; Vounzoulaki et al., 2020). Current ADA guidelines recommend a 75 gram OGTT 4 to 12 weeks postpartum and lifelong screening for prediabetes or type 2 diabetes every one to three years (American Diabetes Association Professional Practice Committee, 2026). In practice, however, many women do not complete the recommended postpartum testing (Smirnakis et al., 2005). The main purpose of postpartum screening is to identify persistent dysglycemia (or abnormal blood sugar level) and women who may need continued monitoring or preventive care.

3 Data

I use the Merative MarketScan Commercial Claims and Encounters (CCAE) Database and MarketScan Lab Database from 2015 through September 2025. CCAE contains health insurance claims for employees and their dependents covered by employer sponsored health insurance. The data include inpatient admissions, claims for inpatient and outpatient services, facility claims, and outpatient pharmacy claims. The medical claims provide information on diagnosis and procedure codes, dates of service, and payments made by the insurer and patient. Pharmacy claims provide information on medications with National Drug Code, days supplied, and quantity dispensed. Enrollment records indicate age, sex, geographic region, metropolitan area, and health plan type. They do not include information on race,

ethnicity, education, income, marital status, or parity.

I link these claims to laboratory results from the MarketScan Lab Database. Laboratory results are available for only a subset of CCAE enrollees. For tests that appear in the Lab Database, I observe the test identifier, service date, numerical result, units, and reference range. The Lab Database mainly captures tests ordered in office based settings. I sometimes observe a claim indicating that a glucose test was performed without observing the corresponding glucose value.[8]

3.1 Sample construction and follow-up

To construct the analytic sample, I first identify pregnancy episodes in the CCAE claims from 2015 through 2025.[9] I then link these pregnancies to numerical fasting glucose results in the MarketScan Lab Database for women ages 15 to 55 at the time of testing. This linkage is important because the LOINC code used to identify fasting glucose, 1558-6, is not specific to the fasting portion of a 100 gram diagnostic OGTT. The code captures fasting serum or plasma glucose measurements more generally. I therefore restrict the sample to fasting glucose measurements that can be linked to an identified pregnancy episode. The cohort consists of fasting glucose results linked to claims-defined pregnancy episodes, so it is not restricted to verified Step 2 tests. I examine the diagnostic rule separately in the subset with all four OGTT measurements observed. I also exclude values below 20 mg/dL or above 500 mg/dL as implausible measurements.[10] This leaves 85,611 fasting glucose testing episodes linked to 79,689 pregnancies among 74,194 women. See Table 3 for the baseline characteristics of the final analytic sample. It is predominantly metropolitan, and most are covered by point of service (POS), health maintenance organization (HMO), or preferred provider organization (PPO) plans.

[8] Appendix B compares pregnancies with a glucose testing claim that do and do not have an observed laboratory value, and shows how the medical claims that can be linked to laboratory results differ from those that cannot.

[9] Appendix E describes the ICD-10-CM code families used to identify pregnancy episodes.

[10] The sample may include some women with pregestational diabetes, even though the clinical definition of GDM applies to women without known preexisting diabetes. The results are robust to the exclusion of women with pregestational diabetes. See Appendix C.

Table 3: Baseline characteristics of the analytic sample

Characteristic	Analytic sample
Fasting glucose testing episodes	85,611
Unique women	74,194
Maternal age (years), mean (SD)	32.2 (5.2)
Calendar year of glucose test, mean (SD)	2020.1 (3.1)
Region	
Northeast	26.9%
North Central	8.4%
South	41.2%
West	23.4%
Unknown	0.0%
Missing	0.1%
Plan type	
Comprehensive	0.1%
EPO	0.3%
HMO	24.8%
POS	42.3%
PPO	19.9%
POS with capitation	0.2%
CDHP	7.0%
HDHP	5.1%
Missing	0.3%
Urbanicity	
Metro	92.1%
Non-metro	2.3%
Missing	5.6%

Notes: The unit of observation is a fasting glucose testing episode (Section 3). A woman may contribute episodes from more than one pregnancy. Sex is omitted because the sample includes only women. The table describes the full analytic sample before any bandwidth restriction. The regression discontinuity estimates at the ±10 mg/dL bandwidth use a subset of these episodes. Outcomes requiring an observed delivery, an observable pregnancy spending window, or a linked infant record are estimated in the corresponding subsamples. Table J.1 tests for discontinuities in the probability of entering each subsample at 95 mg/dL. The data do not include race or ethnicity, education, income, marital status, or parity. The health plan reporting flag identifies whether the enrollment record came from a health plan or an employer (Section 3). I omit this flag from the table because it describes the data source. Table I.6 tests its continuity at the cutoff. Region, plan type, and urbanicity use the categories in the MarketScan data dictionary (Merative, 2021 edition). Metro denotes residence in a metropolitan statistical area. EPO: exclusive provider organization; HMO: health maintenance organization; POS: point of service; PPO: preferred provider organization; CDHP: consumer directed health plan; HDHP: high deductible health plan. Author's calculations from MarketScan claims data.

A woman may contribute more than one pregnancy to the sample. Within a pregnancy, fasting glucose measurements more than 30 days apart are treated as separate testing episodes, and I use the earliest measurement in each episode. Of the 79,689 pregnancies in the sample, 5,490, or 6.9 percent, contribute more than one testing episode. Because delivery and infant records are linked to the pregnancy, the same delivery or infant record can be linked to more than one testing episode. I keep these episodes as separate observations and cluster standard errors by woman. The unit of analysis is a testing episode, so pregnancies with repeated tests receive more weight. Clustering by woman accounts for dependence across those observations but does not remove this weighting or the possibility that later tests follow an earlier GDM diagnosis. Fasting glucose is reported in whole mg/dL values, so the running variable is discrete.

I apply different sample restrictions depending on the outcome. For pregnancy spending, I require continuous enrollment from the test through the end of the pregnancy spending window. For delivery outcomes, I require a delivery to be observed in the claims within 294 days of the test.[11] Measures of delivery procedures also require the entire delivery episode to fall within the woman's enrollment period. Delivery payment can be measured only when I observe an inpatient admission record for the delivery. Infant outcomes require both an observed delivery and a linked infant record.

Pharmacy outcomes have their own requirements. I observe medication fills only for women with pharmacy coverage. To identify medication initiation, I also need to know whether the woman used the medication before the test, so I require pharmacy coverage during the 180 days before the test. The resulting samples overlap because the observation requirements are different for each outcome. Appendix F describes the construction of these samples. Appendix G gives the measurement window and code definition for each outcome.

4 Methods

Crossing the 95 mg/dL fasting glucose threshold increases the probability that a woman receives a recorded GDM diagnosis. I use this discontinuous jump in probability to identify the effect of GDM diagnosis in a regression discontinuity (RD) design. I compare women with fasting glucose just below 95 mg/dL with women just above it. The main identifying assumption is that women close to the cutoff have similar underlying health. In the absence of the 95 mg/dL cutoff, the outcomes would have changed smoothly across the threshold. Under this assumption, any discontinuous variation in the outcome at the cutoff can be

[11] A full term pregnancy lasts about 40 weeks. Because the glucose test can occur at different points in pregnancy, I allow up to 294 days, or 42 weeks, for a delivery to appear in the claims.

attributed to crossing the threshold rather than differences in underlying health (Lee and Lemieux, 2010).

This is a fuzzy RD rather than a sharp RD. Crossing the 95 mg/dL cutoff changes the "probability" of receiving GDM diagnosis. It does not determine diagnosis with certainty. This is in part because of the diagnostic rule itself. As described in Section 2, the conventional Carpenter and Coustan criteria require at least two abnormal OGTT measurements. A woman with fasting glucose at or above 95 mg/dL may therefore remain undiagnosed if the other three measurements are normal. A woman below 95 mg/dL can still be diagnosed if at least two of the other three measurements are abnormal. Fasting glucose alone therefore does not determine whether a woman is diagnosed with GDM. Appendix H examines this relationship among testing episodes for which I observe all four OGTT measurements.

I report both reduced form and instrumental variable (IV) estimates. The reduced form measures the effect of crossing the threshold whether or not a GDM diagnosis follows. The IV estimate divides this outcome change by the corresponding change in the probability of GDM diagnosis. Interpreting the IV estimate as the causal effect of diagnosis requires the additional assumptions described below.

4.1 Estimation

Let i index fasting glucose testing episodes, G_i denote the observed fasting glucose value, $X_i = G_i - 95$ denote the centered running variable, and $Z_i = 1\{X_i \geq 0\}$ indicate a value at or above the cutoff. For an outcome Y_i, I estimate the local linear model

$$Y_i = \alpha + \tau Z_i + \beta_1 X_i + \beta_2 (X_i Z_i) + \varepsilon_i, \qquad |X_i| \leq h, \tag{1}$$

where h is the bandwidth. The interaction allows separate slopes on the two sides of the cutoff, and $\widehat{\tau}$ estimates the reduced form discontinuity.

Let D_i denote the recorded GDM diagnosis indicator defined in Appendix G. Replacing Y_i with D_i in Equation 1 gives the first stage, with the coefficient on Z_i denoted by π. I estimate the effect of diagnosis by two stage least squares, instrumenting D_i with Z_i and retaining an intercept, X_i, and $X_i Z_i$ as included regressors. When the outcome and first stage regressions use the same observations, bandwidth, and weights, the IV estimate equals the Wald ratio:

$$\widehat{\theta}_{\text{IV}} = \frac{\widehat{\tau}}{\widehat{\pi}}. \tag{2}$$

I estimate the first stage separately for each outcome's estimation sample. Each IV estimate therefore applies to the compliers in the sample used for that outcome. Appendix

Tables F.1 and F.2 list the samples and first stages behind the principal outcomes at the fixed ±10 mg/dL bandwidth. For the optimal bandwidth, the reduced form and the fuzzy-RD (IV) estimate may use different selected bandwidths. Each IV estimate is scaled by the first stage at its own bandwidth.

4.2 Assumptions

Let $D_i(z)$ denote potential GDM diagnosis status under threshold assignment $z \in \{0, 1\}$, and let $Y_i(d)$ denote the potential outcome under diagnosis status $d \in \{0, 1\}$. I state the assumptions in the conventional continuous score framework, initially treating recorded diagnosis as an accurate measure of clinical diagnosis. Discrete glucose reporting and incomplete diagnosis recording qualify this interpretation, as discussed below.

A1. Continuity of potential outcomes. For each diagnosis status, the conditional expectation of the potential outcome is assumed to be continuous at the cutoff. In particular,

$$\lim_{\delta \downarrow 0} \mathbb{E}[Y_i(d) \mid X_i = -\delta] = \lim_{\delta \downarrow 0} \mathbb{E}[Y_i(d) \mid X_i = \delta], \qquad d \in \{0, 1\}. \tag{3}$$

Potential outcomes may vary with fasting glucose, but they must not exhibit a discrete change at 95 mg/dL when diagnosis status is held fixed. For the fuzzy RD interpretation, I additionally require continuity in the shares of women with each potential diagnosis response, $(D_i(0), D_i(1))$, and in their mean potential outcomes for groups with positive probability at the cutoff. These conditions require that there be no abrupt change in the "underlying composition" of women at the cutoff (Lee and Lemieux, 2010). Directly testing this assumption is not feasible because I am not able to observe the same woman's outcome both with and without a GDM diagnosis. Instead, I look for jumps in predetermined characteristics and other placebo outcomes around the cutoff. Section 4.3 describes these checks.

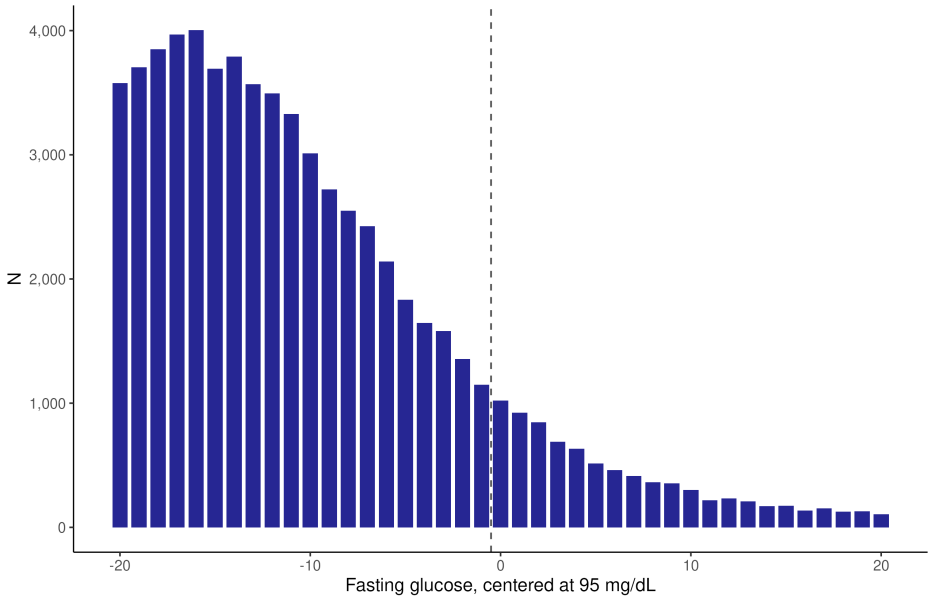
A2. Local support around the cutoff. There must be observations on both sides of 95 mg/dL. In the conventional model with a continuously measured running variable, this condition is expressed as

$$f_X(x) > 0 \qquad \text{for all } x \in (-\eta, \eta) \text{ for some } \eta > 0. \tag{4}$$

Figure 1 shows observations at neighboring values below and above 95 mg/dL. I use separate local linear regressions on either side of the cutoff to estimate the mean of each outcome and the probability of GDM diagnosis at 95 mg/dL. Continuity alone does not

determine how these relationships behave between observed glucose values. Inference must account for possible bias from these linear approximations (Kolesar and Rothe, 2018).

Figure 1: Running-variable distribution near the 95 mg/dL fasting threshold



Notes: Raw count at each integer fasting-glucose value, centered at 95 mg/dL. Author's calculations from MarketScan claims data.

A3. Exclusion restriction. I assume that crossing 95 mg/dL affects outcomes only through GDM diagnosis and its consequences. Let $Y_i(d, z)$ denote the potential outcome under diagnosis status d and threshold assignment z. At the diagnostic margin, the exclusion restriction is

$$Y_i(d, 1) = Y_i(d, 0) = Y_i(d), \qquad d \in \{0, 1\}. \tag{5}$$

A GDM diagnosis triggers nutritional counseling, glucose monitoring, medication, or other care. The assumption would be violated if clinicians responded to the glucose value itself. For example, independently of diagnosis, a clinician might order additional monitoring if fasting glucose reaches 95 mg/dL. Even if the woman's diagnosis remains unchanged, a clinician may order such additional monitoring. If that monitoring affects the outcome, the exclusion restriction would be violated. The assumption still allows a continuous biological relationship between fasting glucose and the outcome. Any discontinuous change at the cutoff must occur through diagnosis and its consequences (Lee and Lemieux, 2010).

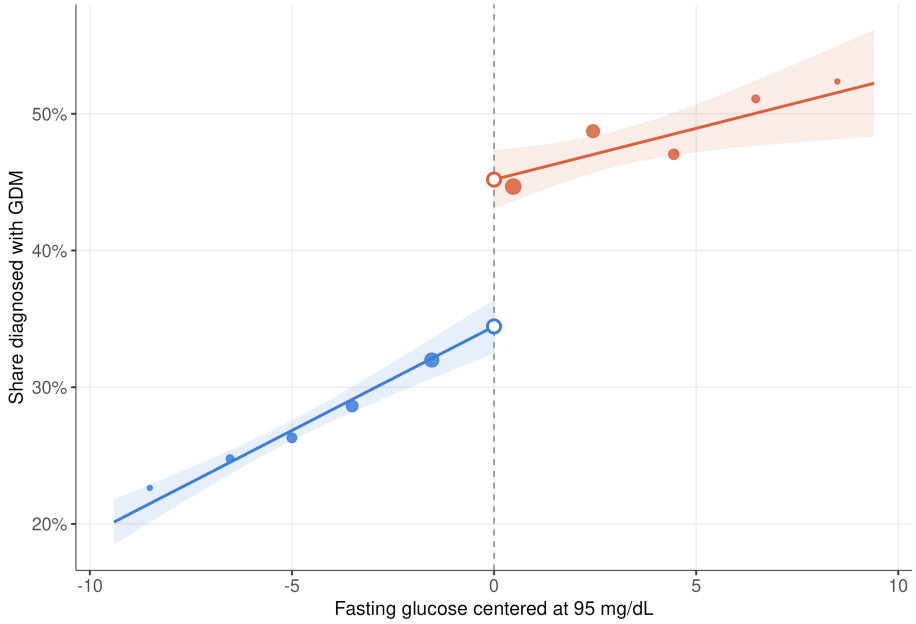
A4. First stage relevance. The probability of receiving a GDM diagnosis must change discontinuously at 95 mg/dL. Given the direction of the diagnostic rule, the relevant condi-

tion is

$$\pi = \lim_{\delta \downarrow 0} \Pr(D_i = 1 \mid X_i = \delta) - \lim_{\delta \downarrow 0} \Pr(D_i = 1 \mid X_i = -\delta) > 0. \tag{6}$$

Section 5.1 and Figure 2 document a positive and statistically significant discontinuity in recorded GDM diagnosis at 95 mg/dL. I assess first stage relevance separately within the sample used for each IV estimate.

Figure 2: First-stage GDM diagnosis jump at the 95 mg/dL fasting threshold



Notes: Points are bin means of the GDM diagnosis rate against the fasting glucose value, centered at 95 mg/dL, full sample of 85,611 fasting glucose testing episodes. Lines are separate local-polynomial fits on either side of the threshold with shaded pointwise 95% confidence bands, estimated with the optimal-bandwidth estimator (triangular kernel, MSE-optimal bandwidth, ±9.4 mg/dL). RD = 10.74 points, $p < 0.001$; the eligible sample is 85,611 testing episodes, of which 23,622 lie within the selected bandwidth; matches the optimal-bandwidth estimate in Table 4. Author's calculations from MarketScan claims data.

A5. Monotonicity. I assume that a woman who would be diagnosed with GDM just below 95 mg/dL would also be diagnosed if her fasting glucose were at or above the cutoff:

$$D_i(1) \geq D_i(0). \tag{7}$$

Crossing the threshold may lead to a diagnosis or leave diagnosis status unchanged. It should not reverse a diagnosis. In other words, there are no defiers. Defiers are women who would be diagnosed with GDM diagnosis below the cutoff but would not be diagnosed if their fasting glucose level were at or above the cutoff.

Under these assumptions in the continuous score framework, the population Wald ratio identifies the local average treatment effect for women whose diagnosis is induced by crossing the threshold:

$$\theta_{\text{LATE}} = \frac{\tau}{\pi} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0),\, X_i = 0]\,. \tag{8}$$

The effect in Equation 8 includes the care induced by diagnosis (Lee and Lemieux, 2010).

Another concern is about the measurement itself. The IV interpretation depends on whether GDM diagnoses appear in MarketScan claims dataset. It is possible that the clinician may verbally tell a pregnant woman that she has GDM, but leaves no corresponding diagnosis record. This would cause the first stage estimate to be underestimated. If the probability of correctly recording a diagnosis is below one and continuous at the cutoff, then the estimated effect of crossing the 95 mg/dL on receiving GDM diagnosis would understate the true estimate. Because the Wald IV estimate divides the reduced form estimate by the first stage estimate, a smaller first stage would make the estimated effect of diagnosis larger in absolute value.[12] The reduced form estimates do not divide by the diagnosis discontinuity and are therefore unaffected by this particular scaling problem.

The main results use local linear regressions with a triangular kernel and an MSE-optimal bandwidth selected separately for each outcome (Calonico, Cattaneo and Titiunik, 2014). I assess sensitivity using a uniform kernel at fixed bandwidths of 5, 10, 15, 20, 25, and 30 mg/dL. The estimates and standard errors reported for the optimal bandwidth are the conventional ones at the selected bandwidth; they are not bias-corrected. The reported conventional confidence intervals do not account for possible approximation bias arising from the integer-valued score. The one exception is the confidence interval for the DSMT/MNT-spending per-complier estimate, where I report the robust bias-corrected interval of Calonico, Cattaneo and Titiunik (2014).

Regression standard errors are clustered by woman to account for dependence across testing episodes and pregnancies contributed by the same individual. This adjustment addresses dependence within women, not the approximation concern described in A2. I omit the IV estimates for an outcome if its first stage F statistic is below 10 at ± 10 mg/dL or at the fixed bandwidth nearest the optimal bandwidth, and at any other bandwidth where the first stage F statistic is below 10; for those outcomes I report reduced form estimates without interpreting IV estimates. An F statistic above 10 does not by itself ensure reliable inference. Reported p values are not adjusted for multiple testing, with one exception: for the ten offspring outcomes at 36 months (asthma, allergic rhinitis, atopic dermatitis, urticaria, an atopic composite, obesity, an obesity composite, body mass index at or above the 95th

[12] Alalouf, Miller and Wherry (2024) discuss a similar measurement concern in their fuzzy RD study of type 2 diabetes at the 6.5% HbA1c threshold.

percentile, hypertension, and type 2 diabetes), I also examine the estimates under Holm and Benjamini and Hochberg adjustments.

4.3 Validity checks Methods

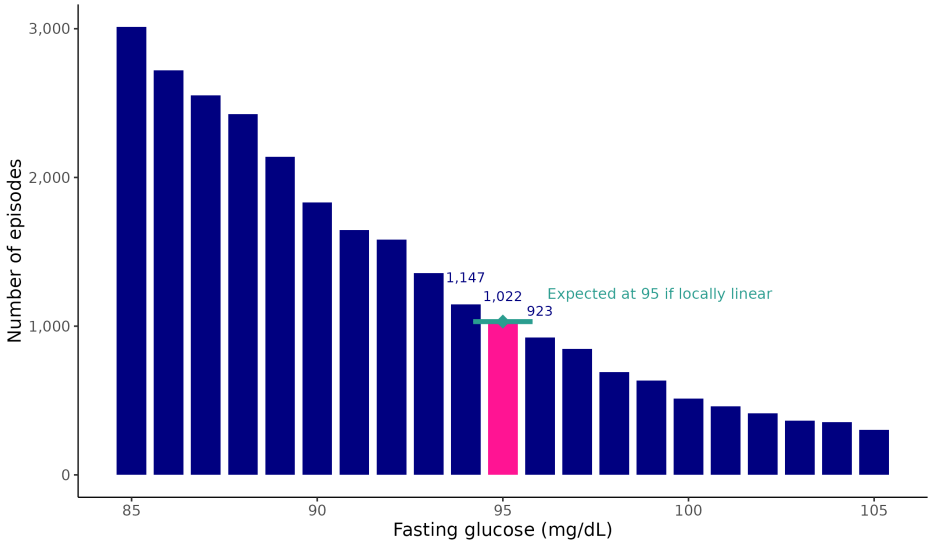
This section provides the methods that I use for validity checks. Results are shown in the next section. One potential threat to identification is the self manipulation problem. Women or clinician could potentially influence which side of the 95 mg/dL cutoff a reported fasting glucose value falls on. If they can manipulate glucose values, then women just above and below the cutoff might differ in ways that also affect their outcomes. Figure 1 shows the distribution of fasting glucose values. I find no clear bunching around the cutoff.

Fasting glucose is reported in integer values. I use the manipulation test developed by Frandsen (2017) for discrete running variables. The test compares the number of observations at 95 mg/dL with those at 94 and 96 mg/dL to see whether the distribution is consistent with no manipulation around the cutoff. This comparison requires a restriction on the local curvature of the probability mass function. The parameter K is an upper bound on the curvature allowed under the null. When $K = 0$, the function is locally linear, and the expected count at 95 mg/dL is one third of the total number of observations at these three values. Larger values of K allow greater departures from local linearity. I report results for $K = 0$, 0.01, 0.02, and 0.05.[13]

As a descriptive benchmark for these choices, a smooth polynomial fit to counts away from the cutoff implies relative curvature of 0.006 to 0.007 at 95 mg/dL. The test does not reject local smoothness at the diagnostic cutoff for any of the reported values of K ($p = 0.746$ at $K = 0$ and $p = 0.755$ at $K = 0.01$). Figure 3 shows the observed count at 95 mg/dL alongside its conditional expectation under local linearity. Table I.3 reports the test results. I also repeat the test at 18 other integer glucose values between 85 and 105 mg/dL. Each test uses three neighboring values without including 95 mg/dL. Appendix Table I.4 reports these tests.

[13] I calculate the binomial p values treating each testing episode as independent. Unlike the regression standard errors, these calculations do not account for dependence between testing episodes from the same woman.

Figure 3: Number of fasting glucose testing episodes at each whole-number fasting glucose value near 95 mg/dL



Notes: The highlighted bar is the cutoff value. The diamond is the count expected at 95 mg/dL if the distribution is locally linear, one third of the 3,092 fasting glucose testing episodes at 94, 95, and 96 mg/dL (1,031); the horizontal bar is the range expected when the relative curvature at the cutoff is at most 0.01 (1,024 to 1,038). The test is reported in Table I.3. Author's calculations from MarketScan claims data.

I also check for discontinuities in predetermined characteristics. I use maternal age, calendar year, region, plan type, urbanicity, and the data contributor indicator as outcome variables in the reduced form model. For categorical variables, I run a separate regression for each category indicator. I also use pregnancy spending before the glucose test as a placebo outcome. Section 5.7 summarizes the results. In Appendix I, Table I.6 reports the estimates for each covariate across bandwidths, and Figure I.4 shows six representative plots.

Lastly, I check whether the probability of entering each outcome sample changes at the cutoff. I test for jumps in the probability that a testing episode has an observed delivery or an observable pregnancy spending window. Among episodes with an observed delivery, I also test for jumps in the probability of observing a delivery payment record or linking an infant record. I then check whether the probability of remaining enrolled through each follow up horizon changes at the cutoff. If crossing the cutoff affects which outcomes I can observe, the estimates may partly reflect differences in who enters the sample.

Because infant outcomes require a linked infant record, I examine selection into this sample in four ways (Appendix J.1). First, I repeat the covariate continuity tests among women with a linked infant. I compare the discontinuities between women with and without a linked infant. Second, I reweight linked observations by the inverse of their estimated

probability of linkage, using characteristics measured before the test to estimate that probability. Third, I use the trimming approach of Lee (2009). I construct approximate bounds on the reduced form effect of crossing the cutoff. I assume that crossing the cutoff cannot prevent an infant from being linked. Under this assumption, some infants linked above the cutoff would not have been linked below it. I remove this excess share from those with the highest outcome values and, in a separate calculation, from those with the lowest. Fourth, I compare estimates for all linked infants with those for infants who have at least one claim during the outcome window. Without a claim, an infant cannot have a recorded outcome code. Reweighting accounts only for differences in observed characteristics. The trimming bounds apply to infants who would be linked on either side of the cutoff. They concern the effect of crossing the cutoff, not the effect of a GDM diagnosis.

For the diagnosis first stage, DSMT/MNT and other medical payments, and the infant outcomes, I repeat the analysis after dropping observations at exactly 95 mg/dL and then excluding progressively wider intervals around the cutoff. This checks how much the estimates depend on observations nearest the threshold. I also estimate discontinuities at placebo cutoffs, using estimation windows that exclude 95 mg/dL, to check whether the specification produces apparent jumps elsewhere. These checks supplement the distributional and covariate tests and do not by themselves rule out sorting. I report the results alongside the main estimates and in the appendices.

5 Results

This section presents the main results on the effects of crossing the 95 mg/dL fasting glucose threshold on recorded GDM diagnosis, subsequent care, health care spending, and maternal and infant outcomes. The "main specification" refers to local linear regression with a triangular kernel and the optimal bandwidth that minimizes estimated mean squared error, following Calonico, Cattaneo and Titiunik (2014). Standard errors are clustered by woman. I assess sensitivity using local-linear regressions with a uniform kernel at fixed bandwidths of 5, 10, 15, 20, 25, and 30 mg/dL. The reduced form estimates the effect of crossing 95 mg/dL. I calculate each IV estimate by dividing the outcome discontinuity by the diagnosis first stage. For the optimal bandwidth, the reduced form and the fuzzy-RD estimate may use different selected bandwidths. For each IV estimate, its numerator and diagnosis first stage use the same observations, bandwidth, and weights. Under the assumptions in Section 4, this ratio estimates the effect of GDM diagnosis among women near the cutoff who are diagnosed because they cross it. The sample for each outcome depends on the enrollment and claims information needed to observe it, as described in Section 3.1. Appendix G provides the full

measurement definitions for every outcome discussed below.

5.1 GDM diagnosis

Figure 2 plots the GDM diagnosis rate against fasting glucose for the full analytic cohort. The diagnosis rate increases discontinuously at the 95 mg/dL cutoff. At the optimal bandwidth, crossing the threshold increases the probability of receiving a GDM diagnosis by 10.74 percentage points. Table 4 reports the first stage estimates across fixed bandwidths of ±5 to ±30 mg/dL, all significant at the 1 percent level.

Table 4: First stage: GDM diagnosis at the 95 mg/dL fasting-glucose cutoff

Specification	Estimate (pp)	(SE)	95% CI	Testing episodes below cutoff	Testing episodes at or above cutoff	Total testing episodes
Local linear, ±5 mg/dL	9.60	(1.81)	[6.06, 13.15]	7,565	4,630	12,195
Local linear, ±10 mg/dL	11.63	(1.27)	[9.15, 14.12]	20,414	6,523	26,937
Local linear, ±15 mg/dL	12.98	(1.06)	[10.90, 15.07]	38,291	7,526	45,817
Local linear, ±20 mg/dL	15.16	(0.95)	[13.29, 17.03]	57,396	8,174	65,570
Local linear, ±25 mg/dL	17.70	(0.89)	[15.95, 19.44]	69,774	8,587	78,361
Local linear, ±30 mg/dL	19.67	(0.85)	[18.00, 21.34]	74,168	8,880	83,048
Optimal bandwidth ($h = 9.4$ mg/dL)	10.74	(1.48)	[7.84, 13.64]	17,401	6,221	23,622

Notes: For each fasting glucose testing episode, the outcome equals one if any GDM diagnosis code is recorded from the test through the later of the first 30 days starting on the test date and delivery, when a delivery is observed, or otherwise through the end of the pregnancy window, and zero otherwise. A diagnosis code can therefore be recorded after delivery. Each row reports a local linear estimate of the effect of crossing 95 mg/dL on diagnosis probability, in percentage points (pp). The fixed bandwidth rows use a uniform kernel, giving equal weight to episodes within the stated bandwidth. For the optimal bandwidth row, I use a triangular kernel and select the bandwidth to minimize the estimated mean squared error (Calonico, Cattaneo and Titiunik, 2014). I report the conventional estimate without bias correction. Standard errors are clustered by woman. The 95% confidence intervals equal the estimate ±1.96 standard errors. Counts give the number of episodes within each row's bandwidth, separately below and at or above 95 mg/dL. The optimal bandwidth is 9.4 mg/dL on each side of the cutoff and includes 23,622 episodes. Bandwidth selection uses the full cohort of 85,611 episodes. Author's calculations from MarketScan claims data.

Samples depend on the records needed to measure each outcome. Pregnancy spending requires an observable spending window, and infant outcomes require a linked infant record. Delivery outcomes require an observed delivery. Cesarean delivery and obstetric anesthesia also require enrollment throughout the delivery episode. Delivery payments require an inpatient record with payment information; women diagnosed with GDM are overrepresented in this sample. Postpartum and later outcomes require continuous enrollment through the follow up window.

Appendix Table G.1 gives the outcome definitions and measurement windows. DSMT

and MNT windows are fixed intervals after the test and continue past delivery. I code these outcomes as zero for women not enrolled in the relevant interval. The DSMT measure for days 90 to 179 can include postpartum claims. Pregnancy payment rows sum plan and patient payments from the test to the end of the pregnancy spending window. The patient payment row includes only the woman's own payments on DSMT/MNT claims.

The first stage is similar in the samples used for the other three outcome families: pregnancy spending, delivery management, and infant health. At each family's own optimal bandwidth, the estimates are 10.72, 11.23, and 11.78 percentage points, respectively (Table I.5). Figure I.3 in Appendix I presents the corresponding plots for these three samples using optimal bandwidths. Additional observation requirements produce narrower estimation samples for medication, delivery payments, and later outcomes. Tables F.1 and F.2 document these requirements and the first stages used for the principal IV estimates. All first stage estimates used for the principal IV results are statistically significant. [14]

5.2 Care utilization

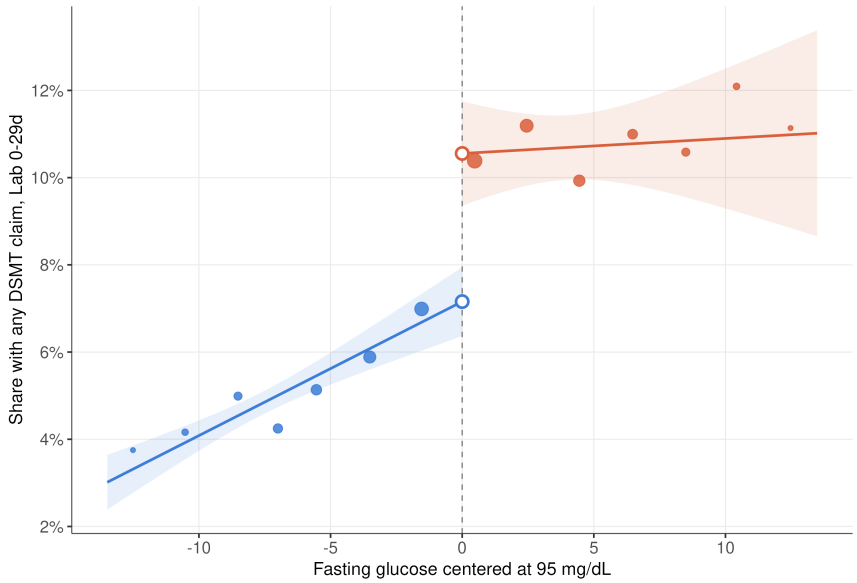
I first examine diabetes education and nutrition claims after the glucose test.

Figure 4 plots the probability of a diabetes self management training (DSMT) claim in the first 30 days after testing. At the optimal bandwidth, crossing the 95 mg/dL threshold increases the probability of a DSMT claim by 3.40 percentage points. Table 5 reports the corresponding reduced form estimates across alternative bandwidths. The response fades to null by 90 days after the lab test. At the optimal bandwidth, the estimates for windows beginning 90 days or later are not statistically significant, as Figure K.1 shows.[15] Figure 5 shows the same pattern for medical nutrition therapy (MNT) claims in the first 30 days after testing. At the optimal bandwidth, crossing the threshold increases the probability of an MNT claim by 1.76 percentage points. Table 6 reports the reduced form estimates for MNT across different bandwidths. I scale these reduced form estimates by the discontinuity in GDM diagnosis to get the per-complier (IV) estimates reported in the same two tables. These show that diagnosis increases the probability of a DSMT claim by 29.1 percentage points and an MNT claim by 15.5 percentage points among women whose diagnosis is induced by crossing the threshold.

[14] Appendix H examines why the first stage is partial using testing episodes in which all four OGTT measurements are observed.

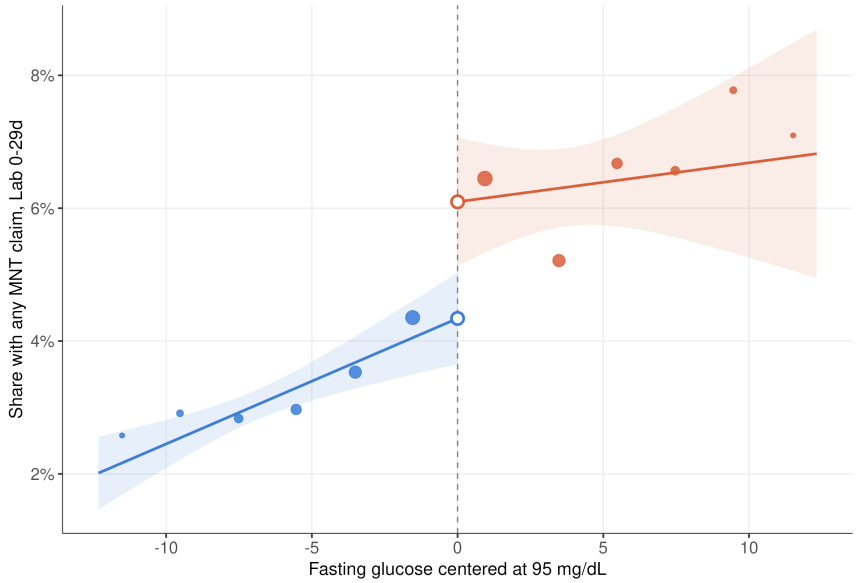
[15] Excluding DSMT claims recorded on the test date leaves the estimate essentially unchanged, at 3.42 percentage points at the optimal bandwidth, showing that the increase is not driven by services billed on the diagnostic test date.

Figure 4: DSMT claims, 0–29 days after the diagnostic lab



Notes: Each point shows the share of testing episodes within a glucose bin that have a DSMT claim in days 0 to 29 after the test. Fasting glucose is centered at 95 mg/dL. Line is a local linear fit with a shaded pointwise 95% confidence band, estimated with the optimal bandwidth of ±13.5 mg/dL and a triangular kernel. The estimated discontinuity is 0.034 ($p < 0.001$).

Figure 5: MNT claims, 0–29 days after the diagnostic lab



Notes: Each point shows the share of testing episodes within a glucose bin that have an MNT claim in days 0 to 29 after the test. Fasting glucose is centered at 95 mg/dL. Line is a local linear fit with a shaded pointwise 95% confidence band, estimated with the optimal bandwidth of ±12.3 mg/dL and a triangular kernel. The estimated discontinuity is 0.018 ($p = 0.004$).

Table 5: Diabetes self-management training, by time since the test

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Any DSMT claim, days 0–29 after test (mean below cutoff: 5.2%)							
Reduced form	2.62	3.32	3.66	3.98	4.76	5.21	3.40
	(p=0.012)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)
Per complier	27.27	28.58	28.20	26.25	26.90	26.51	29.14
	(p=0.007)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							13.5
Any DSMT claim, days 1–29 after test (test date excluded) (mean below cutoff: 5.2%)							
Reduced form	2.59	3.35	3.67	3.98	4.76	5.20	3.42
	(p=0.013)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)
Per complier	26.99	28.82	28.28	26.26	26.88	26.46	29.24
	(p=0.007)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							13.5
Any DSMT claim, days 90–179 after test (mean below cutoff: 0.8%)							
Reduced form	0.05	−0.20	−0.35	−0.37	−0.32	−0.25	−0.20
	(p=0.883)	(p=0.387)	(p=0.091)	(p=0.042)	(p=0.062)	(p=0.125)	(p=0.414)
Per complier	0.49	−1.75	−2.70	−2.45	−1.81	−1.29	−1.76
	(p=0.883)	(p=0.390)	(p=0.094)	(p=0.044)	(p=0.063)	(p=0.126)	(p=0.464)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							10.9
Any DSMT claim, days 180–365 after test (mean below cutoff: 0.3%)							
Reduced form	−0.26	0.04	−0.04	−0.02	0.08	0.13	−0.03
	(p=0.295)	(p=0.829)	(p=0.784)	(p=0.899)	(p=0.531)	(p=0.265)	(p=0.821)
Per complier	−2.69	0.32	−0.31	−0.11	0.44	0.68	−0.93
	(p=0.302)	(p=0.829)	(p=0.784)	(p=0.899)	(p=0.532)	(p=0.267)	(p=0.590)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							15.7

Notes: DSMT: diabetes self-management training. Each outcome is an indicator for a DSMT claim in the stated interval after the glucose test. The row that excludes the test date drops claims billed on the same day as the test. Windows after day 30 assign zero when no claim is observed, including after enrollment or data coverage ends, so they do not establish that actual use converges. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first

stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Table 6: Medical nutrition therapy, by time since the test

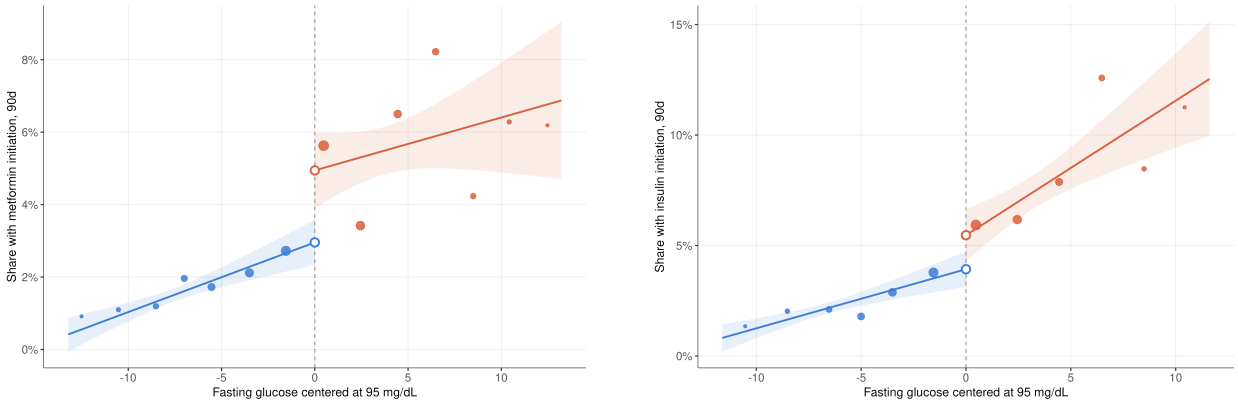
Outcome	±5	±10	±15	±20	±25	±30	Optimal
Any MNT claim, days 0–29 after test (mean below cutoff: 3.2%)							
Reduced form	1.45	1.76	2.07	2.55	2.79	2.87	1.76
	(p=0.079)	(p=0.003)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.004)
Per complier	15.11	15.16	15.95	16.79	15.76	14.59	15.50
	(p=0.072)	(p=0.002)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.006)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							12.3
Any MNT claim, days 90–179 after test (mean below cutoff: 0.9%)							
Reduced form	0.04	−0.17	−0.24	−0.34	−0.33	−0.22	−0.09
	(p=0.919)	(p=0.528)	(p=0.251)	(p=0.085)	(p=0.075)	(p=0.193)	(p=0.760)
Per complier	0.40	−1.44	−1.88	−2.24	−1.84	−1.12	−0.63
	(p=0.919)	(p=0.529)	(p=0.253)	(p=0.087)	(p=0.076)	(p=0.194)	(p=0.813)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							11.1
Any MNT claim, days 180–365 after test (mean below cutoff: 0.8%)							
Reduced form	−0.08	−0.20	−0.18	−0.11	−0.03	0.00	−0.13
	(p=0.831)	(p=0.449)	(p=0.411)	(p=0.562)	(p=0.866)	(p=0.977)	(p=0.601)
Per complier	−0.87	−1.73	−1.39	−0.74	−0.17	−0.02	−1.10
	(p=0.831)	(p=0.450)	(p=0.411)	(p=0.562)	(p=0.866)	(p=0.977)	(p=0.696)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							17.1

Notes: MNT: medical nutrition therapy. Each outcome is an indicator for an MNT claim in the stated interval after the glucose test. Windows after day 30 assign zero when no claim is observed, including after enrollment or data coverage ends, so they do not establish that actual use converges. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth

selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Figure 6 plots the probability of medication initiation against fasting glucose. Panel (a) shows metformin initiation, and panel (b) shows insulin initiation. At each outcome's own optimal bandwidth, crossing 95 mg/dL increases metformin initiation by 1.99 percentage points and any metformin fill by 2.01 points. Insulin initiation increases by 1.54 points, while the 0.87-point estimate for any insulin fill is not statistically significant. Table 7 reports the reduced form and IV estimates across bandwidths. At the optimal bandwidth, diagnosis increases metformin initiation by 15.8 percentage points and any metformin fill by 13.8 percentage points among women whose diagnosis is induced by crossing the threshold, both statistically significant; the insulin per-complier estimates, 12.1 percentage points for initiation and 6.3 points for any fill, are positive but not statistically significant at the optimal bandwidth. The metformin results are robust to bandwidth choice; the insulin reduced form is positive across specifications, but its statistical significance, and that of the corresponding per-complier estimates, is sensitive to bandwidth. Among the other medications examined, glyburide initiation also increases at the cutoff and remains significant at the optimal bandwidth, while GLP-1 receptor agonist fills show no detectable response at any bandwidth.

Figure 6: Diabetes medication initiation, 90 days post-lab



(a) Metformin initiation

(b) Insulin initiation

Notes: Each point shows the share of testing episodes within a glucose bin that start the drug in the 90 days after the test. Fasting glucose is centered at 95 mg/dL. Lines are local linear fits with shaded pointwise 95% confidence bands, estimated with a triangular kernel and an MSE-optimal bandwidth for each panel: (a) ±13.2 and (b) ±11.6 mg/dL. Panel (a) is metformin initiation. The estimated discontinuity is 0.020 ($p = 0.001$). Panel (b) is insulin initiation. The estimated discontinuity is 0.015 ($p = 0.033$).

Table 7: Diabetes medication fills within 90 days of the test

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Metformin, any fill (mean below cutoff: 3.0%)							
Reduced form	0.76	2.31	2.38	2.53	2.88	3.29	2.01
	(p=0.399)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.001)
Per complier	7.70	19.04	17.66	16.46	16.03	16.49	13.82
	(p=0.396)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.037)
N	11,188	24,635	41,901	59,987	71,753	76,016	78,346
Optimal h							13.7
Metformin, initiation (mean below cutoff: 1.7%)							
Reduced form	1.37	2.13	2.49	2.47	2.73	2.99	1.99
	(p=0.119)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.001)
Per complier	14.37	18.25	19.03	17.11	15.74	15.27	15.78
	(p=0.126)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.021)
N	8,494	18,692	31,961	45,715	54,527	57,723	59,433
Optimal h							13.2
Insulin, any fill (mean below cutoff: 2.5%)							
Reduced form	0.18	1.40	1.74	2.58	3.21	3.79	0.87
	(p=0.827)	(p=0.021)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.186)
Per complier	1.86	11.51	12.92	16.79	17.90	18.98	6.34
	(p=0.825)	(p=0.017)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.286)
N	11,188	24,635	41,901	59,987	71,753	76,016	78,346
Optimal h							10.6
Insulin, initiation (mean below cutoff: 2.3%)							
Reduced form	0.87	1.86	2.19	2.74	3.55	4.12	1.54
	(p=0.363)	(p=0.007)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.033)
Per complier	9.16	15.92	16.79	18.97	20.47	21.08	12.14
	(p=0.345)	(p=0.005)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.080)
N	8,494	18,692	31,961	45,715	54,527	57,723	59,433
Optimal h							11.6
Glyburide, initiation (mean below cutoff: 0.9%)							
Reduced form	1.31	1.04	1.32	1.22	1.68	1.82	1.19
	(p=0.028)	(p=0.022)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.007)
Per complier	13.77	8.93	10.08	8.46	9.65	9.28	9.65
	(p=0.037)	(p=0.021)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.025)
N	8,494	18,692	31,961	45,715	54,527	57,723	59,433
Optimal h							13.7

continued on next page

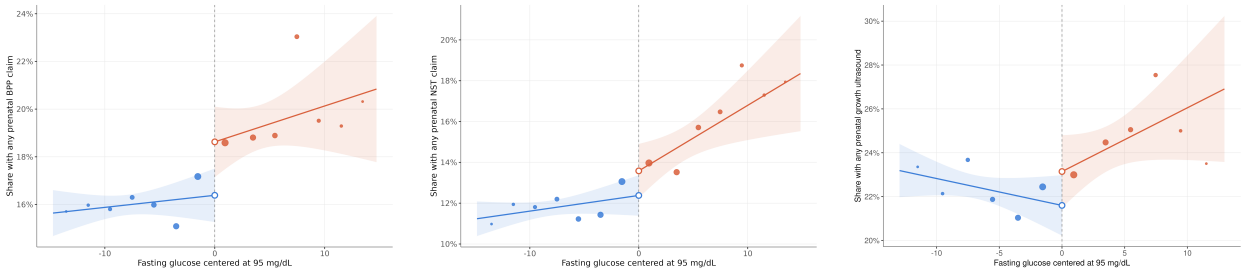
(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
GLP-1 receptor agonist, any fill (mean below cutoff: 0.1%)							
Reduced form	0.14	0.07	0.03	−0.06	−0.14	−0.15	0.10
	(p=0.353)	(p=0.507)	(p=0.785)	(p=0.540)	(p=0.179)	(p=0.136)	(p=0.413)
Per complier	1.40	0.60	0.20	−0.41	−0.80	−0.76	0.89
	(p=0.363)	(p=0.509)	(p=0.785)	(p=0.540)	(p=0.179)	(p=0.136)	(p=0.405)
N	11,188	24,635	41,901	59,987	71,753	76,016	78,346
Optimal h							11.2

Notes: Any fill: at least one pharmacy fill of the drug within 90 days of the test. Initiation: such a fill with no fill of the same drug class in the 180 days before the test. Metformin use increases at the cutoff. The insulin estimates are positive, but their statistical significance depends on the outcome and specification. These claims estimates do not establish the sequence of treatment or the reason an agent was selected. GLP-1 receptor agonist fills are too rare in this population and period to estimate an effect at any bandwidth. The sample for the any-fill outcomes differs from the sample for the initiation outcomes. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Figure 7 plots the probability of each fetal-surveillance test occurring at any point from the test date through delivery, against fasting glucose. Panel (a) shows the probability of any biophysical profile (BPP) claim, panel (b) the probability of any nonstress test (NST) claim, and panel (c) the probability of any follow-up growth ultrasound claim. I observe an upward discontinuity at the 95 mg/dL threshold in panel (a); panels (b) and (c) show no similarly clear increase.

Figure 7: Fetal surveillance response by modality



(a) Any BPP claim, prenatal (b) Any NST claim, prenatal (c) Any growth ultrasound, prenatal

Notes: Each point shows the share of testing episodes within a glucose bin that have a claim for (a) a biophysical profile (BPP), (b) a nonstress test (NST), or (c) a growth ultrasound between the glucose test and delivery. Fasting glucose is centered at 95 mg/dL. Lines are local linear fits with shaded pointwise 95% confidence bands, estimated with a triangular kernel and an MSE-optimal bandwidth for each panel: (a) ±14.7, (b) ±14.8, and (c) ±13.0 mg/dL. The estimated discontinuity is 0.022 ($p = 0.018$) in panel (a), 0.012 ($p = 0.153$) in panel (b), and 0.015 ($p = 0.160$) in panel (c).

Panel A of Table 8 reports estimates across bandwidths for four windows after the glucose test: days 0 to 29, 30 to 89, 90 to 179, and day 180 onward. The increase in BPP claims is concentrated in days 30 to 179, with no detectable increase from day 180 onward. Neither NST nor growth ultrasound shows a statistically significant increase in any window at the optimal bandwidth. NST claims do increase significantly in the first two windows at fixed bandwidths of 25 and 30 mg/dL.

Table 8: Fetal surveillance, by test and event-time window

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Panel A. Any claim, by event-time window (days after the test)							
Any BPP claim, days 0–29 (mean below cutoff: 9.2%)							
Reduced form	0.45	0.94	1.38	0.93	1.39	1.39	1.18
	(p=0.693)	(p=0.235)	(p=0.038)	(p=0.127)	(p=0.014)	(p=0.010)	(p=0.081)
Per complier	4.69	8.12	10.66	6.11	7.83	7.07	7.30
	(p=0.690)	(p=0.228)	(p=0.035)	(p=0.122)	(p=0.012)	(p=0.009)	(p=0.342)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							18.4

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Any BPP claim, days 30–89 (mean below cutoff: 12.5%)							
Reduced form	1.15	1.69	2.10	1.87	1.86	2.11	1.81
	(p=0.385)	(p=0.067)	(p=0.006)	(p=0.007)	(p=0.004)	(p<0.001)	(p=0.037)
Per complier	11.93	14.53	16.19	12.37	10.50	10.75	14.94
	(p=0.371)	(p=0.058)	(p=0.005)	(p=0.005)	(p=0.003)	(p<0.001)	(p=0.093)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							14.3
Any BPP claim, days 90–179 (mean below cutoff: 1.8%)							
Reduced form	1.05	0.77	0.83	0.48	0.52	0.55	0.93
	(p=0.051)	(p=0.048)	(p=0.013)	(p=0.114)	(p=0.068)	(p=0.042)	(p=0.022)
Per complier	10.92	6.62	6.38	3.19	2.91	2.80	9.50
	(p=0.056)	(p=0.047)	(p=0.013)	(p=0.112)	(p=0.066)	(p=0.041)	(p=0.019)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							11.4
Any BPP claim, day 180+ (mean below cutoff: 0.2%)							
Reduced form	0.14	−0.11	−0.05	−0.05	−0.03	0.00	−0.01
	(p=0.427)	(p=0.342)	(p=0.630)	(p=0.537)	(p=0.667)	(p=0.974)	(p=0.951)
Per complier	1.45	−0.97	−0.35	−0.35	−0.19	−0.01	0.58
	(p=0.431)	(p=0.345)	(p=0.630)	(p=0.538)	(p=0.667)	(p=0.974)	(p=0.653)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							11.9
Any NST claim, days 0–29 (mean below cutoff: 5.0%)							
Reduced form	−0.05	0.66	0.32	0.53	0.91	1.14	0.52
	(p=0.956)	(p=0.280)	(p=0.529)	(p=0.251)	(p=0.032)	(p=0.005)	(p=0.339)
Per complier	−0.49	5.64	2.49	3.48	5.16	5.82	4.13
	(p=0.956)	(p=0.276)	(p=0.527)	(p=0.248)	(p=0.031)	(p=0.005)	(p=0.507)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							15.6
Any NST claim, days 30–89 (mean below cutoff: 9.2%)							
Reduced form	0.02	0.75	1.13	0.87	1.31	1.56	0.89
	(p=0.988)	(p=0.357)	(p=0.096)	(p=0.153)	(p=0.021)	(p=0.004)	(p=0.242)
Per complier	0.17	6.43	8.69	5.74	7.39	7.92	7.38
	(p=0.988)	(p=0.348)	(p=0.088)	(p=0.146)	(p=0.018)	(p=0.003)	(p=0.296)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							14.6

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Any NST claim, days 90–179 (mean below cutoff: 1.3%)							
Reduced form	0.68	0.13	0.25	0.16	0.16	0.27	0.37
	(p=0.135)	(p=0.708)	(p=0.376)	(p=0.539)	(p=0.503)	(p=0.245)	(p=0.299)
Per complier	7.11	1.09	1.94	1.06	0.93	1.38	3.65
	(p=0.137)	(p=0.707)	(p=0.374)	(p=0.538)	(p=0.501)	(p=0.244)	(p=0.261)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							11.2
Any NST claim, day 180+ (mean below cutoff: 0.1%)							
Reduced form	0.08	−0.20	−0.07	−0.10	−0.06	0.00	−0.16
	(p=0.546)	(p=0.050)	(p=0.406)	(p=0.191)	(p=0.420)	(p=0.946)	(p=0.068)
Per complier	0.86	−1.72	−0.52	−0.68	−0.32	−0.02	−1.40
	(p=0.548)	(p=0.056)	(p=0.408)	(p=0.194)	(p=0.421)	(p=0.946)	(p=0.083)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							13.3
Any growth ultrasound, days 0–29 (mean below cutoff: 16.2%)							
Reduced form	0.46	0.31	1.00	0.72	1.01	1.07	0.86
	(p=0.741)	(p=0.749)	(p=0.220)	(p=0.329)	(p=0.140)	(p=0.103)	(p=0.298)
Per complier	4.74	2.68	7.72	4.74	5.72	5.43	6.97
	(p=0.739)	(p=0.747)	(p=0.212)	(p=0.324)	(p=0.135)	(p=0.099)	(p=0.485)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							18.4
Any growth ultrasound, days 30–89 (mean below cutoff: 15.4%)							
Reduced form	0.05	0.61	1.07	1.22	0.88	1.10	0.77
	(p=0.973)	(p=0.533)	(p=0.191)	(p=0.095)	(p=0.201)	(p=0.092)	(p=0.434)
Per complier	0.49	5.23	8.21	8.05	4.95	5.60	8.07
	(p=0.973)	(p=0.527)	(p=0.182)	(p=0.088)	(p=0.195)	(p=0.087)	(p=0.407)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							12.4
Any growth ultrasound, days 90–179 (mean below cutoff: 1.9%)							
Reduced form	0.75	0.34	0.43	0.18	0.22	0.27	0.47
	(p=0.174)	(p=0.393)	(p=0.204)	(p=0.564)	(p=0.454)	(p=0.326)	(p=0.234)
Per complier	7.79	2.92	3.32	1.18	1.22	1.38	4.51
	(p=0.177)	(p=0.391)	(p=0.202)	(p=0.563)	(p=0.453)	(p=0.324)	(p=0.259)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							12.8

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Any growth ultrasound, day 180+ (mean below cutoff: 0.2%)							
Reduced form	0.04	0.00	0.08	0.10	0.09	0.10	0.06
	(p=0.824)	(p=0.981)	(p=0.444)	(p=0.310)	(p=0.316)	(p=0.250)	(p=0.636)
Per complier	0.43	−0.03	0.64	0.64	0.51	0.51	0.46
	(p=0.824)	(p=0.981)	(p=0.445)	(p=0.311)	(p=0.317)	(p=0.251)	(p=0.732)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							14.7
Panel B. Counts, rates, and cumulative outcomes							
Number of BPPs, days 90–179 (mean below cutoff: 0.06)							
Reduced form	0.066	0.052	0.055	0.039	0.037	0.039	0.058
	(p=0.009)	(p=0.007)	(p<0.001)	(p=0.010)	(p=0.009)	(p=0.003)	(p=0.003)
Per complier	0.691	0.445	0.421	0.256	0.210	0.198	0.585
	(p=0.014)	(p=0.008)	(p<0.001)	(p=0.010)	(p=0.008)	(p=0.003)	(p=0.003)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							12.0
BPPs per observed week, days 30–89 (mean below cutoff: 0.08)							
Reduced form	0.009	0.020	0.020	0.022	0.022	0.025	0.017
	(p=0.404)	(p=0.011)	(p=0.004)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.019)
Per complier	0.108	0.190	0.167	0.156	0.132	0.139	0.156
	(p=0.377)	(p=0.007)	(p=0.002)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.031)
N	7,626	16,907	28,754	41,296	49,463	52,382	53,941
Optimal h							14.8
Any follow-up growth ultrasound, test to delivery (mean below cutoff: 22.3%)							
Reduced form	−0.06	1.33	2.06	1.72	1.65	1.62	1.54
	(p=0.970)	(p=0.229)	(p=0.027)	(p=0.040)	(p=0.034)	(p=0.030)	(p=0.160)
Per complier	−0.61	11.47	15.86	11.35	9.35	8.26	12.43
	(p=0.970)	(p=0.218)	(p=0.023)	(p=0.035)	(p=0.031)	(p=0.027)	(p=0.272)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							13.0

Notes: BPP: biophysical profile; NST: nonstress test. The windows are measured in days from the glucose test, not from a common gestational age, so they do not establish whether surveillance followed a particular clinical schedule. BPPs per observed week divides the number of BPPs in days 30–89 by the number of weeks of that window before the end of the pregnancy window; it is undefined when the pregnancy window ends before day 30, hence its smaller N. All windows end at delivery or the end of the pregnancy window. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and

Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Panel A of Table 8 reports estimates for BPP, NST, and growth ultrasound across bandwidths in four intervals after the glucose test: days 0 to 29, 30 to 89, 90 to 179, and day 180 onward. Each interval ends at delivery or the end of the pregnancy window. The BPP reduced form estimates are positive at every reported bandwidth in days 30 to 89 and 90 to 179. From day 180 onward, none is statistically significant. Neither NST nor growth ultrasound has a statistically significant increase in any interval at the optimal bandwidth. The NST estimates for days 0 to 29 and 30 to 89 are significant at fixed bandwidths of 25 and 30 mg/dL.

Panel B also reports the probability of any growth ultrasound from the glucose test to the end of the pregnancy window, as plotted in panel (c) of Figure 7. This estimate is positive and statistically significant at fixed bandwidths of 15, 20, 25, and 30 mg/dL, but not at the optimal bandwidth.

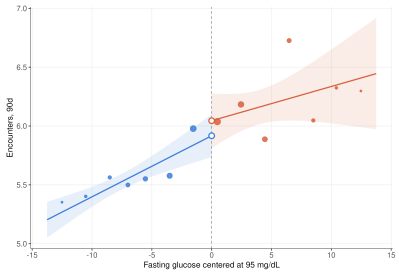
Panel B of Table 8 reports the number of BPPs in days 90–179 and the number of BPPs per observed week. At the optimal bandwidth, crossing the threshold increases the number of BPP claims in days 90 to 179 by 0.058 and the probability of any BPP claim by 0.93 percentage points. Both increases are statistically significant. The corresponding IV estimates imply 0.59 additional BPP claims and a 9.5 percentage point increase in the probability of any BPP claim among women whose diagnosis is induced by crossing the threshold.

The other surveillance estimates are less consistent across specifications. Growth ultrasound increases by 1.54 percentage points at the optimal bandwidth, but the estimate is not statistically significant. The estimates become significant at fixed bandwidths of 15 to 30 mg/dL. Nonstress testing from day 180 onward has a small negative estimate of 0.16 percentage points at the optimal bandwidth, which is not statistically significant.

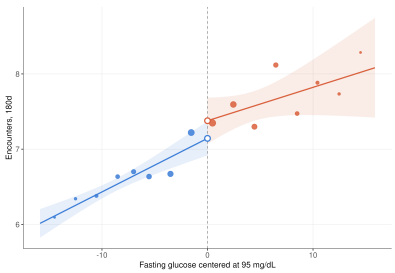
For the fetal surveillance outcomes, I count only claims dated before the pregnancy window ends (Appendix G). If that window has ended by the start of an interval, I keep the testing episode in the sample and set both the claim indicator and count to zero. Panel B of Table 8 reports BPPs per observed week in days 30 to 89, using the narrower sample with pregnancy time remaining in that interval.

Figure 8 examines whether the increases in specific services are accompanied by broader changes in health care use. Panels (a) and (b) plot the number of distinct days with a medical encounter over 90 and 180 days after testing. Panels (c) and (d) show days with a numerical laboratory result, and panels (e) and (f) show days with a pharmacy fill over the same horizons.

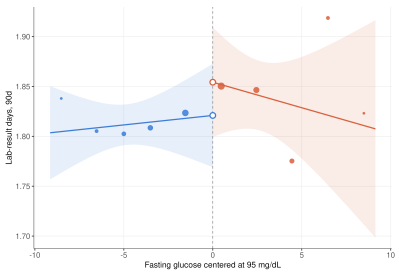
Figure 8: General care intensity: all seven outcomes



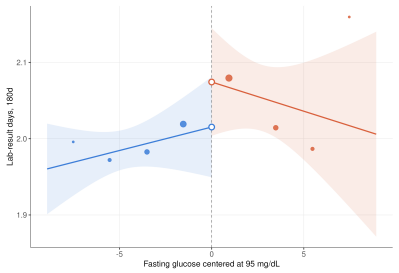
(a) Encounters, 90d



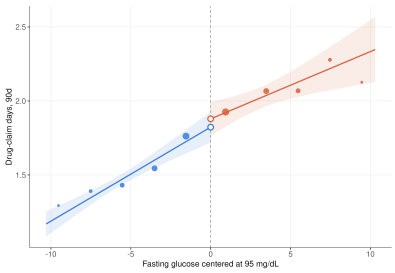
(b) Encounters, 180d



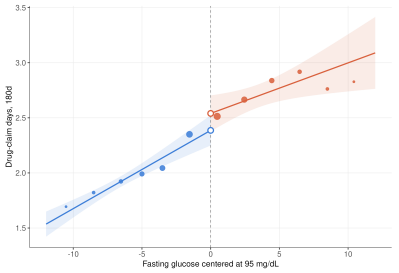
(c) Lab-result days, 90d



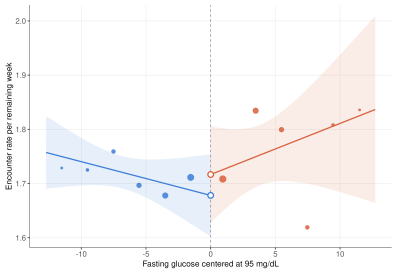
(d) Lab-result days, 180d



(e) Drug-claim days, 90d



(f) Drug-claim days, 180d



(g) Encounter rate per remaining week

Notes: Each point shows the mean number of distinct claim days among testing episodes within a glucose bin: (a) medical encounter days in the 90 days after the test, (b) medical encounter days in the 180 days after the test, (c) laboratory result days in 90 days, (d) laboratory result days in 180 days, (e) drug claim days in 90 days, (f) drug claim days in 180 days, and (g) encounter days per remaining week of pregnancy. Each window ends at delivery or the end of the pregnancy window if that comes first. Fasting glucose is centered at 95 mg/dL. Lines are local linear fits with shaded pointwise 95% confidence bands, estimated with a triangular kernel and an MSE-optimal bandwidth for each panel: (a) ± 13.8, (b) ± 15.9, (c) ± 9.1, (d) ± 8.9, (e) ± 10.3, (f) ± 12.0, and (g) ± 12.7 mg/dL. The estimated discontinuities are 0.128 ($p = 0.393$) in panel (a), 0.234 ($p = 0.228$) in (b), 0.033 ($p = 0.387$) in (c), 0.059 ($p = 0.230$) in (d), 0.056 ($p = 0.477$) in (e), 0.154 ($p = 0.157$) in (f), and 0.039 ($p = 0.518$) in (g).

Panel (g) shows medical encounter days per remaining week of pregnancy. Encounter days and pharmacy fill days generally rise with fasting glucose, but the fitted values differ little at 95 mg/dL. I also see no clear jump in laboratory result days or the encounter rate. All count windows end no later than the pregnancy observation window.

Table 9 reports the reduced form estimates for these seven measures at the optimal bandwidth. None is statistically significant. Over 90 days, crossing the threshold increases medical encounter days by 0.13 and pharmacy fill days by 0.06. The table also reports the corresponding IV estimates of 0.43 and 0.68 days, respectively, neither of which is statistically significant.

Table 9: General health-care use after the test

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Days with a medical encounter, 90 days after test (mean below cutoff: 5.59)							
Reduced form	−0.037	0.138	0.175	0.222	0.274	0.345	0.128
	(p=0.867)	(p=0.373)	(p=0.179)	(p=0.057)	(p=0.013)	(p<0.001)	(p=0.393)
Per complier	−0.387	1.189	1.350	1.468	1.547	1.753	0.429
	(p=0.868)	(p=0.368)	(p=0.174)	(p=0.055)	(p=0.012)	(p<0.001)	(p=0.791)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							13.8
Days with a medical encounter, 180 days after test (mean below cutoff: 6.70)							
Reduced form	−0.033	0.234	0.268	0.297	0.351	0.481	0.234
	(p=0.916)	(p=0.274)	(p=0.135)	(p=0.065)	(p=0.021)	(p<0.001)	(p=0.228)
Per complier	−0.340	2.012	2.062	1.962	1.983	2.444	1.366
	(p=0.917)	(p=0.269)	(p=0.131)	(p=0.063)	(p=0.020)	(p<0.001)	(p=0.545)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							15.9
Days with a laboratory result, 90 days after test (mean below cutoff: 1.82)							
Reduced form	0.046	0.035	−0.011	−0.026	−0.022	−0.008	0.033
	(p=0.320)	(p=0.289)	(p=0.700)	(p=0.312)	(p=0.350)	(p=0.718)	(p=0.387)
Per complier	0.476	0.300	−0.082	−0.169	−0.126	−0.042	0.299
	(p=0.333)	(p=0.294)	(p=0.700)	(p=0.312)	(p=0.349)	(p=0.718)	(p=0.459)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							9.1
Days with a laboratory result, 180 days after test (mean below cutoff: 1.99)							
Reduced form	0.071	0.060	0.010	−0.010	−0.014	0.008	0.059
	(p=0.214)	(p=0.143)	(p=0.766)	(p=0.757)	(p=0.633)	(p=0.783)	(p=0.230)
Per complier	0.744	0.518	0.079	−0.064	−0.079	0.039	0.535
	(p=0.228)	(p=0.149)	(p=0.766)	(p=0.757)	(p=0.633)	(p=0.783)	(p=0.295)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							8.9

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Days with a pharmacy fill, 90 days after test (mean below cutoff: 1.44)							
Reduced form	−0.040	0.123	0.183	0.262	0.336	0.380	0.056
	(p=0.690)	(p=0.084)	(p=0.002)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.477)
Per complier	−0.420	1.054	1.413	1.727	1.901	1.934	0.682
	(p=0.695)	(p=0.079)	(p=0.002)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.304)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							10.3
Days with a pharmacy fill, 180 days after test (mean below cutoff: 1.94)							
Reduced form	−0.007	0.195	0.244	0.323	0.410	0.466	0.154
	(p=0.963)	(p=0.063)	(p=0.006)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.157)
Per complier	−0.072	1.678	1.882	2.133	2.319	2.368	1.210
	(p=0.963)	(p=0.060)	(p=0.005)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.227)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							12.0
Encounter days per remaining week of pregnancy (mean below cutoff: 1.72)							
Reduced form	−0.038	0.055	0.012	0.010	−0.005	0.016	0.039
	(p=0.656)	(p=0.359)	(p=0.819)	(p=0.829)	(p=0.908)	(p=0.689)	(p=0.518)
Per complier	−0.397	0.475	0.089	0.065	−0.027	0.082	0.193
	(p=0.660)	(p=0.355)	(p=0.818)	(p=0.828)	(p=0.908)	(p=0.688)	(p=0.775)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							12.7

Notes: Outcomes count distinct days with any medical encounter, any laboratory result, or any pharmacy fill, and are therefore unaffected by prices. Windows are truncated at the end of the pregnancy window. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

The estimates for encounter days and pharmacy fill days are sensitive to bandwidth choice. Figure K.5 in Appendix K plots the estimates for selected utilization measures across fixed bandwidths. At the optimal bandwidth, neither encounter days nor pharmacy

fill days is statistically significant. The estimated increases in pharmacy fill days become statistically significant at bandwidths of 15 mg/dL and above, while the increase in medical encounter days over 180 days becomes significant at bandwidths of 25 mg/dL and above.

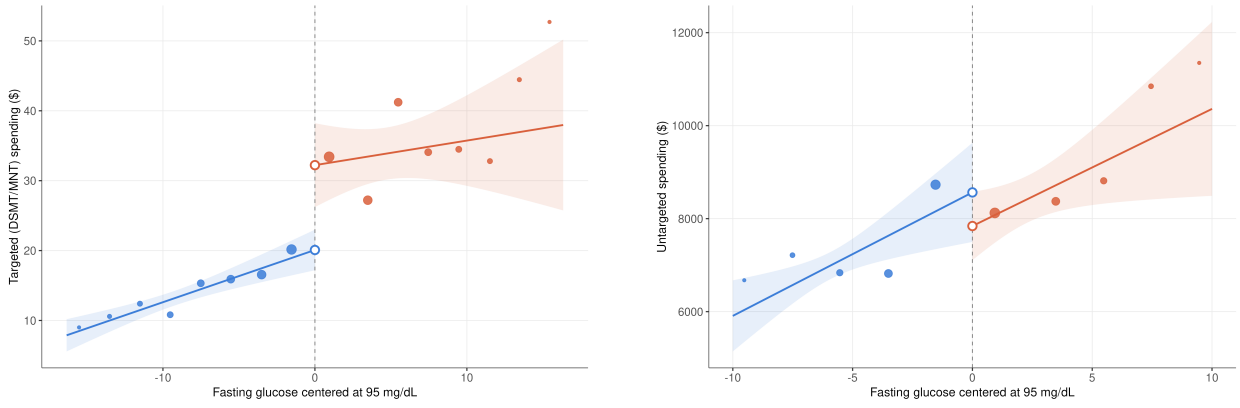
Table K.1 reports estimates for the probability of contact with providers in eight specialty groups. I find no statistically significant change in any group at the optimal bandwidth.

5.3 Pregnancy spending

I next examine whether spending also changes at the cutoff. I measure payments from the glucose test through the day before the pregnancy observation window ends. When a delivery is observed, I exclude claims dated on the delivery date. Delivery admission payments are examined separately in Section 5.5. Appendix G describes the spending windows and payment measures. The eligible sample includes 72,685 fasting glucose testing episodes with continuous enrollment from the test through the end of the spending window. All amounts are reported in nominal dollars.

Figure 9 plots average payments against fasting glucose. Panel (a) shows payments on DSMT and MNT claims, and panel (b) shows payments on other medical claims. Payments for DSMT and MNT jump upward at 95 mg/dL, where recorded use of these services also increases.

Figure 9: Pregnancy spending, DSMT/MNT vs. other medical claims



(a) DSMT/MNT

(b) Total minus DSMT/MNT

Notes: Each point shows mean total spending (plan-paid plus out-of-pocket), in dollars, among testing episodes within a glucose bin: (a) on DSMT/MNT claims and (b) on all other claims, from the test through the end of the pregnancy spending window. Fasting glucose is centered at 95 mg/dL. Lines are local linear fits with shaded pointwise 95% confidence bands, estimated with a triangular kernel and an MSE-optimal bandwidth for each panel: (a) ±16.3 and (b) ±10.0 mg/dL. The estimated discontinuity is \$12.14 ($p < 0.001$) in panel (a) and −\$723 ($p = 0.273$) in panel (b).

Panel A of Table 10 reports the reduced form and IV estimates for total payments and

patient payments on DSMT and MNT claims. At the optimal bandwidth, crossing the threshold increases total payments on these claims by $12.14. The corresponding IV estimate is $107 per additional diagnosis. Patient out of pocket payments increase by $3.46 in reduced form, with an IV estimate of $28 per additional diagnosis.[16] Both payment measures increase significantly. These estimates apply only to education and nutrition services. They do not measure the full cost of care induced by diagnosis or the woman's total out of pocket spending.

Table 10: Payments from the test through delivery: DSMT/MNT claims, non-DSMT/MNT, and total

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Panel A. Payments for DSMT/MNT claims							
DSMT/MNT claims: total payment (plan plus patient) (mean below cutoff: $14.97)							
Reduced form	$11.84	$11.24	$11.70	$12.58	$14.93	$15.65	$12.14
	(p=0.016)	(p=0.002)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)
Per complier	$123.76	$99.06	$92.13	$84.81	$84.02	$79.84	$107.11
	(p=0.016)	(p=0.002)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.002)
N	10,421	22,897	38,886	55,621	66,514	70,460	72,685
Optimal h							16.3
DSMT/MNT claims: patient out-of-pocket payment (mean below cutoff: $4.28)							
Reduced form	$2.29	$3.22	$3.62	$3.94	$4.78	$4.73	$3.46
	(p=0.264)	(p=0.033)	(p=0.007)	(p=0.001)	(p<0.001)	(p<0.001)	(p=0.016)
Per complier	$23.90	$28.36	$28.48	$26.54	$26.88	$24.11	$28.28
	(p=0.255)	(p=0.031)	(p=0.006)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.047)
N	10,421	22,897	38,886	55,621	66,514	70,460	72,685
Optimal h							15.7
Panel B. Non-DSMT/MNT and total pregnancy payments							
Non-DSMT/MNT medical claims: total payment (mean below cutoff: $7,115)							
Reduced form	−$1,062	−$480.43	$472.10	$701.01	$828.09	$1,082	−$723.33
	(p=0.215)	(p=0.390)	(p=0.299)	(p=0.067)	(p=0.022)	(p=0.002)	(p=0.273)
Per complier	−$11,106	−$4,233	$3,717	$4,727	$4,657	$5,516	−$7,893
	(p=0.247)	(p=0.402)	(p=0.293)	(p=0.063)	(p=0.020)	(p=0.002)	(p=0.257)
N	10,421	22,897	38,886	55,621	66,512	70,458	72,683
Optimal h							10.0

continued on next page

[16]For this IV estimate, robust bias corrected inference also indicates a statistically significant increase. The other optimal bandwidth comparisons use conventional inference, as described in Section 4.

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Total pregnancy payment, including pharmacy claims (mean below cutoff: $7,565)							
Reduced form	−$1,010	−$406.36	$495.81	$750.14	$916.64	$1,210	−$569.07
	(p=0.249)	(p=0.482)	(p=0.294)	(p=0.060)	(p=0.015)	(p<0.001)	(p=0.383)
Per complier	−$10,557	−$3,581	$3,903	$5,059	$5,155	$6,172	−$6,328
	(p=0.279)	(p=0.490)	(p=0.287)	(p=0.056)	(p=0.013)	(p<0.001)	(p=0.351)
N	10,421	22,897	38,886	55,621	66,512	70,458	72,683
Optimal h							10.5
Total pregnancy payment, medical claims only (mean below cutoff: $7,130)							
Reduced form	−$1,050	−$469.19	$483.80	$713.59	$843.01	$1,097	−$709.33
	(p=0.220)	(p=0.402)	(p=0.288)	(p=0.063)	(p=0.020)	(p=0.002)	(p=0.282)
Per complier	−$10,982	−$4,134	$3,809	$4,812	$4,741	$5,596	−$7,778
	(p=0.252)	(p=0.413)	(p=0.281)	(p=0.058)	(p=0.018)	(p=0.001)	(p=0.264)
N	10,421	22,897	38,886	55,621	66,512	70,458	72,683
Optimal h							10.0
Pharmacy claims: total payment (mean below cutoff: $435.02)							
Reduced form	$40.68	$62.83	$12.01	$36.55	$73.81	$113.07	$81.18
	(p=0.693)	(p=0.492)	(p=0.886)	(p=0.633)	(p=0.312)	(p=0.119)	(p=0.373)
Per complier	$425.30	$553.60	$94.52	$246.49	$415.36	$576.81	$771.74
	(p=0.694)	(p=0.494)	(p=0.886)	(p=0.633)	(p=0.313)	(p=0.120)	(p=0.392)
N	10,421	22,897	38,886	55,621	66,514	70,460	72,685
Optimal h							10.8

Notes: DSMT: diabetes self management training; MNT: medical nutrition therapy. Payments are in nominal dollars and cover claims from the test through the day before the pregnancy window ends. For observed deliveries, the last included date is the day before delivery. Total payments combine plan and patient payments. DSMT/MNT payments use claims with either procedure code; patient out of pocket payments include only the patient's own payments on those claims. Other medical payments cover outpatient, facility, and inpatient claims, excluding DSMT/MNT and pharmacy claims. Total pregnancy payments include all medical and pharmacy claims. The medical only total excludes pharmacy claims. I assign zero pharmacy spending to episodes with a matched enrollment record that carries the drug coverage field and no pharmacy claim in the spending window. This rule does not verify affirmative pharmacy coverage.

Reduced form rows estimate the effect of crossing 95 mg/dL. Per complier rows estimate the effect of GDM diagnosis for women whose diagnosis is induced by crossing the cutoff. Each per complier estimate divides the reduced form by the diagnosis first stage using the same sample, bandwidth, and weights. The p values appear in parentheses. Columns headed $\pm h$ use local linear regressions with a uniform kernel. The last column uses a triangular kernel and an MSE optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014). I report conventional point estimates. Per complier estimates in the last column use the bandwidth selected for the fuzzy design.

I omit all per complier estimates for an outcome if its first stage F statistic is below 10 at ±10 mg/dL or at

the fixed bandwidth nearest the optimal bandwidth. Otherwise, I omit an estimate at any bandwidth where its own first stage F statistic is below 10. Standard errors are clustered by woman. In the fixed bandwidth columns, N counts testing episodes within the bandwidth. In the last column, N is the full eligible sample. The reported h is the optimal bandwidth. Author's calculations from MarketScan claims data.

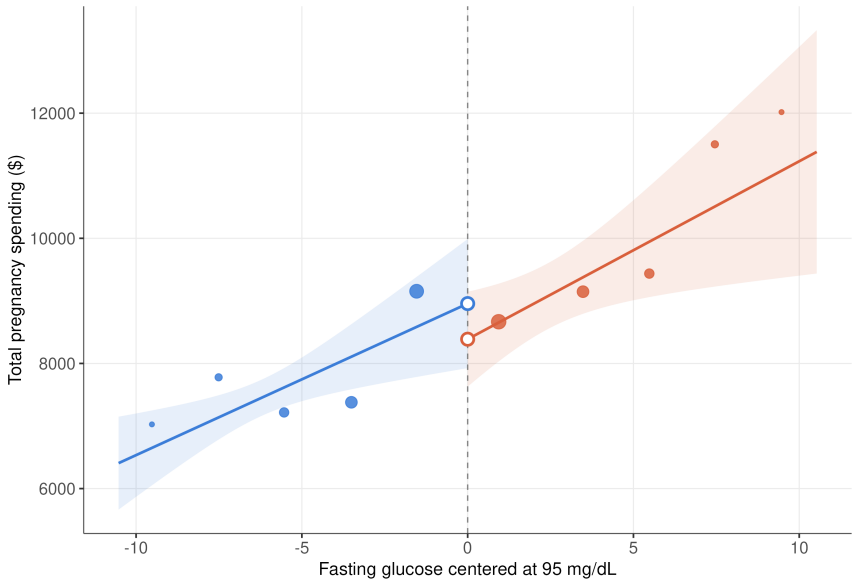
Panel (a) of Figure L.1 in Appendix L shows the reduced form estimates for DSMT/MNT payments across fixed bandwidths. The estimates are positive and statistically significant at every bandwidth examined from 5 to 30 mg/dL. The corresponding IV estimates range from \$80 to \$124 per additional diagnosis.

Excluding observations near the cutoff changes the DSMT/MNT payment estimates. At a fixed bandwidth of 10 mg/dL, dropping the 877 episodes recorded at exactly 95 mg/dL lowers the estimated increase from \$11.24 to \$7.54, and it is no longer statistically significant. Excluding wider intervals around the cutoff also leaves positive but statistically insignificant estimates. At a bandwidth of 20 mg/dL, the estimates range from \$10.29 to \$17.45 and remain significant at every exclusion radius examined.

The placebo tests detect positive discontinuities at 70 mg/dL in two specifications, with estimates of \$4.33 and \$5.19. These are smaller than the corresponding estimates at 95 mg/dL, but positive jumps are not confined to the GDM cutoff. The sensitivity to exclusions near the cutoff and the placebo results leave uncertainty about the size of the DSMT/MNT payment increase.

Panel (b) of Figure 9 shows average payments on medical claims other than DSMT and MNT, excluding pharmacy claims. The fitted line drops at 95 mg/dL. Payments are more dispersed above the cutoff, however, and the size of the drop is imprecisely estimated. Figure 10 plots total payments on medical and pharmacy claims. These payments generally rise with fasting glucose, but I see no clear upward jump at 95 mg/dL.

Figure 10: Total spending, all claim types



Notes: Bin means of total spending (plan-paid plus out-of-pocket) after the glucose test and before the pregnancy endpoint (every claim type, including drugs) against the fasting glucose value, centered at 95 mg/dL. Line is a local-polynomial fit on each side of the threshold with a shaded pointwise 95% confidence band, estimated with the optimal-bandwidth estimator (triangular kernel, MSE-optimal bandwidth, ±10.5 mg/dL). RD = −$569, $p = 0.383$, the optimal-bandwidth estimate reported in the text. The eligible sample contains $N = 72{,}683$ testing episodes; this row does not require an observed pharmacy claim (see Appendix G). The estimate uses the episodes inside its own bandwidth. Table 10 reports bandwidth stability, including the drug-free version of this total. Author's calculations from MarketScan claims data.

Panel B of Table 10 reports the estimates for these broader payment measures. At the optimal bandwidth, total payments including pharmacy change by −$569 in reduced form, with a corresponding IV estimate of −$6,328. Payments on other medical claims change by −$723, with an IV estimate of −$7,893. None of these estimates is statistically significant. The estimates remain compatible with substantial savings or spending increases. The absence of a detectable increase therefore does not establish that the total spending effect is small.

Table 10 reports medical and pharmacy payments in dollars. Appendix Table L.1 in Appendix L reports estimates using alternative transformations of spending. Including DSMT and MNT but excluding pharmacy gives a reduced form of −$709, while pharmacy payments alone increase by $81. Neither dollar estimate is statistically significant. When I use the inverse hyperbolic sine (IHS) of pharmacy payments as the outcome, the estimate is 0.20 at the optimal bandwidth and statistically significant. This coefficient measures a change on the IHS scale and cannot be interpreted as a percentage change in total spending.

Panel (b) of Figure L.1 plots the reduced form estimates for other medical payments across fixed bandwidths. In contrast to DSMT/MNT payments in panel (a), these estimates

change sign as the bandwidth widens. They are negative at bandwidths of 5 and 10 mg/dL, become positive at 15 mg/dL, and are statistically significant at 25 and 30 mg/dL.

Table 10 reports this bandwidth comparison for DSMT/MNT payments, other medical payments, and the totals with and without pharmacy claims. Total payments follow the same pattern as other medical payments, with estimated increases of $917 and $1,210 at bandwidths of 25 and 30 mg/dL, respectively. Both estimates are statistically significant. At a bandwidth of 20 mg/dL, the estimate for other medical payments rises as I exclude wider intervals around the cutoff. The dollar estimates depend on both the bandwidth and the inclusion of observations near 95 mg/dL.

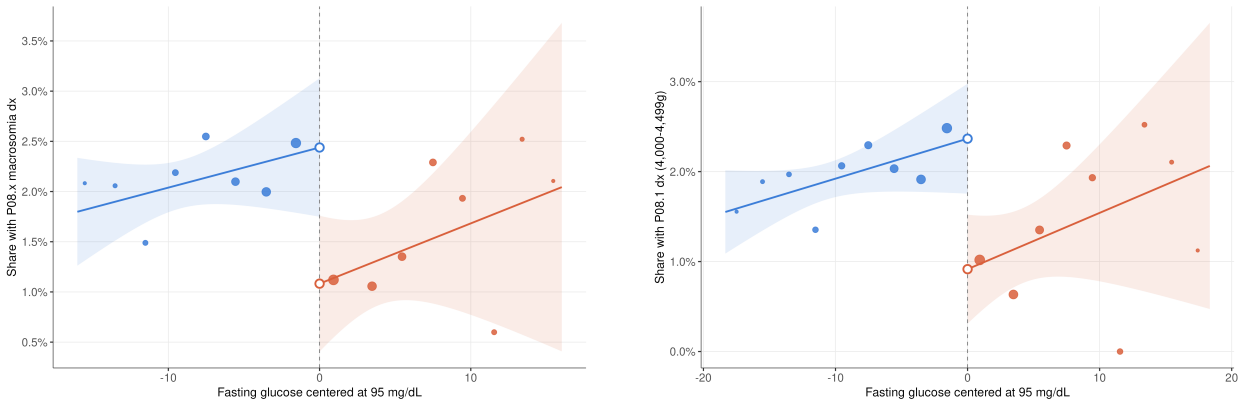
Using $\log(1 + y)$ or IHS for other medical payments gives no statistically significant change at any of the bandwidths examined. These estimates are too imprecise to conclude that spending is unchanged. None of the 33 placebo estimates for other medical payments is statistically significant. The donut and placebo checks cover total payments on DSMT/MNT claims and on other medical claims. They do not cover overall payments, pharmacy payments alone, or patient payments alone. I find an increase in payments for diabetes education and nutrition services, but cannot determine the size or direction of the broader spending effect.

5.4 Infant health

I next examine whether crossing the threshold changes infant outcomes recorded in the claims. The eligible sample includes 37,464 fasting glucose testing episodes with a linked infant record. I measure diagnoses and admissions using the infant's claims from one day before through 30 days after delivery. These measures do not directly capture birth weight or clinical severity. Infant linkage also increases at the cutoff, so the results may partly reflect differences in which infants enter the sample. I examine this selection below.

Figure 11 plots the probability of a recorded macrosomia diagnosis against maternal fasting glucose. Panel (a) combines codes P08.0 and P08.1 and shows a downward jump at 95 mg/dL. Panel (b) uses P08.1 alone for infants coded as heavy for gestational age. The similar drop in this panel indicates that P08.1 accounts for the decline in the combined measure.

Figure 11: Infant macrosomia (ICD-10 P08.0 and P08.1) at the fasting threshold



(a) P08.0 and P08.1, combined

(b) P08.1 (4,000–4,499g)

Notes: Each point shows the share of testing episodes with a linked infant, within a glucose bin, in which the infant has a claim with a diagnosis code for (a) P08.0 or P08.1 combined or (b) P08.1 (4,000 to 4,499 g). The codes proxy for a heavy newborn, and birth weight is not measured. Fasting glucose is centered at 95 mg/dL. Lines are local linear fits with shaded pointwise 95% confidence bands, estimated with a triangular kernel and an MSE-optimal bandwidth for each panel: (a) ±16.0 and (b) ±18.3 mg/dL. The estimated discontinuity is −0.014 ($p = 0.006$) in panel (a) and −0.015 ($p < 0.001$) in panel (b).

Table 11 reports the reduced form estimates for the combined indicator and each code separately. At the optimal bandwidth, crossing the threshold reduces the probability of a recorded macrosomia diagnosis by 1.36 percentage points. This reduction is statistically significant. The corresponding IV estimate in Table 11 is a reduction of 10.8 percentage points among women whose GDM diagnosis is induced by crossing the threshold. For P08.1 alone, the reduced form and IV estimates imply reductions of 1.45 and 12.6 percentage points, respectively. P08.0 is too uncommon to estimate its response precisely.

Table 11: Infant outcomes: macrosomia and NICU admission

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Macrosomia (P08.0 or P08.1) (mean below cutoff: 2.3%)							
Reduced form	−1.66	−1.29	−1.37	−1.42	−1.36	−1.47	−1.36
	(p=0.066)	(p=0.021)	(p=0.003)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.006)
Per complier	−16.09	−10.40	−9.75	−8.45	−6.59	−6.17	−10.81
	(p=0.093)	(p=0.028)	(p=0.005)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.035)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							16.0

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Heavy for gestational age, 4,000–4,499 g (P08.1) (mean below cutoff: 2.1%)							
Reduced form	−1.82	−1.51	−1.46	−1.50	−1.40	−1.50	−1.45
	(p=0.040)	(p=0.006)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p<0.001)
Per complier	−17.64	−12.10	−10.43	−8.97	−6.78	−6.30	−12.60
	(p=0.068)	(p=0.010)	(p=0.002)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.014)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							18.3
Exceptionally large, ≥4,500 g (P08.0) (mean below cutoff: 0.2%)							
Reduced form	–	0.28	0.17	0.08	0.02	0.00	0.34
		(p=0.184)	(p=0.372)	(p=0.632)	(p=0.922)	(p=0.995)	(p=0.121)
Per complier	–	2.25	1.23	0.48	0.07	0.00	3.26
		(p=0.199)	(p=0.377)	(p=0.633)	(p=0.922)	(p=0.995)	(p=0.109)
N	–	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							12.0
NICU admission (revenue-code indicator) (mean below cutoff: 2.4%)							
Reduced form	−1.84	−1.43	−0.99	−1.19	−1.48	−1.45	−1.33
	(p=0.035)	(p=0.012)	(p=0.033)	(p=0.003)	(p<0.001)	(p<0.001)	(p=0.011)
Per complier	−17.85	−11.53	−7.08	−7.09	−7.18	−6.07	−11.88
	(p=0.071)	(p=0.020)	(p=0.039)	(p=0.004)	(p<0.001)	(p<0.001)	(p=0.015)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							14.6
Normal-newborn nursery admission (revenue-code indicator) (mean below cutoff: 14.1%)							
Reduced form	−1.42	−2.42	−1.87	−1.81	−2.63	−3.31	−2.11
	(p=0.482)	(p=0.079)	(p=0.102)	(p=0.074)	(p=0.006)	(p<0.001)	(p=0.158)
Per complier	−13.75	−19.40	−13.36	−10.77	−12.73	−13.90	−18.66
	(p=0.479)	(p=0.081)	(p=0.101)	(p=0.073)	(p=0.006)	(p<0.001)	(p=0.082)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							10.5

Notes: Estimates use fasting glucose testing episodes with a linked infant record. I measure infant outcomes from one day before through 30 days after delivery. Macrosomia combines ICD-10-CM codes P08.0 and P08.1 on infant claims. NICU (neonatal intensive care unit) and nursery admissions use facility revenue codes. The estimated macrosomia decline is concentrated in P08.1. P08.0 is too uncommon for precise estimation, and cells with too few events contain a dash. These measures do not directly capture birth weight.

Reduced form rows estimate the effect of crossing 95 mg/dL. Per complier rows estimate the effect of GDM diagnosis among women whose diagnosis is induced by crossing the cutoff. All effect estimates are in percentage points, with p values in parentheses. Columns headed $\pm h$ use local linear regressions with a uniform kernel. The last column uses a triangular kernel and an MSE optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014). I report conventional estimates without bias correction. Each per complier estimate divides the outcome discontinuity by the diagnosis first stage using the same sample, bandwidth,

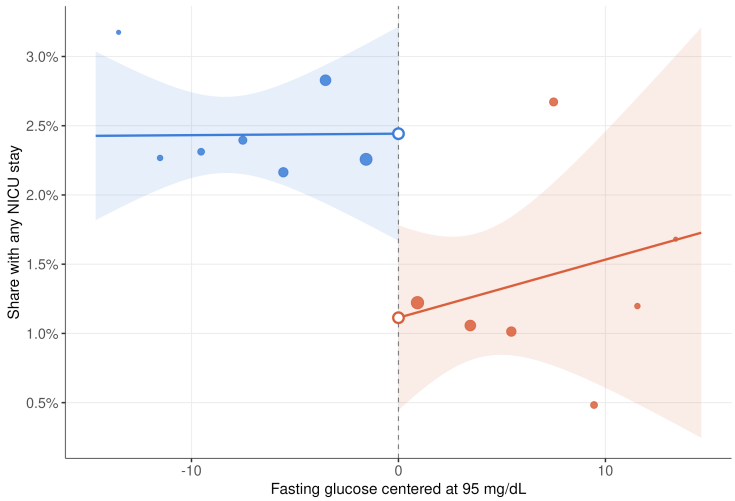
and weights. In the last column, per complier estimates use the bandwidth selected for the fuzzy design. I omit all per complier estimates for an outcome if its first stage F statistic is below 10 at ± 10 mg/dL or at the fixed bandwidth nearest the optimal bandwidth. I also omit an individual per complier estimate wherever its own first stage has $F < 10$. Standard errors are clustered by woman. In the fixed bandwidth columns, N counts testing episodes within the bandwidth. The last column reports the full eligible sample as N. The reported h is the optimal bandwidth in mg/dL. Author's calculations from MarketScan claims data.

Panel (a) of Figure N.1 in Appendix N plots the combined macrosomia estimate across fixed bandwidths. Table 11 reports the corresponding numerical estimates. I estimate a decline in recorded macrosomia at every fixed bandwidth from 5 to 30 mg/dL. Only the estimate at 5 mg/dL is not statistically significant.

Dropping episodes at exactly 95 mg/dL makes the estimated decline larger: at the fixed 10 mg/dL bandwidth, it increases from 1.29 to 1.72 percentage points and remains statistically significant (Table N.2). Excluding wider intervals around the cutoff still yields negative estimates, but with less precision. With a bandwidth of 20 mg/dL, the estimated reductions range from 1.27 to 1.66 percentage points across the exclusion radii examined. Table N.3 reports the discontinuities at alternative cutoffs. The combined indicator and P08.1 both have a positive and statistically significant estimate at 120 mg/dL. These two estimates coincide at that cutoff and have the opposite sign from the discontinuity at 95 mg/dL.

Figure 12 plots the probability of admission to a neonatal intensive care unit (NICU), identified from facility revenue codes, against maternal fasting glucose. The fitted admission probability falls at the threshold, showing a downward discontinuity similar to that observed for macrosomia.

Figure 12: NICU admission



Notes: Points are bin means against the fasting glucose value, centered at 95 mg/dL, infant-linked subset. The line is a local-polynomial fit with shaded pointwise 95% confidence bands, estimated with the optimal-bandwidth estimator (triangular kernel, MSE-optimal bandwidth ±14.6 mg/dL). RD = −0.013 ($p = 0.011$), matching the outcome's optimal-bandwidth row in Table 11. The eligible sample contains $N = 37{,}464$ testing episodes. Author's calculations from MarketScan claims data.

The NICU row of Table 11 reports a reduction of 1.33 percentage points at the optimal bandwidth, a statistically significant reduction. The corresponding IV estimate implies a reduction of 11.9 percentage points. Panel (b) of Figure N.1 plots the reduced form across fixed bandwidths and shows that the estimates remain negative throughout the range examined.

Dropping the 372 testing episodes recorded at exactly 95 mg/dL changes the estimated reduction in NICU admission from 1.43 to 1.19 percentage points at the fixed 10 mg/dL bandwidth (Table N.2). The NICU estimate is no longer statistically significant after this exclusion. At a bandwidth of 20 mg/dL, the estimated NICU decline is 0.86 percentage points after excluding observations within 3 mg/dL of the cutoff and 0.45 points after excluding those within 5 mg/dL. I find no statistically significant NICU discontinuity at the placebo cutoffs (Table N.3). I also examine the number of days with a NICU claim in Table N.1, but about one NICU claim day is recorded per infant with a NICU admission. Therefore, it nearly duplicates the admission indicator.

For the other diagnoses recorded around delivery, I estimate a 1.32 percentage point reduction in preterm birth or extreme immaturity at the optimal bandwidth (Table N.1). This reduction is statistically significant. The estimated reduction in jaundice is 2.10 percentage points but is too imprecise to distinguish from zero. The estimates for low birth weight, neonatal hypoglycemia or other glucose disorders, respiratory distress, and birth injury are not statistically significant and do not rule out effects on these outcomes.

Selection and claims observation. A linked infant enrollment record does not guarantee that I observe any claims for the infant. A diagnosis or admission appears in the claims only if the infant has a claim. I measure infant outcomes using claims from one day before to 30 days after delivery. Within 10 mg/dL of the cutoff, 3,184 of 10,217 testing episodes with a linked infant have any infant claim in this window. Infants without a claim cannot have a recorded event under these definitions, so the estimates concern recorded diagnoses and services, not fully observed clinical outcomes. Appendix J.1 examines linkage and claim observation separately.

I first repeat the covariate continuity tests within the linked sample at a fixed bandwidth of 10 mg/dL (Table J.3). None of the 23 tests detects a statistically significant jump. When I compare the discontinuities between linked and unlinked observations, only pregnancy spending before the test has a statistically significant difference. I include pregnancy spending before the test among the covariates in the reweighting described next.

I re-estimate the infant outcomes after weighting linked observations by the inverse of their estimated probability of linkage (Table J.4). At the optimal bandwidth, the reweighted reduced forms imply reductions of 1.38 percentage points for the combined macrosomia indicator, 1.53 points for P08.1, and 1.09 points for NICU admission. These magnitudes are close to the unweighted estimates, but none is statistically significant. All three reweighted estimates are significant at fixed bandwidths of 10, 15, and 20 mg/dL, but not at 5 mg/dL. The covariate checks and reweighting do not rule out selection on unobserved characteristics. The standard errors for the reweighted estimates treat the estimated weights as known.

Panel A of Table J.5 uses the trimming approach of Lee (2009).[17] I assume that crossing the cutoff never prevents an infant from being linked. Under this assumption, the higher linkage rate above the cutoff means that some infants in that sample would not have been linked below it. At the fixed 10 mg/dL bandwidth, the estimated excess share among linked infants above the cutoff is about 7.0 percent. I remove this share from the highest outcome values above the cutoff and, in a separate calculation, from the lowest. The resulting approximate bounds imply reductions of 1.23 to 2.31 percentage points for macrosomia and 1.36 to 2.59 points for NICU admission. Both reductions remain statistically distinguishable from zero after accounting for sampling uncertainty.

Panel B repeats the calculation without observations at exactly 95 mg/dL. The macrosomia reduction remains statistically significant, but the NICU reduction does not. For inference, I resample women and recompute the trimming share and bounds in each draw. The bounds are approximate because I apply the trimming share estimated at the cutoff

[17] The linked share is the probability of a linked infant among observed deliveries. Crossing the cutoff does not change the probability of observing delivery (Table J.1). See Section 5.5 for more detail.

throughout the estimation window. They concern the reduced-form effect of crossing the threshold among infants who would be linked on either side. They do not bound the effect of diagnosis among compliers or directly validate the IV magnitudes.

Table J.6 compares all linked infants with those who have at least one claim during the infant outcome window. At the fixed 10 mg/dL bandwidth, the estimated change in the probability of observing any infant claim is a decline of 0.6 percentage points, which is not statistically significant. Restricting the sample to infants with a claim gives larger estimated reductions of 4.48 percentage points for macrosomia and 4.86 points for NICU admission. The baseline frequencies of both outcomes are also higher in this sample. These larger estimates apply only to infants with an observed claim; the additional restriction does not resolve the selection concern.

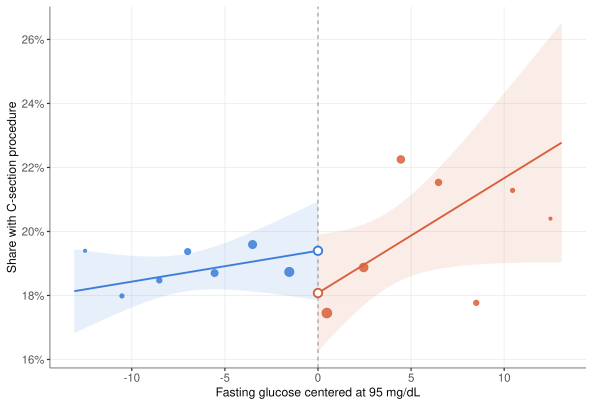
Panel C of Table J.5 applies trimming to claim observation. This calculation requires a separate assumption: crossing the cutoff cannot increase the probability that a linked infant has a claim. After accounting for sampling uncertainty, these bounds do not distinguish the effects from zero.

5.5 Maternal health and delivery

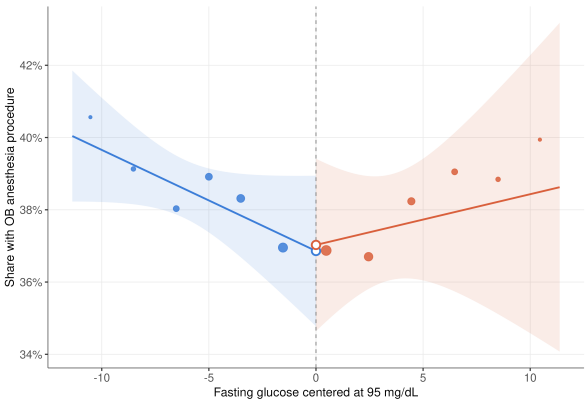
I next examine whether delivery procedures, admission payments, or maternal complications change at the cutoff. For the delivery analyses, I require an observed delivery and the enrollment and claims information needed to measure each outcome. Table F.2 lists these requirements and the resulting estimation samples. A linked infant record is not required.

Figure 13 plots delivery procedures and admission payments against maternal fasting glucose. In panel (a), the fitted probability of a recorded cesarean delivery is slightly lower just above 95 mg/dL. The probability of obstetric anesthesia changes little at the cutoff in panel (b). I see no clear upward jump in either procedure.

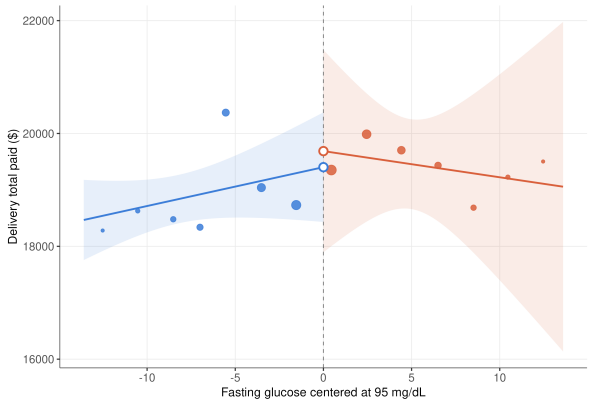
Figure 13: Delivery management and cost



(a) C-section procedure



(b) Obstetric anesthesia procedure



(c) Delivery total paid

Notes: Points are bin means against the fasting glucose value, centered at 95 mg/dL. Lines are local-polynomial fits with shaded pointwise 95% confidence bands, estimated with the optimal-bandwidth estimator (triangular kernel, each panel's own MSE-optimal bandwidth: (a) ±13.1; (b) ±11.4; (c) ±13.6 mg/dL), matching each outcome's optimal-bandwidth row in Table 12. The eligible samples contain N = 62,795 testing episodes for panels (a)-(b) and N = 30,906 for panel (c). None of the three panels shows a discontinuity in delivery procedures or payment at the threshold. Appendix M reports bandwidth stability for the C-section and total-paid rows. Author's calculations from MarketScan claims data.

Table 12 reports the reduced form estimates. Among the 62,795 eligible fasting glucose testing episodes whose delivery episode lies entirely within enrollment, the optimal-bandwidth estimates indicate a decrease of 1.32 percentage points in cesarean delivery and an increase of 0.17 percentage points in obstetric anesthesia; neither is statistically significant. Table 12 also reports the corresponding IV estimates: a decrease of 11.7 percentage points in cesarean delivery and an increase of 1.2 percentage points in anesthesia. Both are imprecise.

Table 12: Delivery procedures and payments

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Cesarean delivery (mean below cutoff: 18.8%)							
Reduced form	−2.22	−0.89	−0.05	−0.29	−0.39	−0.60	−1.32
	(p=0.210)	(p=0.473)	(p=0.962)	(p=0.751)	(p=0.653)	(p=0.475)	(p=0.279)
Per complier	−21.75	−7.38	−0.38	−1.92	−2.20	−3.01	−11.68
	(p=0.239)	(p=0.479)	(p=0.962)	(p=0.752)	(p=0.654)	(p=0.477)	(p=0.291)
N	8,463	18,939	32,847	47,739	57,491	61,028	62,795
Optimal h							13.1
Obstetric anesthesia (mean below cutoff: 38.8%)							
Reduced form	0.91	0.26	0.50	1.55	1.26	0.50	0.17
	(p=0.679)	(p=0.866)	(p=0.699)	(p=0.177)	(p=0.240)	(p=0.630)	(p=0.918)
Per complier	8.92	2.15	3.82	10.11	7.12	2.50	1.17
	(p=0.675)	(p=0.866)	(p=0.697)	(p=0.172)	(p=0.235)	(p=0.628)	(p=0.936)
N	8,463	18,939	32,847	47,739	57,491	61,028	62,795
Optimal h							11.4
Delivery admission, total payment (mean below cutoff: $19,054)							
Reduced form	$1,146	$553.92	$13.93	−$4.28	$23.26	$82.56	$284.66
	(p=0.442)	(p=0.590)	(p=0.988)	(p=0.996)	(p=0.973)	(p=0.900)	(p=0.784)
Per complier	$8,290	$3,445	$75.07	−$19.95	$89.11	$279.15	$3,813
	(p=0.449)	(p=0.590)	(p=0.988)	(p=0.996)	(p=0.973)	(p=0.900)	(p=0.617)
N	3,592	8,330	15,109	22,871	28,131	30,033	30,906
Optimal h							13.6

Notes: Estimates use fasting glucose testing episodes with an observed delivery. Cesarean delivery and obstetric anesthesia also require enrollment throughout the delivery episode. Payment estimates require an admission record with payment information. Within ±10 mg/dL, women with payment information have a higher recorded GDM diagnosis rate than those without it. None of the procedure or payment estimates is statistically significant at any reported bandwidth. The payment estimates are imprecise, and some procedures may be missing from the claims.

Reduced form rows report the effect of crossing 95 mg/dL. Per complier rows report the effect of GDM diagnosis among women diagnosed because they cross the cutoff. Procedure effects are in percentage points; payment effects are in nominal dollars. Each per complier estimate divides the outcome discontinuity by the GDM diagnosis first stage using the same sample, bandwidth, and weights. Columns headed $\pm h$ use local linear regressions with a uniform kernel. The last column uses a triangular kernel and an MSE optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014). I report conventional estimates without bias correction. Per complier estimates in the last column use the bandwidth selected for the fuzzy design.

I omit all per complier estimates for an outcome if its first stage has $F < 10$ at ±10 mg/dL or at the fixed bandwidth nearest the optimal bandwidth. I also omit an individual per complier estimate if the first stage at its own bandwidth has $F < 10$. Standard errors are clustered by woman, and p values appear in parentheses. In the fixed bandwidth columns, N counts testing episodes within the bandwidth. The last

column reports the full eligible sample as N. The reported h is the optimal bandwidth in mg/dL. Author's calculations from MarketScan claims data.

Procedure recording differs substantially between deliveries with and without an observed admission payment (Appendix J). Across all observed deliveries, cesarean codes appear in 36.7 percent of those with a payment record and 1.1 percent of those without one. The corresponding shares for obstetric anesthesia are 82.2 and 3.5 percent. These differences suggest that a missing procedure code may reflect incomplete records. Because I code a delivery with no procedure claim as having no procedure, incomplete capture can affect the estimated discontinuity. Its direction depends on whether procedure recording changes at the cutoff.

To address this, I repeat the procedure analysis using only episodes with an admission record and enrollment covering the entire delivery episode. This leaves 8,254 observations within the fixed 10 mg/dL bandwidth. The estimated reductions are 2.5 percentage points for cesarean delivery and 0.8 points for anesthesia. Neither is statistically significant. Even in this sample with more complete procedure information, the estimates are too imprecise to conclude that delivery care is unchanged.

Panel (c) of Figure 13 plots average payments on observed delivery admissions against fasting glucose. The fitted payment levels differ little at 95 mg/dL. Table 12 reports an estimated increase of $285 at the optimal bandwidth. The eligible sample has 30,906 testing episodes with an observed admission payment. The optimal bandwidth IV estimate in Table 12 is an increase of $3,813. Neither estimate is statistically significant.

Figure M.2 in Appendix M shows the reduced form estimates across fixed bandwidths. Panel (a) reports cesarean delivery and panel (b) reports admission payments. Neither outcome has a statistically significant discontinuity at any bandwidth shown.

Not all delivery admissions have payment records. Table J.2 examines whether this restriction is "selective." In other words, I check whether the availability of a record changes at the cutoff, and how admissions with a record differ from those without. Panel A shows no statistically significant jump in the probability of observing such a record at the optimal bandwidth. When I additionally require enrollment throughout the delivery episode, the optimal bandwidth payment estimate changes from a $285 increase to a $271 decrease (Panel D). Neither estimate is statistically significant. These results apply to a selected set of admissions and are too imprecise to rule out economically meaningful increases or decreases in payments.

I also examine claims attached to each delivery episode. Among 19,065 testing episodes with an observed delivery, I estimate an increase of 1.3 days, which is not statistically significant. At bandwidths of 20 to 30 mg/dL, the estimated increases are larger, about 2.9

to 3.9 days, and statistically significant. The number of distinct days with a claim also rises at wider bandwidths. A delivery claims episode ends with the last attached claim, however, so these findings do not establish that women spend more time in the hospital.

Recorded gestational hypertension generally becomes more common as fasting glucose rises, but the jump at 95 mg/dL is imprecisely estimated (Figure M.1 in Appendix M). Table M.1 reports estimates for gestational hypertension, preeclampsia, severe preeclampsia or HELLP syndrome, and polyhydramnios. I count diagnoses recorded from the glucose test through the end of the pregnancy window, using all 85,611 eligible testing episodes. None of these estimates is statistically significant at the optimal bandwidth. The largest is a 0.67 percentage point increase in gestational hypertension. The estimates for gestational hypertension and preeclampsia become significant at some of the widest fixed bandwidths.

There are too few cases of the rare maternal complications for precise estimates. I also find no detectable changes in readmission, labor dystocia, operative vaginal delivery, or postpartum hypertension in the reported specifications. Meaningful effects on maternal complications remain possible given the uncertainty in these estimates. The maternal and delivery results also do not show which services account for the infant outcome findings.

5.6 Postpartum care and longer term outcomes

I follow women after delivery to examine subsequent care, spending, and recorded health conditions. For these analyses, I require an observed delivery and enrollment through the relevant follow up window. Medication outcomes have additional pharmacy requirements, and follow up for offspring outcomes depends on the child's enrollment record. The samples become smaller as follow up lengthens because enrollment ends or there is not enough time between delivery and September 30, 2025, when the data end.

Panel A of Table 13 reports the results for medical encounters in days 43 to 84 after delivery. There are 54,291 testing episodes eligible for this analysis. At the optimal bandwidth, I estimate that crossing the cutoff raises the probability of an encounter by 0.52 percentage points, with an IV estimate of 1.11 points. Neither estimate is statistically significant. I also find no significant change in glucose testing, recorded diabetes or prediabetes, or most measures of medication and monitoring supply use. I examine testing separately from diagnoses because a diagnosis code does not show whether recommended screening took place.

Table 13: Postpartum care and long-run maternal disease: incident type 2 diabetes and prediabetes

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Panel A. Postpartum care							
Any encounter, days 43–84 after delivery (mean below cutoff: 83.1%)							
Reduced form	0.32	0.47	1.46	1.05	1.00	1.33	0.52
	(p=0.858)	(p=0.705)	(p=0.160)	(p=0.255)	(p=0.246)	(p=0.109)	(p=0.663)
Per complier	2.87	4.13	11.66	7.06	5.77	6.82	1.11
	(p=0.858)	(p=0.706)	(p=0.168)	(p=0.259)	(p=0.249)	(p=0.112)	(p=0.924)
N	7,319	16,378	28,445	41,287	49,718	52,771	54,291
Optimal h							13.7
Panel B. Incident type 2 diabetes							
Incident type 2 diabetes (E11), 12 months (mean below cutoff: 1.8%)							
Reduced form	−0.65	0.65	0.95	0.78	0.99	0.94	0.51
	(p=0.557)	(p=0.393)	(p=0.166)	(p=0.209)	(p=0.094)	(p=0.103)	(p=0.490)
Per complier	–	6.47	8.66	6.04	6.70	5.46	4.71
		(p=0.395)	(p=0.171)	(p=0.211)	(p=0.095)	(p=0.103)	(p=0.548)
N	4,180	9,347	16,201	23,437	28,210	29,944	30,771
Optimal h							13.7
Incident type 2 diabetes (E11), 24 months (mean below cutoff: 2.9%)							
Reduced form	−1.51	0.64	0.29	0.36	0.62	0.91	0.42
	(p=0.385)	(p=0.614)	(p=0.804)	(p=0.726)	(p=0.527)	(p=0.334)	(p=0.722)
Per complier	–	–	–	–	–	–	–
N	2,368	5,335	9,228	13,327	16,051	17,030	17,463
Optimal h							14.5
Incident type 2 diabetes (E11), 36 months (mean below cutoff: 4.5%)							
Reduced form	−1.50	0.22	1.23	1.91	0.85	0.56	−0.66
	(p=0.604)	(p=0.913)	(p=0.473)	(p=0.200)	(p=0.567)	(p=0.697)	(p=0.769)
Per complier	–	–	–	–	–	–	–
N	1,441	3,278	5,741	8,272	9,980	10,585	10,843
Optimal h							10.3
Panel C. Incident prediabetes							
Incident prediabetes (R73.0), 12 months (mean below cutoff: 4.7%)							
Reduced form	1.60	1.34	1.52	1.93	2.18	2.50	1.31
	(p=0.295)	(p=0.222)	(p=0.106)	(p=0.023)	(p=0.007)	(p=0.001)	(p=0.223)
Per complier	–	13.37	13.83	14.90	14.70	14.47	13.76
		(p=0.236)	(p=0.117)	(p=0.028)	(p=0.009)	(p=0.002)	(p=0.157)
N	4,180	9,347	16,201	23,437	28,210	29,944	30,771
Optimal h							13.6

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Incident prediabetes (R73.0), 24 months (mean below cutoff: 9.0%)							
Reduced form	2.45	2.04	1.52	2.67	3.40	3.84	1.79
	(p=0.370)	(p=0.297)	(p=0.358)	(p=0.070)	(p=0.013)	(p=0.003)	(p=0.335)
Per complier	–	–	–	–	–	–	–
N	2,368	5,335	9,228	13,327	16,051	17,030	17,463
Optimal h							14.2
Incident prediabetes (R73.0), 36 months (mean below cutoff: 12.5%)							
Reduced form	−1.18	0.58	2.43	3.10	3.83	4.19	−0.09
	(p=0.780)	(p=0.842)	(p=0.316)	(p=0.158)	(p=0.059)	(p=0.029)	(p=0.978)
Per complier	–	–	–	–	–	–	–
N	1,441	3,278	5,741	8,272	9,980	10,585	10,843
Optimal h							11.1

Notes: All panels use fasting glucose testing episodes with an observed delivery and continuous enrollment through the relevant follow up window. Panel A measures any medical encounter in days 43 to 84 after delivery. Panels B and C measure first recorded diagnoses of type 2 diabetes and prediabetes from day 43 after delivery through the stated horizon. For these outcomes, I exclude women with a qualifying diagnosis code before the glucose test. Fewer episodes qualify at longer horizons because enrollment ends or the required follow up extends beyond the available data. Too few episodes qualify at 60 months for an informative estimate, so I omit that horizon.

At 36 months, the descriptive shares with recorded type 2 diabetes are 3.24 percent below the cutoff and 11.96 percent above it. The corresponding shares for prediabetes are 9.34 and 17.99 percent.

Reduced form rows estimate the effect of crossing 95 mg/dL. Per complier rows estimate the effect of GDM diagnosis among women whose diagnosis is induced by crossing the cutoff. All effects are in percentage points, with p values in parentheses. Columns headed $\pm h$ use local linear regressions with a uniform kernel. The last column uses a triangular kernel and an MSE optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014). I report conventional estimates without bias correction. Each per complier estimate divides the outcome discontinuity by the GDM diagnosis first stage using the same observations, bandwidth, and weights. In the last column, this calculation uses the bandwidth selected for the fuzzy design.

I omit all per complier estimates for an outcome if its first stage has $F < 10$ at ±10 mg/dL or at the fixed bandwidth nearest the optimal bandwidth. I also omit an individual per complier estimate if its first stage at that bandwidth has $F < 10$. Standard errors are clustered by woman. In the fixed bandwidth columns, N counts testing episodes within the bandwidth. In the last column, N is the full eligible sample. The reported h is the optimal bandwidth in mg/dL. Author's calculations from MarketScan claims data.

DSMT claims and insulin fills do increase over the 180 days beginning on the delivery date. At the optimal bandwidth, crossing the cutoff raises the probability of a DSMT claim by 3.34 percentage points and an insulin fill by 2.53 points. Both increases are statistically significant.

The largest descriptive gap in DSMT claims appears in the first 30 days after delivery:

9.0 percent of women above the cutoff have a claim, compared with 4.5 percent below it. From day 90 onward, claim frequencies fall below 1 percent in each 30 day interval. These comparisons suggest that the difference in recorded education use is concentrated soon after delivery, but they are descriptive rather than RD estimates. The available timing results do not establish whether insulin fills follow the same pattern.

Table O.1 separates postpartum spending into total payments, payments on medical claims other than DSMT and MNT, and payments on DSMT and MNT claims. Panel A covers spending through the first year after delivery, and Panel B covers the first two years cumulatively. Both spending windows begin on the day after the delivery claims episode ends, unlike the 180 day care window above, which begins on the delivery date.

For the first year, I estimate a $935 reduction in total payments at the optimal bandwidth, using 30,758 eligible testing episodes. The IV estimate is a reduction of $10,632. Neither is statistically significant, and the estimates cannot distinguish substantial savings from an increase in spending.

The estimated reductions in other medical payments and DSMT/MNT payments are $909 and $0.66, respectively, at the optimal bandwidth. Neither is statistically significant. Although the DSMT/MNT point estimate is close to zero, it is too imprecise to conclude that these payments are unchanged. Its uncertainty is substantial relative to the increase in DSMT/MNT payments during pregnancy.

The spending results depend on bandwidth. At the fixed 20 mg/dL bandwidth, the estimated reductions in total and other medical payments are larger and statistically significant. At bandwidths of 20 mg/dL and above, the estimated decline in total payments ranges from about $1,370 to $1,540. None of the reported IHS transformed payment estimates is statistically significant.

Through two years, none of the three dollar measures in Panel B shows a statistically significant change at the optimal bandwidth. The estimated reduction in total payments is $755, based on an eligible sample of 17,460 testing episodes. The diagnosis first stage is weak under the criterion in Section 4, so I do not interpret an IV estimate for this outcome. The reduced form estimate still leaves room for economically meaningful increases or decreases in cumulative spending.

Table 13 also reports results for recorded type 2 diabetes and prediabetes at 12, 24, and 36 months after delivery. I count diagnoses recorded after postpartum day 42 among women without a qualifying diagnosis before the glucose test. Each horizon requires enrollment through that point. At the fixed 10 mg/dL bandwidth, the sample falls from 9,347 observations at 12 months to 5,335 at 24 months and 3,278 at 36 months.

The optimal bandwidth estimates for recorded type 2 diabetes are increases of 0.51 per-

centage points at 12 months and 0.42 points at 24 months, and a decrease of 0.66 points at 36 months. None is statistically significant, nor are the prediabetes estimates at the optimal bandwidth. Given their imprecision, I cannot rule out meaningful increases or decreases in later recorded disease.

The diagnosis first stage in the samples at 24, 36, and 60 months falls below the strength criterion in Section 4. I therefore report reduced form results without interpreting effects per additional diagnosis. Appendix J examines whether continued enrollment changes at the cutoff. Most checks find no statistically significant change, although the result at 60 months varies with the specification. Differential attrition remains a possibility.

I also examine hypertension, cardiovascular disease, dyslipidemia, metabolic syndrome, depression, and chronic kidney disease from 12 to 60 months after delivery (Table O.2). None of the 24 estimates is statistically significant at the optimal bandwidth. Dyslipidemia at 12 months is significant at the fixed 10 mg/dL bandwidth, but not at the optimal bandwidth or at later horizons.

For recurrence and fertility, I examine outcomes at 24, 36, and 48 months after the index glucose test (Appendix G). Among episodes with complete follow up, I estimate that crossing the cutoff raises the probability of a subsequent delivery by 4.3 percentage points within 24 months, 8.4 points within 36 months, and 9.1 points within 48 months. The increases at 36 and 48 months are statistically significant.

These estimates are sensitive to observations near the cutoff. Dropping episodes at exactly 95 mg/dL reduces the estimate at 36 months to 7.5 percentage points, and it is no longer statistically significant. Excluding observations within 3 mg/dL of the cutoff reverses the signs of the estimates and leaves them imprecise. I therefore cannot conclude that crossing the cutoff increases subsequent fertility.

I find no statistically significant change in recorded recurrence at any of these horizons. Restricting the sample to women with an observed subsequent delivery gives an estimated increase of 5.9 percentage points. But subsequent delivery may itself respond to crossing the cutoff. This estimate therefore cannot separate a change in recurrence from a change in which women have another observed delivery.

The infant spending samples are much smaller, with 1,753 testing episodes through one year and 1,093 through two years. Neither total nor inpatient payments in dollars show a statistically significant change in the reported specifications. Some estimates for patient payments are negative, but the reductions are not consistent across bandwidths.

IHS transformed inpatient payments decline when I measure spending from birth. After excluding the first 30 days, most estimates are no longer statistically significant. This pattern suggests that any spending change may be concentrated in the neonatal period, but it does

not show that fewer NICU admissions account for the payment results. I treat these analyses as exploratory because the samples are small and require both infant linkage and continued observation.

At the fixed 10 mg/dL bandwidth, the reduced form estimates for recorded asthma and allergic rhinitis by 36 months are increases of 12.86 and 13.93 percentage points, respectively. This sample contains 698 testing episodes with sufficient infant follow up. Both estimates are statistically significant before adjustment for multiple testing, as are the corresponding optimal bandwidth estimates.

Table O.3 reports the estimates across bandwidths. Allergic rhinitis is significant in five of the seven specifications, while asthma is significant at fixed bandwidths of 5 and 10 mg/dL and at the optimal bandwidth. Neither outcome remains statistically significant after the reported Holm or Benjamini and Hochberg adjustment for the ten outcomes examined at 36 months.

I find no detectable change in recorded atopic dermatitis or urticaria. Too few children have recorded type 2 diabetes, hypertension, or obesity through 96 months to estimate effects precisely. The small, selected samples and reliance on diagnosis codes prevent firm conclusions about effects on children's health beyond the neonatal period.

5.7 Validity checks and additional analyses

Distribution and covariate checks. I use the test developed by Frandsen (2017) for discrete running variables to check for manipulation near 95 mg/dL. The test compares the number of testing episodes at 95 mg/dL with those at 94 and 96 mg/dL. Figure 3 plots the counts near the cutoff and marks the expected count at 95 mg/dL under a locally linear probability mass function. Of the 3,092 testing episodes at these three values, 1,022, or 33.1 percent, are recorded at 95 mg/dL. This is close to the 33.3 percent expected under local linearity.

Table I.3 reports the Frandsen test results for different values of the curvature parameter K. The test does not reject local smoothness under local linearity ($K = 0$, $p = 0.746$) or when allowing some curvature ($K = 0.01$, $p = 0.755$). It also does not reject at any of the 18 comparison values. The smallest of the 18 p-values at $K = 0$ is 0.10 (Appendix Table I.4).

I also test for discontinuities in observed characteristics. None of the 23 specifications for baseline covariates and pregnancy spending before the test detects a discontinuity at the 5 percent level. The estimated discontinuity in spending before the test is $-\$275$ ($p = 0.339$). Table I.6 and Figure I.4 report these checks. These tests cannot establish continuity of unobserved potential outcomes.

The share of episodes with a diabetes diagnosis before the test rises from about 2 percent at fasting glucose values of 75 to 87 mg/dL to 4 to 5 percent near the cutoff. At the optimal bandwidth, the estimated discontinuity in this share is −0.36 percentage points ($p = 0.510$). None of the other indicators of prior diabetes has a statistically significant jump (Appendix Table C.1).

Excluding episodes with any diabetes diagnosis code before the test gives a first stage of 10.72 percentage points, close to the 10.74 points in the full sample. Excluding episodes that meet the broader definition of prior diabetes gives a first stage of 11.20 percentage points. Appendix Table C.2 reports the estimates at the optimal bandwidth for each sample.

After excluding episodes with a prior diabetes diagnosis code, the reduced form for NICU admission is −1.37 percentage points ($p = 0.012$), close to −1.33 in the full sample. The estimate for heavy for gestational age changes from −1.45 to −1.57 percentage points ($p < 0.001$). The estimated increase in DSMT/MNT payments changes from \$12.14 to \$11.91 ($p < 0.001$).

The insulin initiation estimate falls from 1.54 percentage points ($p = 0.033$) to 1.18 points ($p = 0.124$) after excluding episodes with a prior diabetes diagnosis code. It is 1.31 points ($p = 0.094$) under the broader exclusion. Neither estimate remains statistically significant. For preterm birth or extreme immaturity, the estimate changes from −1.32 percentage points ($p = 0.038$) to −1.19 points ($p = 0.065$) after the diagnosis code exclusion and to −1.29 points ($p = 0.040$) under the broader exclusion.

Entry into the outcome samples. I find no statistically significant discontinuity in entry into the pregnancy spending, delivery, or delivery payment samples at either the fixed ±10 mg/dL bandwidth or the optimal bandwidth. At wider bandwidths, however, entry into the delivery sample falls at ±25 and ±30 mg/dL, while entry into the delivery payment sample rises at ±20 and ±25 mg/dL. These changes are statistically significant. Most enrollment checks also show no statistically significant discontinuity, with one borderline result at 60 months that does not appear at the optimal bandwidth. Infant linkage does increase at the cutoff. The probability of linking an infant record rises by 3.6 percentage points at the fixed ±10 mg/dL bandwidth ($p = 0.022$) and by 3.5 points at the optimal bandwidth ($p = 0.033$). This raises a selection concern for the infant outcomes (Appendix J). Section 5.4 and Appendix J.1 examine it.

Sensitivity of the first stage. I also examine whether the diagnosis jump depends on observations recorded at or near 95 mg/dL. Excluding fasting glucose testing episodes at exactly 95 mg/dL reduces the first stage from 10.74 to 6.90 percentage points at the optimal

bandwidth. The estimate remains positive and statistically significant. When I exclude progressively wider intervals around the cutoff, the estimates remain significant with fixed bandwidths of 10 and 20 mg/dL. Under the optimal bandwidth specification, however, the estimates become less precise and are no longer statistically significant at the largest exclusion radii.

The estimated diagnosis jump also increases as the fixed bandwidth widens. Wider bandwidths include observations farther from 95 mg/dL, and larger exclusions require extrapolating the fitted lines across a wider gap around the cutoff. The diagnosis jump remains positive after dropping observations at exactly 95 mg/dL. Its estimated size depends on the bandwidth and the range of glucose values excluded.

Finally, I estimate diagnosis discontinuities at placebo cutoffs away from 95 mg/dL. None of the 12 placebo estimates is statistically significant at the 5 percent level. Appendix I reports the complete sensitivity and placebo results.

6 Conclusion

I use the 95 mg/dL fasting glucose cutoff to estimate the effects of GDM diagnosis among commercially insured women in the United States. Diagnosis increases diabetes education and medical nutrition therapy. Medication use and biophysical profile testing also increase. For education and nutrition services, I estimate an increase of $107 in total payments and about $28 in patient out of pocket payments per additional diagnosis.

I find no detectable increase in total pregnancy spending at the optimal bandwidth. This estimate is too imprecise to conclude that the additional care has little or no cost: it allows substantial spending increases as well as savings.

For women diagnosed because they cross the threshold, I estimate lower probabilities of recorded infant macrosomia and NICU admission. These findings come from the linked infant sample. Because linkage increases at the cutoff, selection into that sample remains a concern. I find no detectable increase in cesarean delivery, obstetric anesthesia, or delivery payments, but incomplete procedure records and imprecise estimates leave open a role for changes in delivery care. The analysis cannot attribute the infant benefits to education, nutrition therapy, or any other individual service.

The infant findings suggest benefits from diagnosing women near this fasting glucose threshold. Evaluating a different threshold requires evidence on the women whose diagnosis would change under that rule and the care they would receive. My estimates concern the 95 mg/dL fasting cutoff within the Carpenter and Coustan protocol. They do not identify the optimal cutoff or establish that another screening strategy would produce similar benefits.

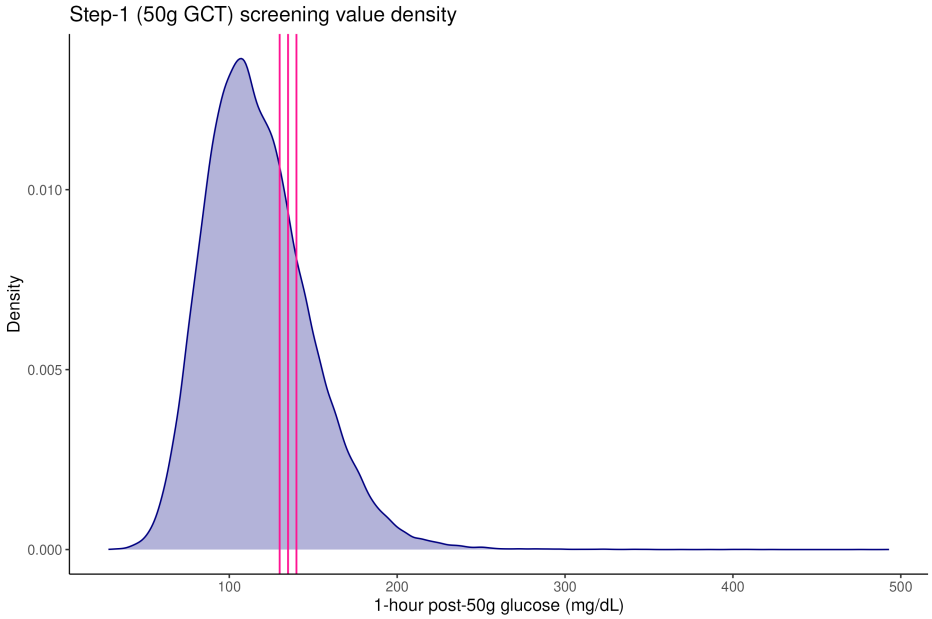
Appendix A: The Screening Step (Step 1)

The analysis in the main text uses pregnancy-linked fasting glucose measurements, which include the fasting value of the diagnostic glucose tolerance test (Step 2) but cannot be identified uniquely as such (Section 3). Most women are first screened with a nonfasting 50-gram glucose challenge test (Step 1), and a value at or above a screening cutoff leads to the diagnostic test (Section 2). Guidelines accept cutoffs of 130, 135, or 140 mg/dL, and the choice is left to the clinician or institution (American College of Obstetricians and Gynecologists, 2018; American Diabetes Association, 2019*b*). This appendix describes the distribution of screening values in the claims data and how subsequent testing and diagnosis respond at each cutoff. The paper does not use the screening margin as a separate design.

Figure A.1 shows the distribution of screening values with the three accepted cutoffs marked. Because the values are whole numbers, I test for sorting around each cutoff with the test for a discrete running variable (Section 4.3). It does not reject at any of them ($p = 0.779$, 0.511, and 0.822 at 130, 135, and 140 mg/dL with $K = 0$; Table I.3), and neither does the continuous-density test of Cattaneo, Jansson and Ma (2020) ($p = 0.16$, 0.84, and 0.21). The last digits of the screening values are evenly distributed (9.9 to 10.2 percent each), so the values do not bunch at round numbers in general. Applied at the other whole-number values from 115 to 155 mg/dL, the test rejects at the 5 percent level in three of 32 windows, those centered on 149, 150, and 151 mg/dL. All three contain a small spike at 150 mg/dL (489 screening values, against 433 at 149 and 397 at 151), which is well above the highest cutoff; none of the windows centered within 2 mg/dL of a cutoff rejects.

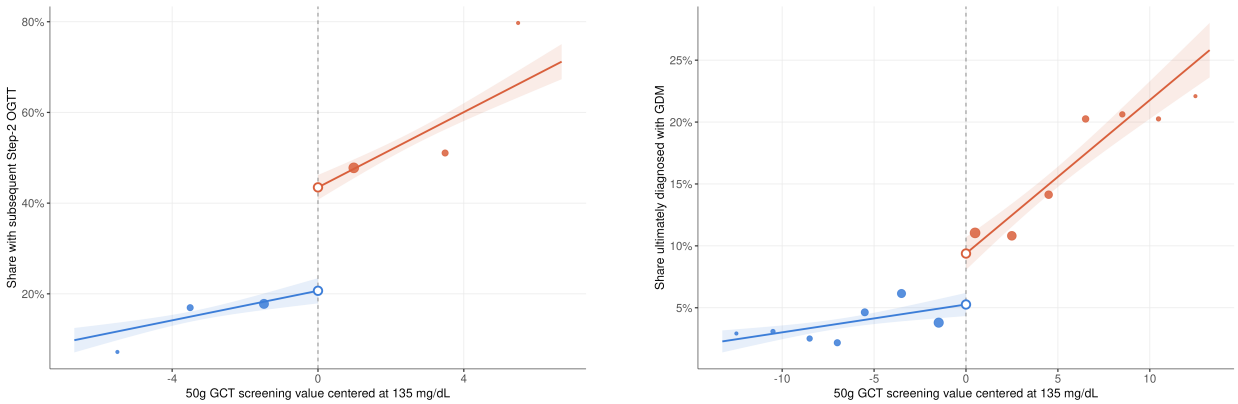
Figure A.2 shows the probability of a subsequent diagnostic test and of an eventual GDM diagnosis around the 135 mg/dL cutoff. Crossing the cutoff raises the probability of a subsequent diagnostic test by 22.8 percentage points (SE 2.0) and the probability of an eventual GDM diagnosis by 4.1 percentage points (SE 0.8), both estimated with the optimal-bandwidth estimator. The response differs across cutoffs. The probability of a subsequent diagnostic test rises by 9.9 points at 130 mg/dL, 22.8 points at 135 mg/dL, and 20.2 points at 140 mg/dL, and the probability of a GDM diagnosis rises by 4.1 points at 135 mg/dL and 5.0 points at 140 mg/dL but shows no detectable change at 130 mg/dL (0.1 points, $p = 0.81$). This pattern suggests that 135 and 140 mg/dL are the cutoffs most often applied in these data.

Figure A.1: Distribution of 50-gram glucose challenge test values



Notes: Kernel density of the 1-hour glucose value after a 50-gram glucose load (76,501 screening values). The vertical lines mark the three accepted screening cutoffs (130, 135, and 140 mg/dL). Author's calculations from MarketScan claims data.

Figure A.2: Screening at the 135 mg/dL cutoff: subsequent diagnostic test and GDM diagnosis



(a) Subsequent diagnostic test observed

(b) GDM diagnosed

Notes: Points are bin means against the 50-gram glucose challenge test value, centered at 135 mg/dL. Lines are local-polynomial fits with shaded pointwise 95% confidence bands, estimated with the optimal-bandwidth estimator (triangular kernel, own MSE-optimal bandwidth per panel: (a) ±6.7; (b) ±13.3 mg/dL). Panel (a) is an indicator that a subsequent 100-gram glucose tolerance test is observed in the claims (N = 73,561); panel (b) is an indicator that a GDM diagnosis is recorded through delivery, in the sample for which the pregnancy window is observable (N = 76,501). Standard errors are clustered by woman. Author's calculations from MarketScan claims data.

Appendix B: Laboratory Result Availability

Numerical laboratory values are not observed for all pregnancies with claims evidence of glucose testing. Table B.1 asks who the women with an observed laboratory value are, by comparing pregnancy episodes with a billed glucose-test claim that do and do not have a value in the Lab Database. The two groups differ most in the health-plan-reporting flag (77.0 percent of pregnancy episodes with a value against 25.5 percent of those without; standardized difference 1.20). They also differ in plan type (point-of-service plans are 35.4 percent of pregnancy episodes with a value against 10.4 percent of those without) and region (the North Central region is 7.1 percent against 21.3 percent). Differences in maternal age (0.16) and across calendar years (roughly 0.10 at most) are small, and only 2 of 150 metropolitan-area codes exceed 0.20. The pattern is similar under the wider claim definition. The estimates in this paper therefore describe women whose laboratory values are captured. The analytic cohort should be interpreted as the subset of commercially insured pregnancies for which MarketScan captures a numerical fasting glucose result, not as a representative sample of all glucose-tested pregnancies in CCAE.

Table B.1: Pregnancy episodes with a glucose-test claim, with and without a laboratory value in the Lab Database

	Narrow claim definition			Wide claim definition		
	With value	Without value	Std. diff.	With value	Without value	Std. diff.
Pregnancy episodes with a billed glucose-test claim	69,062	791,858		74,192	1,905,549	
Share of pregnancy episodes with a laboratory value		8.0%			3.7%	
Maternal age (years), mean	31.8	31.0	0.16	31.8	30.5	0.25
Enrollment record supplied by a health plan	77.0%	25.5%	1.20	77.5%	30.5%	1.07
Region						
Northeast	21.8%	17.3%	0.11	22.1%	15.7%	0.16
North Central	7.1%	21.3%	-0.41	7.3%	20.0%	-0.38
South	51.2%	41.9%	0.19	50.9%	48.1%	0.06
West	19.8%	19.2%	0.02	19.7%	16.0%	0.10
Unknown	0.0%	0.3%	-0.05	0.0%	0.2%	-0.05
Plan type						
Basic/major medical	0.0%	0.0%	-0.01	0.0%	0.0%	-0.01
Comprehensive	0.1%	1.7%	-0.17	0.1%	1.6%	-0.16
EPO	0.2%	0.8%	-0.08	0.2%	0.8%	-0.08
HMO	21.1%	12.7%	0.23	21.3%	13.0%	0.22
POS	35.4%	10.4%	0.62	35.4%	11.5%	0.59
PPO	29.7%	49.8%	-0.42	28.8%	49.0%	-0.42
POS with capitation	1.5%	0.3%	0.13	2.4%	0.6%	0.15
CDHP	6.7%	10.0%	-0.12	6.5%	9.8%	-0.12
HDHP	4.8%	12.5%	-0.28	4.7%	11.8%	-0.26
Missing	0.4%	1.9%	-0.14	0.6%	1.9%	-0.13
Metropolitan-area codes with |std. diff.| above 0.20		2 of 150			2 of 157	
Largest |std. diff.| across calendar years		0.10			0.06	

Notes: Pregnancy episodes, 2015–2025, whether or not a delivery is observed, with a billed procedure claim for glucose testing. The narrow definition counts standalone glucose and glucose-tolerance procedure codes; the wide definition adds basic and comprehensive metabolic panels, which are ordered for routine prenatal care and are much less specific to gestational-diabetes screening. "With value" pregnancy episodes also have a numerical fasting-glucose laboratory result in the Lab Database; "Without value" pregnancy episodes do not. Entries are the share of each group in the category (age: the group mean); the standardized difference is the difference between the two groups scaled by their pooled standard deviation, signed as with a value minus without a value. Values above roughly 0.10 are conventionally read as meaningful imbalance. Metropolitan area has too many categories to tabulate, so it is summarized by the number of codes with a standardized difference above 0.20 in absolute value, among codes with at least 30 pregnancy episodes in each group. Region and plan type follow the MarketScan data dictionary. EPO: exclusive provider organization; HMO: health maintenance organization; POS: point of service; PPO: preferred provider organization; CDHP: consumer-directed health plan; HDHP: high-deductible health plan. Descriptive only; these comparisons do not identify any causal effect. Author's calculations from MarketScan claims data.

Appendix C: Pre-existing Diabetes

The analytic sample is defined by a fasting glucose result linked to a pregnancy. The sample may therefore include women with pregestational diabetes. The women with diabetes before the lab test make up about 3 percent of testing episodes (3.2 percent have a diabetes diagnosis code before the test, and 4.7 percent meet a broader definition that adds pregnancy pre-existing diabetes codes, long-term insulin or oral hypoglycemic status codes, and prior insulin or glyburide fills).

This appendix measures how much pre-existing diabetes the sample contains and whether it affects the results. There is no single claims code for pre-existing type 1 or type 2 diabetes. I use diabetes diagnosis codes (E10 for type 1, E11 for type 2, and E08, E09, and E13 for other or secondary diabetes), the pregnancy codes for pre-existing diabetes (O24.0, O24.1, O24.3, and O24.8), status codes for long-term insulin (Z79.4) and oral hypoglycemic (Z79.84) use, and insulin, glyburide, and metformin fills in the 180 days before the test. The window starts at the beginning of the continuous enrollment span that contains the test and ends the day before it. Gestational diabetes (O24.4) is the treatment and is never a flag. The broad flag combines any diabetes diagnosis code, a pregnancy pre-existing diabetes code, Z79.4 or Z79.84, and an insulin or glyburide fill. Metformin and GLP-1 fills are not part of it, because metformin is also used for polycystic ovary syndrome and GLP-1 agonists for weight loss. Diagnosis codes are ICD-10-CM only, so claims dated before October 2015 are not captured, and the prior-fill flags require observed pharmacy coverage, so all shares are lower bounds.

Table C.1 reports the prevalence of each flag. Table C.2 re-estimates the first stage and the principal outcomes after excluding flagged episodes.

Table C.1: Pre-existing diabetes before the fasting glucose test

Evidence before the test	All episodes	≥365 days of look-back	Below 95 (±10)	At or above 95 (±10)	Jump at 95 (pp)	p
Any diabetes diagnosis code (E08–E11, E13)	3.17	4.34	3.18	5.37	−0.36	0.510
At least two claim dates ≥ 7 days apart	1.37	1.93	1.20	2.52	−0.02	0.945
Type 1 (E10)	0.38	0.48	0.22	0.63	−0.12	0.499
Type 2 (E11)	2.91	4.03	2.98	5.07	−0.26	0.616
Other or secondary (E08, E09, E13)	0.19	0.27	0.13	0.26	0.02	0.831
Pregnancy pre-existing diabetes code (O24.0, O24.1, O24.3, O24.8)	1.11	1.40	1.14	2.55	0.22	0.512
Long-term insulin use status (Z79.4)	0.54	0.76	0.52	1.43	−0.27	0.383
Long-term oral hypoglycemic status (Z79.84)	0.49	0.72	0.63	1.04	0.30	0.227
Insulin fill in the prior 180 days	1.58	1.73	1.91	4.36	0.06	0.922
Metformin fill in the prior 180 days	3.66	3.86	4.51	6.47	0.64	0.337
Broad flag	4.70	6.43	5.47	9.26	−0.32	0.655
Episodes with a pre-test window	85,457	46,654	20,381	6,512		$N = 85{,}457$

Notes: Share of fasting glucose testing episodes, in percent, with evidence of diabetes in claims before the test, from the start of the continuous enrollment span containing the test through the day before it. Diagnosis codes are ICD-10-CM; claims dated before October 2015 use ICD-9 and are not captured, so look-back flags are undercounted for early tests. Only 55 percent of episodes have at least 365 days of look-back; the second column restricts to them. The third and fourth columns restrict to episodes within ±10 mg/dL of 95 mg/dL, below and at or above the cutoff. The last two columns report the optimal-bandwidth (rdrobust) discontinuity at 95 mg/dL in percentage points, with its p-value based on standard errors clustered by woman. Fill rows are limited to episodes with observable pharmacy coverage. The broad flag is any diabetes diagnosis code, a pregnancy pre-existing diabetes code, Z79.4 or Z79.84, or an insulin or glyburide fill in the prior 180 days; metformin and GLP-1 fills are not included. Author's calculations from MarketScan claims data.

Table C.2: Excluding women with pre-existing diabetes evidence: first stage and principal outcomes

	Full sample		Excluding any diabetes dx		Excluding broad flag		
Outcome	Estimate	p	Estimate	p	Estimate	p	N (full)
Panel A. First stage and care after the test (percentage points)							
Any GDM diagnosis (first stage)	10.74	<0.001	10.72	<0.001	11.20	<0.001	85,611
DSMT claim, days 0–29 (excluding same-day claims)	3.42	<0.001	3.48	<0.001	3.52	<0.001	85,611
MNT claim, days 0–29	1.76	0.004	1.81	0.004	1.92	0.003	85,611
Prenatal biophysical profile	2.24	0.018	2.07	0.035	2.36	0.016	85,611
Metformin initiation, 90 days	1.99	0.001	2.08	<0.001	2.14	<0.001	59,433
Insulin initiation, 90 days	1.54	0.033	1.18	0.124	1.31	0.094	59,433
Glyburide initiation, 90 days	1.19	0.007	1.25	0.006	1.24	0.009	59,433
Panel B. Payments from the test through delivery (dollars)							
DSMT/MNT total payment	$12.14	<0.001	$11.91	<0.001	$12.47	<0.001	72,685
Panel C. Infant outcomes (percentage points)							
Heavy for gestational age (P08.1)	−1.45	<0.001	−1.57	<0.001	−1.54	0.002	37,464
NICU admission (revenue code)	−1.33	0.011	−1.37	0.012	−1.32	0.017	37,464
Preterm or extreme immaturity (P07)	−1.32	0.038	−1.19	0.065	−1.29	0.040	37,464
Neonatal jaundice (P59)	−2.10	0.059	−2.18	0.059	−2.13	0.071	37,464
Panel D. Delivery (percentage points)							
C-section procedure	−1.32	0.279	−1.47	0.281	−1.37	0.325	62,795
Obstetric anesthesia procedure	0.17	0.918	0.03	0.987	0.26	0.882	62,795

Notes: Optimal-bandwidth (rdrobust) reduced forms at the 95 mg/dL cutoff, estimated separately in each sample, with p-values based on standard errors clustered by woman. The first column pair is the full sample of 85,611 testing episodes; the second excludes episodes with any diabetes diagnosis code (E08–E11, E13) before the test (82,899 remain in the first-stage sample); the third excludes episodes with the broad flag defined in Table C.1 (81,595 remain). Episodes with no observed pre-test window are kept in both exclusion samples. The last column reports the full-sample N; each outcome's sample in the exclusion columns is smaller by the excluded episodes that belong to it. Author's calculations from MarketScan claims data.

Appendix D: Glucose Testing During Pregnancy

The analysis sample consists of women with a fasting glucose result, so it does not show how many pregnant women are tested for gestational diabetes at all. This appendix uses the claims to measure how common glucose testing is among delivered pregnancies. Because the claims record billed procedures and not the purpose of a test, the shares below describe claims evidence of testing, not the true screening rate.

The sample starts from the 2,730,453 deliveries identified in the claims, of which 2,417,404 match to a detected pregnancy episode. I keep the 1,753,281 that are continuously enrolled from 280 days before delivery through delivery, so that testing at any point in the pregnancy would appear in the claims, and exclude 40,684 pregnancies with a diagnosis of pre-existing diabetes recorded before the pregnancy or before 24 weeks, which leaves 1,712,597. Because episode matching is poor for 2015, the first year of ICD-10 coding, I report the 1,653,007 deliveries in 2016–2025. The start of pregnancy is set 280 days before delivery.

Table D.1 reports the share with at least one qualifying claim between the start of pregnancy and delivery. Claims for a glucose draw (CPT 82947, 82948, or 82950), the codes used for Step 1 glucose challenge tests but also for other glucose measurements, appear in 77.8 percent of pregnancies, and the post-dose glucose code (CPT 82950) alone in 71.3 percent. Adding the glucose tolerance test codes (CPT 82951 and 82952) and glucose by monitoring device (CPT 82962) gives 81.7 percent, of which the glucose tolerance test codes account for 18.5 percent. The broadest definition, which adds hemoglobin A1c and routine metabolic and health panels that include a glucose measurement, reaches 90.3 percent. That last number measures recorded glucose-related testing, and routine prenatal panels are included in it, so it includes tests used for other purposes and can miss tests outside the available claims.

The results are lower when I limit the window to the period when guidelines recommend universal screening. For claims from 24 weeks of gestation to delivery, 74.7 percent of pregnancies have a glucose draw claim and 69.2 percent a post-dose glucose claim. When I anchor gestational age on the weeks recorded in pregnancy diagnosis codes, which is possible for the 1.2 million pregnancies with a plausible start, 61.9 percent have a glucose draw claim and 57.9 percent a post-dose glucose claim between 24 and 28 weeks.

Two checks bear on how well the claims capture testing. First, the Lab Database records the 50 gram screen value (LOINC 1504-0) for a subset of enrollees. Among the 418,607 pregnancies in the stratum in which laboratory results are most likely to be observed, 96.5 percent of those with a recorded screen value also have a glucose draw claim, so the claims rarely miss a test that the laboratory data confirm. Second, women with pre-existing diabetes should not receive a routine screening test. They are indeed less likely to have a post-dose glucose claim from 24 weeks to delivery (34.9 percent against 69.2 percent), although not absent from the claims. A claim can also reflect a diagnostic test or glucose monitoring, and a test billed through a source not visible in the claims is missed.

Table D.1: Claims evidence of glucose testing for gestational diabetes during pregnancy

Evidence of testing in claims	Deliveries	Share (percent)
Glucose-draw claims: CPT 82947, 82948, or 82950	1,285,986	77.8
Post-dose glucose only (CPT 82950)	1,178,210	71.3
Glucose-draw and tolerance-test claims: CPT 82947, 82948, 82950, 82951, 82952, or 82962	1,349,703	81.7
Glucose tolerance test only (CPT 82951 or 82952)	306,253	18.5
Broader glucose-related testing claims: all glucose CPT above, HbA1c (CPT 83036 or 83037), or a metabolic or health panel that includes glucose (CPT 80048, 80050, 80053, or 80069)	1,492,034	90.3

Notes: Share of delivered pregnancies with at least one claim for the listed procedure codes between the start of pregnancy and delivery, in claims for office, hospital, facility, and other services. The sample is 1,653,007 deliveries in 2016–2025 with continuous enrollment throughout the pregnancy and no diagnosis of pre-existing diabetes before 24 weeks of gestation. The start of pregnancy is set 280 days before delivery. CPT 82950 is glucose after a glucose dose, the procedure code for the 50 gram challenge test; CPT 82947 and 82948 are blood glucose (laboratory and test strip), which are also used outside screening; CPT 82951 and 82952 are the multiple-specimen glucose tolerance test (Step 2, or the one-step test); CPT 82962 is glucose by monitoring device; CPT 83036 and 83037 are hemoglobin A1c, which is used for early-pregnancy diabetes screening and diabetes monitoring; and CPT 80048, 80050, 80053, and 80069 are routine metabolic, general health, and renal panels that include a glucose measurement. The last row therefore measures any glucose-related test and is an upper bound on testing for gestational diabetes, not a screening rate. A pregnancy without a claim may have been tested through a source not visible in claims. Author's calculations from MarketScan claims data.

Appendix E: Identification of Pregnancy Episodes and Deliveries

Pregnancies and deliveries are identified from claims alone; the data contain no birth certificates or other external pregnancy registry. This appendix describes three steps: the construction of pregnancy episodes, the linkage of each fasting-glucose testing episode to a pregnancy episode, and the identification of a delivery following the test. Pregnancy episodes determine which laboratory results enter the analytic cohort. Deliveries are identified from a separate delivery definition and are dated relative to the glucose test rather than to the pregnancy episode.

Claims and pregnancy-anchor codes

Pregnancy episodes are built from ICD-10-CM diagnosis codes in every diagnosis position on inpatient admission, inpatient service, outpatient service, and facility claims for female enrollees. Procedure codes are not used for this step. Codes are sorted into the families listed in Table E.1. Nine families act as *anchors*: supervision of normal and high-risk pregnancy, pregnancy with abortive outcome, hypertensive and other maternal disorders of pregnancy, maternal care for the fetus and placenta, labor and delivery complications, the delivery encounter, and the outcome of delivery. Gestational-week codes (Z3A) help date and separate pregnancies but cannot by themselves define one. Postpartum complications and obstetric sequelae are attached to an existing pregnancy episode when they occur during it or within 42 days after its last anchor claim; they never start a pregnancy episode.

Grouping claim days into pregnancy episodes

All qualifying claims for a woman are first collapsed to claim days. Claim days carrying an anchor or a gestational-week code are then read in date order. A new pregnancy episode begins when any of three conditions holds: more than 90 days have passed since the previous qualifying claim day; the new day would extend the pregnancy episode beyond 320 days from its first day; or the recorded gestational week is inconsistent with the one recorded earlier in the pregnancy episode, meaning it falls, or it differs from the week implied by the elapsed calendar time by more than three weeks. Sequences that contain only gestational-week codes are discarded. The pregnancy episode's first anchor claim must fall within one of the woman's enrollment spans. The resulting file contains 6,422,059 pregnancy episodes among 3,175,135 women.

Table E.1: ICD-10-CM code families used to identify pregnancy episodes

ICD-10-CM codes	Description	Role
O00–O08	Pregnancy with abortive outcome	Anchor
O09	Supervision of high-risk pregnancy	Anchor
O10–O16	Hypertensive disorders in pregnancy	Anchor
O20–O29	Other maternal disorders in pregnancy	Anchor
O30–O48	Maternal care for fetus, placenta, and amniotic cavity	Anchor
O60–O77	Preterm labor and labor and delivery complications	Anchor
O80–O84	Encounter for delivery	Anchor
Z37	Outcome of delivery	Anchor
Z34	Supervision of normal pregnancy	Anchor
Z3A	Weeks of gestation	Dating and splitting only
O85–O92	Postpartum complications	Attached only
O94–O9A	Obstetric sequelae and other obstetric conditions	Attached only

Notes: A pregnancy episode must contain at least one claim day with an anchor code. Gestational-week codes enter the grouping rule and date the start of gestation but cannot form a pregnancy episode by themselves. Attached families are assigned to the nearest pregnancy episode that they fall within or follow by no more than 42 days, and never start a pregnancy episode.

Start of gestation

When a pregnancy episode contains gestational-week codes, each such claim day implies a start of gestation equal to the claim date minus the recorded week times seven days; when several week codes appear on the same day, the midpoint is used. The pregnancy episode's estimated start is the median of these implied dates. A pregnancy episode is flagged as having conflicting gestational-age information when codes on the same day differ by more than one week or when its implied start dates span more than 21 days. The flag does not change the estimate. Without any gestational-week code, the estimated start is the date of the first anchor claim. The estimated start is used to validate the construction below; the linkage and outcome windows are defined from the claim dates and the laboratory date.

Linking a fasting-glucose test to a pregnancy episode

A fasting-glucose testing episode is linked to a pregnancy episode when the laboratory date falls between 90 days before the pregnancy episode's first anchor claim and 42 days after its last anchor claim. When more than one pregnancy episode qualifies, I choose the pregnancy episode closest to the laboratory date, with distance equal to zero when the test falls inside the pregnancy episode. Any remaining tie is broken in favor of the pregnancy episode that began earlier. Testing episodes that cannot be linked are excluded from the analytic cohort. This rule can include tests outside the observed anchor span and does not by itself establish

that the test occurred before delivery.

Delivery and the end of the pregnancy window

Deliveries are identified from a separate claims definition: an ICD-10-CM delivery code (O80, O82, O75, or Z37), a delivery diagnosis-related group (765–768, 774, or 775), or a claim assigned to major diagnostic category 14 (pregnancy, childbirth, and the puerperium). Delivery claims for the same woman within 30 days of each other form one delivery episode, dated by its earliest claim, which must fall within an enrollment span. The delivery assigned to a testing episode is the earliest one dated on or after the laboratory date. A delivery is counted as observed when it occurs within 294 days of the test. The pregnancy window ends on the delivery date when a delivery is observed. Otherwise it ends at the earliest of 294 days after the test, the end of the woman's enrollment span, or the end of the data.

Validation

I check the pregnancy episode construction against a separate list of deliveries confirmed by a delivery diagnosis code or delivery diagnosis-related group, excluding deliveries identified only by the major diagnostic category. A delivery is matched when it falls between the first anchor claim of one of the woman's pregnancy episodes and 42 days after its last. Of 2,730,453 such deliveries, 2,431,503 (89.1 percent) are matched, above the prespecified minimum of 75 percent. Almost all unmatched deliveries occurred before October 2015, when diagnoses were still coded in ICD-9-CM; from October 2015 onward, 0.8 percent of deliveries are unmatched. Because delivery codes are themselves anchors, this check mainly confirms that the grouping rules neither drop nor split pregnancies that end in a coded delivery. The second check concerns dating. Among matched deliveries whose pregnancy episode has a gestational-week code, 73.7 percent have an estimated start of gestation between 200 and 320 days before delivery, above the prespecified minimum of 70 percent. In pregnancy episodes without conflicting gestational-age codes, the median interval from estimated start to delivery is 251 days.

Limitations

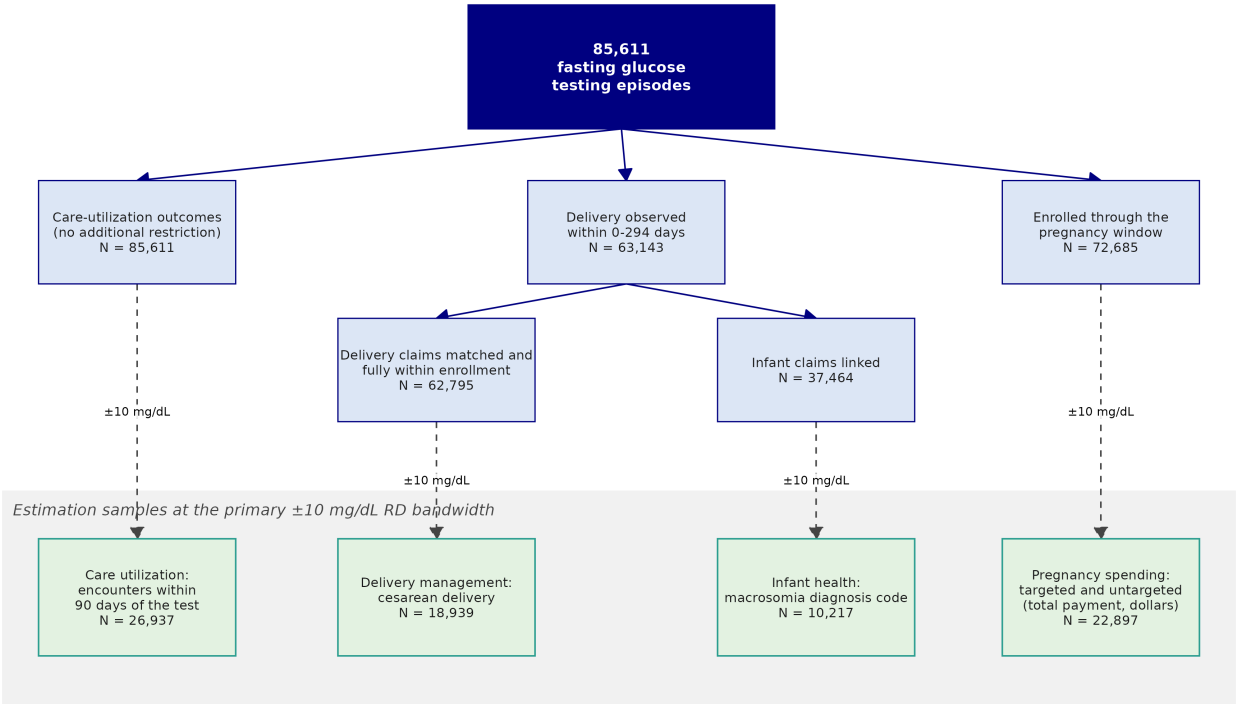
A pregnancy episode is a claims construct, not a count of pregnancies. Nearly half of all pregnancy episodes (49.2 percent) span 30 days or less, and the gestational-week rule can split one pregnancy into several pregnancy episodes when week codes are recorded inconsistently. Pregnancy losses enter through the abortive-outcome anchor codes, but early losses without an obstetric claim are missed, and a loss cannot always be distinguished from a pregnancy episode that ends when coverage ends. A delivery can be identified from the major diagnostic category alone, which also covers obstetric admissions that are not deliveries. The validation checks above do not test that part of the definition, so results that use a delivery date may

be sensitive to it. Finally, a woman who leaves a contributing plan before delivery has no observed delivery; Section 3.1 separates these cases from pregnancies observed through the full 294-day window without a delivery.

Appendix F: Sample Construction Detail

Figure F.1 diagrams the construction of the analytic cohort described in Section 3, together with each downstream restriction and estimation sample, and reports the sample size at each step.

Figure F.1: Sample construction: the analytic cohort and its downstream estimation samples



Notes: Branches below the base sample are overlapping restrictions applied by outcome family, not a partition, so the branch sample sizes do not sum to the base N. Green boxes (bottom row) restrict each sample to the fixed ±10 mg/dL RD bandwidth, which is a specification choice rather than an eligibility restriction; blue boxes report full-sample eligibility counts. Author's calculations from MarketScan claims data.

The analytic cohort is the set of 85,611 fasting-glucose testing episodes, from 79,689 pregnancies among 74,194 women, described in Section 3.1. Each of the four sample families starts from this cohort and applies the observation requirements of one outcome family; outcomes within a family can carry further requirements, so a family can contain several estimation samples. The eligible sample contains all episodes that meet an outcome's observation requirements. I supply this sample to the bandwidth-selection procedure, which chooses a bandwidth separately for each outcome. The local point estimate uses the episodes within the selected bandwidth on each side of the cutoff. A bandwidth is a specification choice, not an observation requirement. The tables below report the eligible sample sizes and the sample sizes within ±10 mg/dL. The samples overlap; they do not partition the cohort.

Appendix Tables F.1 and F.2 list the estimation samples behind the principal outcomes, with the additional requirement that defines each, its sample size, and its first stage. Outcomes that share a sample share a first stage. The first table covers care, pregnancy payments, and infant outcomes (Panels A–C); the second covers maternal outcomes at delivery and postpartum outcomes (Panels D–E).

Table F.1: Estimation samples and first stages for the principal outcomes: care, pregnancy payments, and infant outcomes

Sample	Additional requirement	Principal outcomes	Eligible sample	Within ±10 mg/dL	Women	First stage (pp)	F
Panel A. Care after the test							
P1	Analytic cohort	DSMT claim, days 0–29; DSMT claim, days 90–179; MNT claim, days 0–29; Biophysical profile, days 90–179; Encounter days, 90 days; Pharmacy-fill days, 90 days	85,611	26,937	24,411	11.6 (1.3)	84.1
P2	Observable pharmacy coverage	Metformin fill, 90 days; Insulin fill, 90 days	78,346	24,635	22,335	12.1 (1.3)	83.9
P3	Observable pharmacy coverage before and after the test	Insulin initiation, 90 days; Metformin initiation, 90 days	59,433	18,692	16,918	11.7 (1.5)	58.0
Panel B. Pregnancy payments							
P4	Continuous enrollment from the test through the end of the spending window	DSMT/MNT total payment; DSMT/MNT patient payment; Non-DSMT/MNT total payment; Total pregnancy payment	72,683–72,685	22,897	21,040	11.3 (1.4)	69.9
Panel C. Infant outcomes							
P5	Linked infant record	Macrosomia; Heavy for gestational age (P08.1); NICU admission	37,464	10,217	9,551	12.5 (2.1)	33.8

Notes: Each row is one estimation sample (Panels A–C; Panels D–E are in Appendix Table F.2). The first stage is the discontinuity in the probability of a recorded gestational diabetes diagnosis at 95 mg/dL, estimated by local linear regression with a uniform kernel and a bandwidth of ±10 mg/dL on the observations that enter the outcome regression (standard error in parentheses; women are the clusters). Outcomes listed in the same row use identical observations and therefore share a first stage. Per-complier estimates that use optimal bandwidths are scaled by the first stage at each outcome's own bandwidth and are not listed here. Eligible sample is the number of fasting-glucose testing episodes that meet the additional requirement, before any bandwidth restriction; it is the sample used by the optimal-bandwidth estimates. Within ±10 mg/dL is the number of those episodes within 10 mg/dL of the cutoff, and women is the number of distinct women among them. The additional requirement is in addition to membership in the analytic cohort (Appendix F describes the sample construction). The complete registry, covering all 1,212 estimated per-complier rows and 63 distinct samples across the outcomes reported in the paper and its appendices, is in the replication materials. Author's calculations from MarketScan claims data.

Table F.2: Estimation samples and first stages for the principal outcomes: maternal outcomes at delivery and postpartum outcomes

Sample	Additional requirement	Principal outcomes	Eligible sample	Within ±10 mg/dL	Women	First stage (pp)	F
Panel D. Maternal outcomes at delivery							
P6	Delivery episode entirely within enrollment	Cesarean delivery; Obstetric anesthesia	62,795	18,939	17,462	12.0 (1.5)	64.4
P7	Inpatient admission record within the delivery episode (payment observed)	Delivery payment	30,906	8,330	8,077	16.1 (2.4)	46.3
Panel E. Postpartum and later outcomes							
P8	Observed delivery; continuous enrollment through day 84 after delivery	Encounter, days 43–84 after delivery	54,291	16,378	15,084	11.4 (1.6)	49.8
P9	Observed delivery; continuous enrollment through 365 days after delivery	Total payment, postpartum year 1	30,758	9,344	8,600	10.1 (2.2)	21.7
P10	Observed delivery; continuous enrollment through 36 months after delivery	Incident type 2 diabetes, 36 months	10,843	3,278	3,010	9.4 (3.8)	6.2

Notes: Each row is one estimation sample (Panels D–E; Panels A–C are in Appendix Table F.1). The first stage is the discontinuity in the probability of a recorded gestational diabetes diagnosis at 95 mg/dL, estimated by local linear regression with a uniform kernel and a bandwidth of ±10 mg/dL on the observations that enter the outcome regression (standard error in parentheses; women are the clusters). Outcomes listed in the same row use identical observations and therefore share a first stage. Per-complier estimates that use optimal bandwidths are scaled by the first stage at each outcome's own bandwidth and are not listed here. Eligible sample is the number of fasting-glucose testing episodes that meet the additional requirement, before any bandwidth restriction; it is the sample used by the optimal-bandwidth estimates. Within ±10 mg/dL is the number of those episodes within 10 mg/dL of the cutoff, and women is the number of distinct women among them. The additional requirement is in addition to membership in the analytic cohort (Appendix F describes the sample construction). The complete registry, covering all 1,212 estimated per-complier rows and 63 distinct samples across the outcomes reported in the paper and its appendices, is in the replication materials. Author's calculations from MarketScan claims data.

Care utilization

The care-utilization sample applies no observation requirement beyond membership in the cohort. Its outcomes are claims-based measures of care from the test forward, such as encounters, drug claim-days, diabetes education and nutrition services, provider contact, and fetal surveillance. Each is measured over a window that starts at the date of the diagnostic

test, so none requires an observed delivery. The care-intensity counts are truncated at the end of the pregnancy window. The diabetes education and nutrition event-time windows (0–29, 30–89, 90–179, and 180–365 days) are fixed offsets from the test date that are not truncated at delivery, so the later windows include postpartum days. Fetal surveillance excludes claims dated on or after the end of the pregnancy window. For these claims-based measures, a woman with no qualifying claim, no enrollment, or no remaining pregnancy time in a window is recorded as having no service in it; the medication outcomes described below are the exception, because they are missing when pharmacy coverage is not observed. All 85,611 fasting glucose testing episodes are eligible. A few outcomes carry further requirements and therefore use smaller samples (Section 5.2): medication any-fill outcomes require observable pharmacy coverage, leaving 78,346 testing episodes; medication initiation additionally requires the pharmacy history needed to identify prior use, leaving 59,433 testing episodes; and the fetal-surveillance rates per observed week are defined only for testing episodes with pregnancy time remaining in the band. The diabetes education windows are not restricted to testing episodes in which the window is observed.

Delivery management

Delivery outcomes require an observed delivery. A delivery is observed when it is identified in the claims within 294 days after the test, which holds for 63,143 of the 85,611 fasting glucose testing episodes (73.8 percent). Delivery procedures are further restricted to deliveries whose delivery episode lies entirely inside the woman's enrollment, because the absence of a procedure claim cannot be interpreted once enrollment ends during the delivery. This leaves 62,795 testing episodes. A delivery whose delivery episode extends beyond the end of enrollment is coded missing; within enrollment, a delivery with no procedure claim is coded as having no procedure. Delivery payment is observed only for deliveries with an inpatient admission record dated within the delivery episode. That is the case for 30,906 deliveries (Appendix J).

Infant health

Infant outcomes require an additional linkage between the mother and an infant enrollment record. I identify an infant as an enrollee sharing the mother's family identifier whose birth year matches the delivery year, using the enrollee's relationship to the primary subscriber (child or dependent) where it is available. When more than one candidate is available, I select the infant whose enrollment begins closest to the delivery date, and unresolved ties remain unlinked. This procedure links an infant to 37,464 testing episodes, or 59.3 percent of the delivery observed sample. Infant outcomes therefore use these 37,464 testing episodes.

Pregnancy spending

For pregnancy spending outcomes, I require continuous enrollment from the fasting glucose test through the end of the pregnancy spending window. The window ends at delivery when a delivery is observed. Otherwise, it ends at the earliest of 294 days after the test, the end of the woman's enrollment span, or the end of the available data. This produces a spending window observable subsample of 72,685 fasting glucose testing episodes, or 84.9 percent of the full cohort.

Unobserved deliveries

An unobserved delivery does not necessarily indicate that the pregnancy ended without a birth. In some cases, enrollment or data coverage ends before the full delivery observation window closes. In other cases, no delivery is recorded despite continued enrollment. The claims data do not identify a single explanation for these cases. Appendix J tests whether crossing the 95 mg/dL fasting threshold changes the probability of entering the delivery observed, spending window observable, or infant linked samples.

Outcomes measured after delivery

For outcomes measured after delivery, I require continuous enrollment through the end of the relevant follow-up window. A woman who leaves MarketScan before that point is excluded from the corresponding outcome rather than assigned a value of zero. Among women who remain observable throughout the window, the absence of a qualifying claim is recorded as no service or no diagnosis. Consequently, the estimation sample becomes smaller at longer follow-up horizons as women leave MarketScan or reach the end of the available data.

Appendix G: Outcome Definitions

Table G.1 summarizes the measurement window and code list for each outcome family, with one row per outcome family rather than per individual outcome.

Table G.1: Outcome definitions: measurement windows and code lists

Domain	Outcome family	Measurement window	Code list
Diagnosis	Any GDM diagnosis	Headline measure: test date through the later of (a) the first 30 days starting on the test date (shortened at the end of enrollment or of the data) and (b) delivery, when a delivery is observed within 294 days of the test, or otherwise the earliest of 294 days after the test, the end of enrollment, or the end of the data. Component measures used in the first-stage checks: a diagnosis within the first 30 days starting on the test date (observable only for fasting glucose testing episodes enrolled through that window), and a diagnosis through delivery or, with no observed delivery, through 294 days (observable when a delivery is observed or the full 294-day window is covered by enrollment and data)	Gestational diabetes diagnosis code (ICD-10-CM O24.4x); the headline measure is defined for every testing episode, with no enrollment-continuity requirement

continued on next page

(continued from previous page)

Domain	Outcome family	Measurement window	Code list
Care/ utilization	Diabetes self-management training (DSMT)	Any claim (and the number of claim days) from the test date through 180 days after the test, truncated at the pregnancy window end. The event-time buckets (-30 to -1 (placebo), 0–29, 30–89, 90–179, and 180–365 days from the test) are fixed day offsets from the test date: they are not truncated at delivery or at the pregnancy window end, and they are coded zero, not missing, for fasting glucose testing episodes with no enrollment in the bucket window or with the window beyond the end of the data (only the placebo bucket is missing, when enrollment does not begin 30 days before the test). A separate prenatal bucket runs from the test to the day before the pregnancy window end	Diabetes self-management training (HCPCS G0108/G0109)
Care/ utilization	Medical nutrition therapy (MNT)	Any claim (and the number of claim days) from the test date through 180 days after the test, truncated at the pregnancy window end; same event-time buckets, and the same fixed offsets and zero coding, as DSMT above	Medical nutrition therapy (CPT 97802-97804)

continued on next page

(continued from previous page)

Domain	Outcome family	Measurement window	Code list
Care/ utilization	Antidiabetic medication fill / initiation	Fill: 90 and 180 days after the test, truncated at the pregnancy window end. Initiation: a fill in that window with no fill of the same class in the 180 days before the test	Insulin fill; Metformin fill; Glyburide/glibenclamide fill; GLP-1 receptor agonist fill
Care/ utilization	Medication sequence among users	Among women with any qualifying fill; first observed agent, monotherapy vs. combination, and insulin added after an earlier metformin fill	First agent filled (insulin, metformin, or glyburide; same-day ties flagged); monotherapy vs. combination therapy; insulin added after an earlier metformin fill
Care/ utilization	Glucose self-monitoring supplies/devices	90 and 180 days after the test, truncated at the pregnancy window end; medical-benefit claims, not gated on drug-benefit coverage	Glucose test strips (HCPCS A4253); Lancets (HCPCS A4258/A4259); Glucometer (HCPCS E0607/E2100/ E2101); CGM supplies/devices (HCPCS A4238/A4239/E2103/ K0554); CGM interpretation services (CPT 95249-95251)

continued on next page

(continued from previous page)

Domain	Outcome family	Measurement window	Code list
Care/ utilization	Fetal surveillance	Prenatal (test to the day before the pregnancy window end) and four day bands after the test: 0–29, 30–89, 90–179, and 180–window-end days; any indicator, claim count, and count per observed week. Claims on or after the pregnancy window end are excluded, so a band that extends past the observed pregnancy covers only its observed part; the any indicator and the claim count are coded zero, not missing, when no pregnancy time remains in the band, and only the per-week rate is missing	Nonstress test (CPT 59025); biophysical profile (CPT 76818/ 76819); follow-up fetal growth ultrasound (CPT 76816)
Care/ utilization	General care intensity	90 and 180 days after the test, truncated at the pregnancy window end	Distinct days with any medical encounter (outpatient/inpatient/ facility/inpatient-service claims); distinct days with a numerical result in the laboratory-results file (not laboratory procedure claims); distinct days with a pharmacy fill
Care/ utilization	Provider contact, by specialty	Test date through the earlier of 180 days after the test or the pregnancy window end	Any claim whose rendering-provider specialty (MarketScan provider-type field) falls in each of eight groups: endocrinology, dietitian, obstetrics/ gynecology, midwife, primary-care physician, advanced-practice primary care, pediatrics/neonatal, other/ unclassified

continued on next page

(continued from previous page)

Domain	Outcome family	Measurement window	Code list
Care/ utilization	Postpartum metabolic screening	43-84 days after delivery; also 6- and 12-month windows	Any observed glucose or A1c laboratory result in the laboratory-results file (a broad measure of recorded metabolic testing, not completion of the recommended postpartum 75-gram OGTT)
Care/ utilization	Postpartum diabetes/prediabetes diagnosis	43-84 days after delivery	Postpartum diabetes diagnosis (ICD-10 E10-E13; may partly reflect more intensive postpartum testing); Postpartum prediabetes diagnosis (ICD-10 R73; may partly reflect more intensive postpartum testing)
Spending	Spending after the glucose test and before the pregnancy endpoint	Test date through the day before the pregnancy window end (delivery date when a delivery is observed within 294 days of the test; otherwise the earliest of 294 days after the test, the end of enrollment, or the end of the data); requires enrollment covering the full window	Sum of total payments (plan plus patient) on outpatient, facility, and inpatient claims, excluding pharmacy claims; inpatient spending is taken from the admission-level record only (itemized service lines excluded to avoid double counting). Nominal dollars, no price or inflation adjustment. Fasting glucose testing episode totals with a negative reported amount are floored at zero.
Spending	Spending after the glucose test and before the pregnancy endpoint, including pharmacy claims	Same window as above	Spending above plus total payments on pharmacy claims. Pharmacy spending is set to zero (not missing) for a fasting glucose testing episode with a matched enrollment record carrying the drug-coverage field and no pharmacy claim in the window; it is left missing, and so is the drug-inclusive total, only when no enrollment record can be matched or pharmacy payments are not available in the claim years

continued on next page

(continued from previous page)

Domain	Outcome family	Measurement window	Code list
Spending	DSMT/MNT spending	Test date through the day before the pregnancy window end, the same window and enrollment requirement as whole-pregnancy spending above (not a fixed number of days after the test)	Total payments on outpatient and facility claim lines carrying a DSMT (HCPCS G0108/G0109) or MNT (CPT 97802–97804) code
Spending	Delivery cost	Delivery episode (claims within 30 days of the index delivery claim collapsed into one delivery episode)	Total payment and plan net payment on the delivery admission
Delivery/ maternal	Delivery procedures	Delivery episode	Cesarean delivery (CPT 59510/ 59514/59515/59618/59620/59622); obstetric anesthesia (CPT 01960–01969)
Delivery/ maternal	Rare/severe maternal complications	Test date through the pregnancy window end	Maternal care for excessive fetal growth (O36.6x); postpartum hemorrhage (O72); shoulder dystocia (O66.0); third- or fourth-degree perineal laceration (O70.2/O70.3); HELLP syndrome or eclampsia (O14.2/O15); ICD-9-CM equivalents before October 2015
Infant	Macrosomia	Delivery date minus 1 day through delivery plus 30 days, on the linked infant's own claims	P08.0 (exceptionally large newborn, ≥4,500 g) and P08.1 (other heavy for gestational age, 4,000–4,499 g), reported separately and combined; P08.2 (post-term infant, not heavy for gestational age) excluded. A diagnosis-code proxy, not a measured birth weight.
Infant	Birth weight / gestational-age extremes	Same infant claim window as macrosomia	Low birth weight diagnosis (P07.0/ P07.1); Preterm/extreme immaturity diagnosis (P07.2/P07.3)

continued on next page

(continued from previous page)

Domain	Outcome family	Measurement window	Code list
Infant	Neonatal metabolic/ respiratory/injury diagnoses	Same infant claim window as macrosomia	Infant-of-diabetic-mother label diagnosis (P70.0/P70.1; records maternal diabetes exposure, not an infant health outcome); Neonatal hypoglycemia/diabetes-of-newborn diagnosis (P70.2-P70.9); Neonatal respiratory distress diagnosis (P22.x); Birth injury diagnosis (P10-P15)
Infant	NICU / nursery admission	Same infant claim window as macrosomia; identified from facility revenue codes on outpatient and inpatient service claims	NICU stay (revenue code 0172–0174); Normal newborn nursery stay (revenue code 0170/0171; reported separately from NICU)
Later maternal disease	Cumulative diagnosis, 12/24/36/ 60/96 months post-delivery	Delivery + 43 days through delivery + 365/ 730/1096/1826/2922 days; outcomes are first recorded diagnoses after postpartum day 42, and women with a qualifying code before the index test are excluded; requires continuous enrollment through the horizon	Type 2 diabetes diagnosis (E11.x); Prediabetes/impaired glucose diagnosis (R73.01/R73.02/R73.03/ R73.09); Hypertension diagnosis (I10-I13); Ischemic heart disease diagnosis (I20-I25); Heart failure diagnosis (I50); Stroke diagnosis (I63/I64); Cardiovascular disease composite (ischemic heart disease, heart failure, or stroke); Dyslipidemia diagnosis (E78.0-E78.5); Metabolic syndrome diagnosis (E88.81); Depression diagnosis (F32/F33; postpartum-specific codes F53.0/ O90.6 not used); Chronic kidney disease diagnosis (N18.x)
Later maternal disease/ recurrence	Subsequent pregnancy / next delivery	24/36/48 months after the index test	Any subsequent pregnancy episode or delivery identified in claims (see Section 3 on surveillance-sensitive diagnosis-code outcomes)

continued on next page

(continued from previous page)

Domain	Outcome family	Measurement window	Code list
Offspring, long run	Atopic disease	Birth through 365/730/ 1096/1826/2922 days on the child's own enrollment; cumulative (no exclusion for earlier diagnoses)	Asthma diagnosis (J45.x); Atopic dermatitis/eczema diagnosis (L20.x); Allergic rhinitis diagnosis (J30.1-J30.9; J30.0 vasomotor/ non-allergic rhinitis excluded); Urticaria diagnosis (L50.x); Any atopic disease diagnosis (composite: asthma, atopic dermatitis, allergic rhinitis, or urticaria)
Offspring, long run	Childhood obesity	Same horizons as atopic disease above; too few cases for estimation at the longer horizons	Obesity diagnosis (E66.0/E66.01/ E66.09/E66.1/E66.2/E66.8/E66.9; E66.3 overweight excluded); BMI ≥95th percentile for age status code (Z68.54, pediatric obese range; Z68.51-Z68.53 underweight/normal/ overweight excluded); Any offspring obesity diagnosis (composite: E66 obesity diagnosis or Z68.54 BMI ≥95th percentile status code)
Offspring, long run	Other offspring diagnoses (type 2 diabetes, hypertension, obesity)	Same horizons as atopic disease above; estimated at every horizon but statistically underpowered	Offspring type 2 diabetes diagnosis (E11.x); Offspring hypertension diagnosis (I10-I13); Obesity diagnosis (E66.0/E66.01/E66.09/E66.1/E66.2/ E66.8/E66.9; E66.3 overweight excluded); BMI ≥95th percentile for age status code (Z68.54, pediatric obese range; Z68.51-Z68.53 underweight/normal/overweight excluded); Any offspring obesity diagnosis (composite: E66 obesity diagnosis or Z68.54 BMI ≥95th percentile status code)

Notes: One row per outcome family rather than per outcome; families measured over several windows or horizons (DSMT/MNT event-time windows, fetal surveillance, later maternal and offspring disease) list all windows once. The pregnancy window ends at delivery when a delivery is observed and otherwise at the censoring date defined in Section 3. DSMT: diabetes self-management training; MNT: medical nutrition therapy; NICU: neonatal intensive care unit; CGM: continuous glucose monitor; CPT: Current Procedural Terminology; HCPCS: Healthcare Common Procedure Coding System. Diagnosis-code outcomes are claims-based proxies; see Section 3 for measurement caveats. Author's calculations from MarketScan claims data.

The timing comparisons in Section 5.2 (for example, the DSMT and MNT event-time windows) concern recorded claims rather than fully observed service use, and two features of these definitions limit how a smaller or null estimate in a later window should be read. First, the windows are fixed calendar-day offsets from the test date and are not truncated at delivery, so a window such as 180–365 days can fall largely or entirely in the postpartum period, when DSMT and MNT claims are not expected regardless of GDM status; a null estimate there may simply reflect that neither group is receiving this pregnancy-specific service any longer, not that the gap between groups has closed. Second, the indicators are coded as zero, not missing, whenever no qualifying claim is observed, including when a woman's enrollment has ended or the window extends beyond the available data. A woman who is no longer observed is therefore indistinguishable, in this coding, from a woman who is observed and has no claim. For both reasons, the smaller estimates in later periods do not establish that actual service use converges once care is no longer observed.

Appendix H: The Diagnostic Rule

As discussed in Section 4, crossing the fasting glucose threshold does not by itself determine GDM diagnosis, even when clinicians strictly apply the Carpenter and Coustan criteria. The conventional rule requires at least two abnormal OGTT measurements. To illustrate how this rule relates to recorded diagnosis, I examine diagnosis rates by the number of measurements that meet or exceed their respective thresholds.

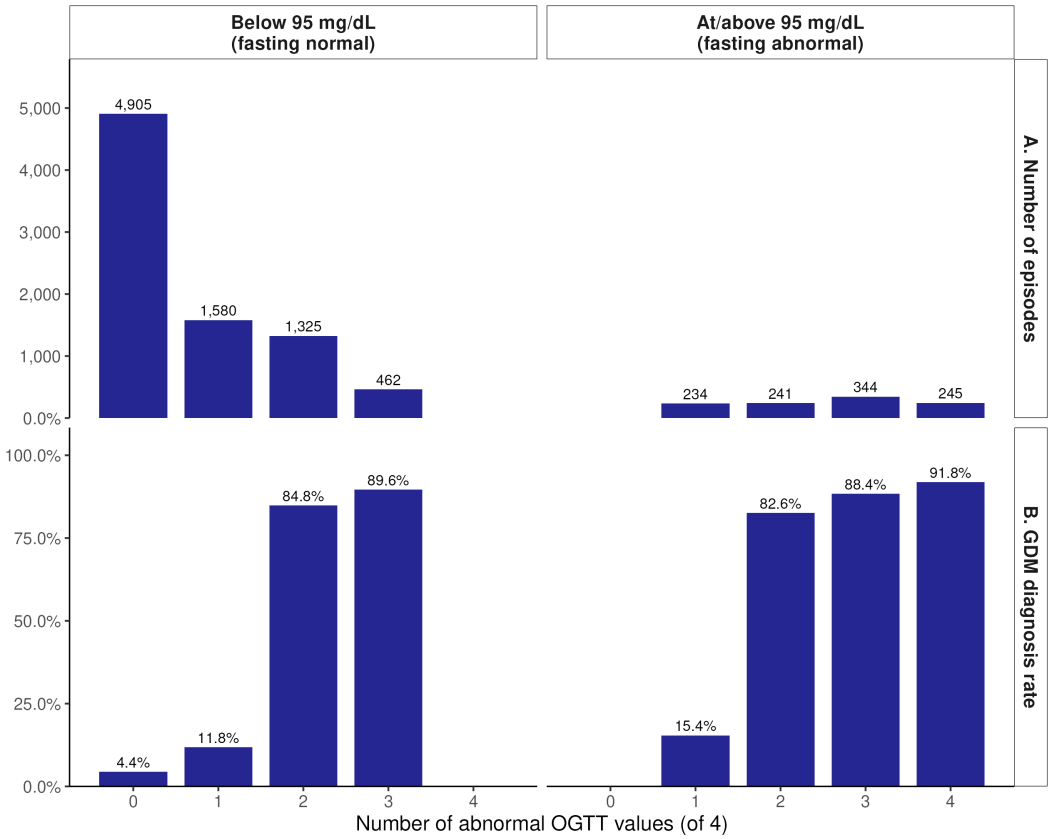
The subset with all four OGTT measurements provides evidence on why fasting glucose changes diagnosis probability without determining it. All four measurements, the fasting value and the 1-hour, 2-hour, and 3-hour results of the 100-gram test recorded in the same testing episode, are available for 9,336 fasting glucose testing episodes, or 10.9 percent of the analytic cohort. In this subset, diagnosis rates rise with the number of abnormal measurements on either side of the fasting cutoff (Table H.1; Figure H.1). These comparisons are descriptive. Table H.2 instead estimates the diagnosis discontinuity within groups defined by the other three measurements. When exactly one of those measurements is abnormal, crossing the fasting threshold increases diagnosis probability by 63.1 percentage points (SE 5.6; $p < 0.001$; $N = 753$). The corresponding estimates are 5.2 points when none is abnormal ($p = 0.23$) and 5.3 points when at least two are abnormal ($p = 0.16$). The largest response therefore occurs where the fasting result is decisive under the conventional rule described in Section 2. This analysis pertains to testing episodes with a complete observed panel. In 80.9 percent of testing episodes, none of the other three measurements is available; an unavailable measurement is not classified as normal.

Table H.1: GDM diagnosis rate by number of abnormal OGTT component values

Abnormal component values	Fasting glucose <95 mg/dL		Fasting glucose ≥95 mg/dL	
	N	Diagnosed	*N*	Diagnosed
0 of 4 abnormal	4,905	4.4%	0	0.0%
1 of 4 abnormal	1,580	11.8%	234	15.4%
2 of 4 abnormal	1,325	84.8%	241	82.6%
3 of 4 abnormal	462	89.6%	344	88.4%
4 of 4 abnormal	0	0.0%	245	91.8%
Fasting value only (no other component observed)	60,619	17.3%	8,642	42.0%

Notes: Rows "0 of 4" through "4 of 4" are fasting glucose testing episodes with all four 100-g OGTT component values observed (fasting, 1-hour, 2-hour, 3-hour), classified by how many exceed the Carpenter–Coustan thresholds (95, 180, 155, and 140 mg/dL); a GDM diagnosis requires at least two. Because fasting glucose ≥95 mg/dL is itself abnormal, no testing episode can have zero abnormal values above the cutoff, and no testing episode can have four below it; those two cells therefore show zero by construction. The last row covers the 80.9 percent of testing episodes in which only the fasting value is observed; their diagnosis rate below the cutoff, about 4.0 times the 0-of-4 rate, shows that an unobserved component cannot be treated as normal. The 7,014 testing episodes (8.2 percent) with some but not all of the other three components observed are not shown. Diagnosed is the share with a GDM diagnosis code. Full sample, N = 85,611. Author's calculations from MarketScan claims data.

Figure H.1: GDM diagnosis by count of abnormal OGTT components



Notes: Subset with a complete OGTT panel only (all four 100-gram OGTT components observed, $N = 9{,}336$ of the full sample of 85,611 fasting glucose testing episodes). Panel A: number of testing episodes at each total count of abnormal values (0-4, fasting included). Panel B: share with a recorded GDM diagnosis. The two sides of the cutoff are not mirror images because the fasting value itself contributes to the count: below 95 mg/dL the count ranges 0-3 (fasting is normal), at or above it the count ranges 1-4 (fasting is abnormal). Author's calculations from MarketScan claims data.

Table H.2: GDM diagnosis at the 95 mg/dL cutoff within strata defined by the other Carpenter-Coustan values

Stratum	Local linear, ±10 mg/dL						Optimal bandwidth		
	Mean below cutoff	Jump (pp)	SE	95% CI	*p*-value	*N*	Estimate (SE)	*p*-value	*h*
No other value abnormal	5.4%	5.21	4.34	[−3.29, 13.72]	0.230	1,240	5.71 (4.44)	0.199	11.5
Exactly one other value abnormal	13.2%	63.10	5.57	[52.18, 74.02]	<0.001	753	65.87 (6.00)	<0.001	10.6
Two or more other values abnormal	88.5%	5.30	3.80	[−2.15, 12.75]	0.164	1,055	5.07 (4.42)	0.252	9.5

Notes: Each row is a separate regression discontinuity estimate of the effect of crossing the 95 mg/dL fasting-glucose cutoff on the probability of any GDM diagnosis code from the test through the later of the first 30 days and delivery, or through the end of the pregnancy window when no delivery is observed, estimated within one stratum of the complete-panel sample: the 9,336 pregnancies in which all four Carpenter-Coustan values (fasting, 1-hour, 2-hour, and 3-hour) are observed. The strata are defined by how many of the three other values are at or above their Carpenter-Coustan thresholds (1-hour 180, 2-hour 155, and 3-hour 140 mg/dL): none, exactly one, or two or more. The Carpenter-Coustan criteria require at least two of the four values to be abnormal, so a fasting value at or above 95 mg/dL is decisive for the diagnosis only where exactly one other value is abnormal; with no other value abnormal it cannot complete a diagnosis, and with two or more the woman already qualifies. These are regression discontinuity estimates within strata, whereas the component counts in Table H.1 are descriptive. The strata are defined by test values observed after the fasting value was drawn, so they describe the mechanism and are not a population defined before the test. The jump is the change in the diagnosis probability at the cutoff, in percentage points (pp), from a uniform-kernel local-linear regression within ±10 mg/dL; mean below cutoff is the diagnosis rate, in percent, among women in the stratum within 10 mg/dL below the cutoff; N is the number of pregnancies within the bandwidth. The optimal bandwidth columns report a separate estimate that uses a triangular kernel and a mean-squared-error-optimal bandwidth h (in mg/dL) selected within the stratum (Calonico, Cattaneo and Titiunik, 2014); the estimate shown is the conventional estimate. Confidence intervals are the estimate plus or minus 1.96 standard errors, clustered by woman. Author's calculations from MarketScan claims data.

Appendix I: Identification-Robustness Detail

The donut-hole and placebo-cutoff checks in this appendix are estimated on the headline diagnosis indicator (Section 4), which is defined for every fasting glucose testing episode. The same checks on the two enrollment-conditional components of the indicator – a GDM diagnosis code within a fixed 30-day window after the lab, and a GDM diagnosis code within the pregnancy window (ending at delivery or at 294 days), each defined only for testing episodes whose enrollment covers the full window – show the same pattern and are summarized in Section 5.1. Table I.1 reports the full donut-hole sweep summarized in Section 5.1: local-linear and optimal-bandwidth estimates of the GDM diagnosis first stage at donut radii 0 through 5 mg/dL, with a ±20 mg/dL reference bandwidth. Table I.2 reports the placebo-cutoff grid at offsets of ±15 and ±25 mg/dL, with a ±10 mg/dL bandwidth.

Table 4 reports the first-stage estimates at each bandwidth for the headline indicator.

Table I.1: First stage: donut-hole estimates

Donut radius	Local linear, ±20 mg/dL	Optimal bandwidth	N
0 mg/dL	0.1478 (p<0.001)	0.0690 (p<0.001)	64,548
1 mg/dL	0.1741 (p<0.001)	0.1203 (p<0.001)	62,478
2 mg/dL	0.1845 (p<0.001)	0.1125 (p=0.008)	60,274
3 mg/dL	0.1924 (p<0.001)	0.1038 (p=0.238)	58,003
5 mg/dL	0.2394 (p<0.001)	0.1914 (p=0.415)	53,375

Notes: The outcome is the headline indicator of Table 4: any GDM diagnosis code from the test through the later of the first 30 days and delivery, or through the end of the pregnancy window when no delivery is observed, defined for every fasting glucose testing episode. Each row excludes testing episodes with fasting glucose within the stated radius (mg/dL) of 95 mg/dL; radius 0 excludes only testing episodes at exactly 95 mg/dL. The local-linear estimate is significant at 5 of 5 radii; it moves from 0.1478 at radius 0 to 0.2394 at radius 5. Excluding a wider interval changes the observations used to fit the local lines and increases the distance over which they are extrapolated, so this pattern does not by itself rule out heaping or sorting. The optimal-bandwidth estimate is significant at 3 of 5 radii (not significant at radii 3, 5 mg/dL) and positive at 5 of 5. Local-linear estimates use a ±20 mg/dL bandwidth, rather than the ±10 mg/dL used elsewhere, so that the sample remains adequate at a 5 mg/dL radius, following Barreca et al. (2011). Local-linear estimates weight observations equally within the bandwidth; optimal-bandwidth estimates use a triangular kernel and a mean-squared-error-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014); these are the conventional estimates at that bandwidth, without bias correction. p-values, in parentheses, are based on standard errors clustered by woman. N is the local-linear estimation sample. Author's calculations from MarketScan claims data.

Table I.2: First stage: placebo cutoffs

Cutoff offset (mg/dL)	Local linear, ±10 mg/dL	Optimal bandwidth	N
-25	-0.0149 (p=0.124)	-0.0260 (p=0.080)	40,453
-15	0.0044 (p=0.427)	-0.0057 (p=0.442)	64,042
0 (true cutoff)	0.1163 (p<0.001)	0.1074 (p<0.001)	26,937
15	-0.0329 (p=0.316)	-0.0303 (p=0.416)	4,471
25	0.0126 (p=0.802)	0.0232 (p=0.655)	1,775

Notes: The outcome is the headline indicator of Table 4: any GDM diagnosis code from the test through the later of the first 30 days and delivery, or through the end of the pregnancy window when no delivery is observed, defined for every fasting glucose testing episode. Each row estimates the discontinuity at a cutoff shifted by the stated offset from 95 mg/dL; offset 0 is the true cutoff, and its local-linear estimate reproduces the ±10 mg/dL estimate in Table 4. The estimate at the true cutoff is positive and significant in 3 of 3 specifications. Of the 8 placebo-cutoff estimates shown in the table (local linear at ±10 mg/dL and optimal bandwidth, at 4 non-zero offsets), none is significant at the 5% level. Across the full grid of 12 placebo-cutoff estimates (local-linear bandwidths of ±5 and ±10 mg/dL and the optimal-bandwidth estimator), none is significant at the 5% level. Local-linear estimates weight observations equally within the bandwidth; optimal-bandwidth estimates use a triangular kernel and a mean-squared-error-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014); these are the conventional estimates at that bandwidth, without bias correction. p-values, in parentheses, are based on standard errors clustered by woman. N is the local-linear estimation sample. Author's calculations from MarketScan claims data.

Table I.3: Manipulation test for a whole-number running variable at each cutoff

Cutoff	Episodes at $c-1$ / c / $c+1$	Share at c	p, $K=0$	p, $K=0.01$	p, $K=0.02$	p, $K=0.05$
Fasting glucose, 95 mg/dL	1,147 / 1,022 / 923	0.331	0.746	0.755	0.779	0.903
50-gram screen, 130 mg/dL	819 / 800 / 802	0.330	0.779	0.785	0.802	0.895
50-gram screen, 135 mg/dL	774 / 714 / 698	0.327	0.511	0.522	0.553	0.725
50-gram screen, 140 mg/dL	616 / 608 / 586	0.336	0.822	0.826	0.837	0.882

Notes: Each row tests whether the number of observations at the whole-number cutoff value c is consistent with a locally smooth distribution, using only c and its two neighbors (Frandsen, 2017). Conditional on the total in those three values, the count at c is binomial with success probability between $(1-K)/(3-K)$ and $(1+K)/(3+K)$ if the relative curvature of the distribution there, $|f(c-1)+f(c+1)-2f(c)|/(f(c-1)+f(c+1))$, is at most K; $K=0$ is a locally linear distribution, for which the share at c is one third. p-values are computed from the exact binomial distribution. The test has power only against a shift of mass between adjacent values at the cutoff. Applying the same test to the 18 whole-number values from 85 to 105 mg/dL that do not touch the cutoff cell, the null is rejected at the 5% level in 0 at $K=0$ and in 0 at $K=0.01$; the placebo windows overlap, so these counts are correlated.
Observations are pregnancy episodes, so a woman with more than one episode contributes more than once; the test treats them as independent. Author's calculations from MarketScan claims data.

Table I.4 repeats the test at the 18 other whole-number glucose values from 85 to 105 mg/dL that do not touch the cutoff cell. None rejects at the 5 percent level at any value of K. Neighboring values share observations, so the 18 tests are correlated.

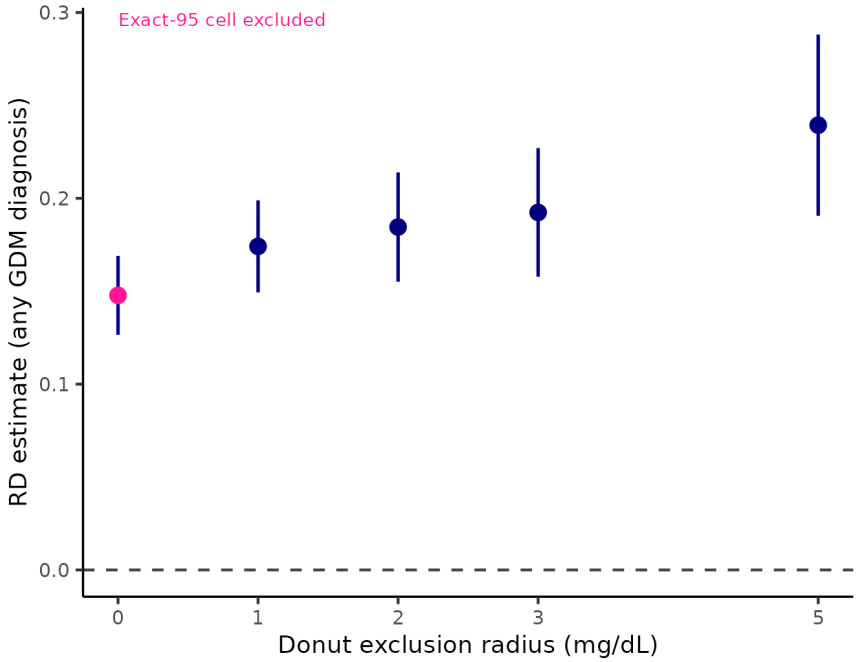
Table I.4: Manipulation test at whole-number glucose values away from the fasting cutoff

Value c (mg/dL)	Episodes at $c-1$ / c / $c+1$	Share at c	p, $K=0$	p, $K=0.01$	p, $K=0.02$	p, $K=0.05$
95 (cutoff)	1,147 / 1,022 / 923	0.331	0.746	0.755	0.779	0.903
85	3,330 / 3,013 / 2,720	0.332	0.867	0.880	0.912	0.992
86	3,013 / 2,720 / 2,551	0.328	0.339	0.383	0.499	0.894
87	2,720 / 2,551 / 2,425	0.331	0.735	0.756	0.811	0.972
88	2,551 / 2,425 / 2,140	0.341	0.183	0.221	0.315	0.736
89	2,425 / 2,140 / 1,833	0.334	0.853	0.863	0.882	0.969
90	2,140 / 1,833 / 1,647	0.326	0.258	0.288	0.371	0.753
91	1,833 / 1,647 / 1,581	0.325	0.239	0.261	0.339	0.705
92	1,647 / 1,581 / 1,357	0.345	0.100	0.119	0.170	0.474
93	1,581 / 1,357 / 1,147	0.332	0.881	0.887	0.902	0.971
97	923 / 847 / 690	0.344	0.248	0.270	0.303	0.515
98	847 / 690 / 634	0.318	0.127	0.137	0.166	0.351
99	690 / 634 / 514	0.345	0.299	0.310	0.339	0.503
100	634 / 514 / 460	0.320	0.255	0.255	0.282	0.448
101	514 / 460 / 413	0.332	0.909	0.911	0.915	0.954
102	460 / 413 / 364	0.334	0.976	0.976	0.977	0.983
103	413 / 364 / 354	0.322	0.430	0.436	0.453	0.563
104	364 / 354 / 302	0.347	0.353	0.375	0.379	0.482
105	354 / 302 / 219	0.345	0.473	0.478	0.492	0.560

Notes: Each row repeats the test in Table I.3 with the whole-number value c shown in the first column as the cutoff cell, using only c and its two neighbors (Frandsen, 2017); the values 94 and 96, which neighbor 95, and 95 itself are excluded, leaving 18 values. The first row is the test at 95 mg/dL. The null is rejected at the 5% level in 0 of 18 values at $K=0$ and in 0 of 18 at $K=0.01$. The smallest p-value is 0.100 at $K=0$ and 0.119 at $K=0.01$; the median is 0.346 and 0.379. Neighboring values share observations, so the tests are correlated.

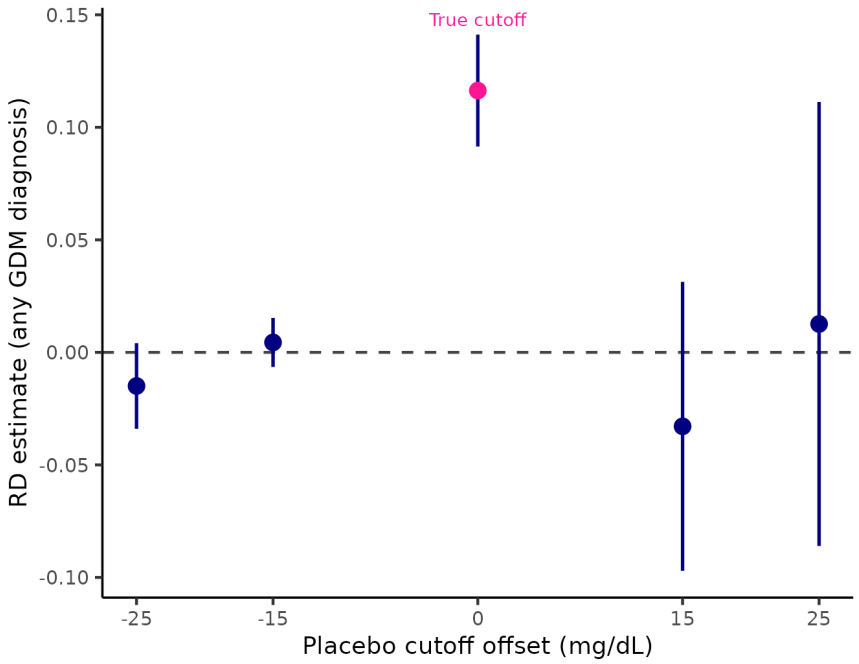
Observations are pregnancy episodes, so a woman with more than one episode contributes more than once; the test treats them as independent. Author's calculations from MarketScan claims data.

Figure I.1: Diagnosis first stage under donut-hole exclusion, by radius



Notes: Local-linear reduced-form estimate (point) and 95% confidence interval (bar) for the discontinuity in the probability of GDM diagnosis, by donut exclusion radius, ±20 mg/dL reference bandwidth, for the headline diagnosis indicator; radius 0 excludes only fasting glucose testing episodes at exactly 95 mg/dL. The estimate does not decay toward zero as observations nearest the cutoff are excluded; the points match the local-linear column of Table I.1. Author's calculations from MarketScan claims data.

Figure I.2: Diagnosis first stage at placebo cutoffs, by offset



Notes: Local-linear reduced-form estimate (point) and 95% confidence interval (bar) for the discontinuity in the probability of GDM diagnosis at placebo cutoffs away from the true 95 mg/dL threshold, ±10 mg/dL bandwidth, for the headline diagnosis indicator. The estimate at the true cutoff (offset 0) is significant in every specification; none of the 12 placebo estimates for this indicator (four offsets, at ±5 and ±10 mg/dL and under the optimal-bandwidth estimator) is significant at the 5% level (the smallest p-value is 0.066, at −25 mg/dL with a ±5 mg/dL bandwidth). Author's calculations from MarketScan claims data.

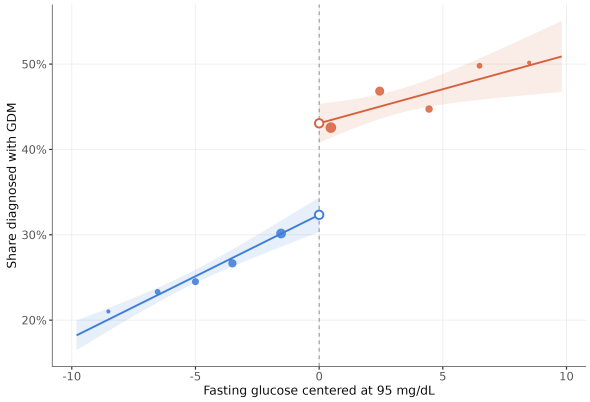
Table I.5 reports the GDM diagnosis first stage estimated separately in the sample used for each of the four outcome families, as summarized in Section 5.1, and Figure I.3 shows the first stage in the three families other than the full cohort, whose first stage is the headline Figure 2. The jump in the pregnancy spending, delivery-observed, and infant-linked samples is similar to the jump in the full cohort.

Table I.5: First stage by sample family

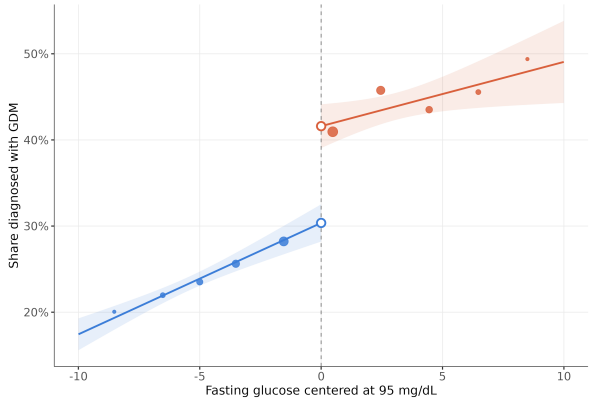
Sample family	±5	±10	±15	±20	±25	±30	Optimal bandwidth
Pregnancy spending	0.0957 (p<0.001)	0.1135 (p<0.001)	0.1270 (p<0.001)	0.1483 (p<0.001)	0.1777 (p<0.001)	0.1960 (p<0.001)	0.1072 (p<0.001)
Delivery management (delivery observed)	0.1031 (p<0.001)	0.1198 (p<0.001)	0.1303 (p<0.001)	0.1528 (p<0.001)	0.1766 (p<0.001)	0.1975 (p<0.001)	0.1123 (p<0.001)
Infant health (infant linked)	0.1032 (p<0.001)	0.1245 (p<0.001)	0.1402 (p<0.001)	0.1678 (p<0.001)	0.2063 (p<0.001)	0.2379 (p<0.001)	0.1178 (p<0.001)

Notes: Discontinuity in the probability of any GDM diagnosis, estimated separately in the sample used for each of the other three outcome families. The full-cohort first stage, used for care utilization, is reported separately. Outcomes with further requirements (pharmacy coverage, delivery payments, follow-up windows) use narrower samples within these families. Columns headed $\pm h$ report local-linear estimates at a bandwidth of h mg/dL; the final column reports the optimal-bandwidth estimate. Local-linear estimates weight observations equally within the bandwidth; optimal-bandwidth estimates use a triangular kernel and a mean-squared-error-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014); these are the conventional estimates at that bandwidth, without bias correction. p-values, in parentheses, are based on standard errors clustered by woman. The estimate is positive and significant at the 5 percent level in all 3 samples at every bandwidth from ±5 to ±30 mg/dL and under the optimal-bandwidth estimator (21 specifications). Author's calculations from MarketScan claims data.

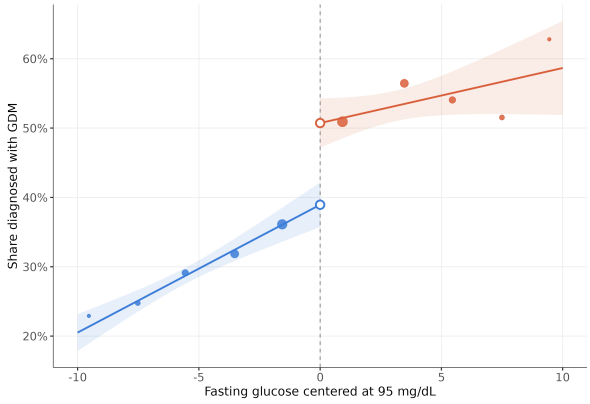
Figure I.3: First-stage GDM diagnosis jump, by sample family



(a) Pregnancy spending



(b) Delivery management (delivery observed)



(c) Infant health (infant linked)

Notes: Points are bin means of the GDM diagnosis rate against the fasting glucose value, centered at 95 mg/dL. Lines are local-polynomial fits with shaded pointwise 95% confidence bands, estimated with the optimal-bandwidth estimator (triangular kernel, each panel's own MSE-optimal bandwidth: (a) ±9.8; (b) ±10.0; (c) ±10.0 mg/dL), drawn in the same way as the full-cohort Figure 2. RD = +0.107, +0.112, and +0.118 (all $p < 0.001$) for panels (a)-(c) respectively, matching each sample's optimal-bandwidth row in Table I.5. The eligible samples, N = 72,685 (panel a), N = 63,143 (panel b), and N = 37,464 (panel c), are larger than the numbers of episodes within ±10 mg/dL (22,897, 19,065, and 10,217, respectively), and each estimate uses the fasting glucose testing episodes inside its own bandwidth. Author's calculations from MarketScan claims data.

Table I.6 reports the full covariate-continuity check summarized in Section 5.7: local-linear discontinuity estimates at every bandwidth from ±5 to ±30 mg/dL, plus the optimal-bandwidth estimator, for each observed level of maternal age, lab year, region, plan type, urbanicity, and the health-plan-reporting flag, and for pre-test pregnancy spending, a placebo outcome estimated on the 68,928 fasting glucose testing episodes whose enrollment covers the pre-test window, with fasting glucose as the running variable. The table's last three rows report, at the fixed ±10 mg/dL bandwidth only, the count of significant metropolitan-area codes, a recorded prior 50-gram screen, and the days from that screen to the test. Figure I.4 shows six representative panels, each plotted at its own MSE-optimal bandwidth. For categorical covariates the panel shows the modal level, a selection rule fixed before any estimate was examined.

Table I.6: Covariate continuity at the 95 mg/dL cutoff

Covariate: level	±5	±10	±15	±20	±25	±30	Optimal
Region: Missing	0.0009 (0.366)	-0.0007 (0.402)	-0.0002 (0.792)	-0.0002 (0.727)	-0.0001 (0.812)	-0.0004 (0.550)	-0.0003 (0.645)
Region: Northeast	0.0002 (0.990)	-0.0068 (0.568)	-0.0042 (0.669)	-0.0069 (0.440)	-0.0128 (0.125)	-0.0140 (0.077)	-0.0055 (0.630)
Region: North Central	-0.0024 (0.825)	0.0008 (0.919)	0.0029 (0.655)	0.0022 (0.712)	0.0010 (0.851)	0.0026 (0.623)	0.0012 (0.862)
Region: South	-0.0120 (0.521)	-0.0002 (0.989)	0.0003 (0.975)	0.0013 (0.894)	0.0074 (0.422)	0.0079 (0.370)	0.0012 (0.925)
Region: West	0.0134 (0.386)	0.0066 (0.546)	0.0008 (0.931)	0.0033 (0.690)	0.0042 (0.580)	0.0037 (0.615)	0.0029 (0.769)
Region: Unknown	-0.0002 (0.810)	0.0002 (0.680)	0.0004 (0.414)	0.0003 (0.432)	0.0003 (0.460)	0.0002 (0.493)	0.0002 (0.816)
Plan type: Missing	0.0004 (0.780)	-0.0008 (0.432)	-0.0007 (0.483)	-0.0007 (0.406)	-0.0002 (0.803)	-0.0006 (0.420)	-0.0006 (0.503)
Plan type: Comprehensive	-0.0000 (0.979)	-0.0006 (0.548)	0.0002 (0.839)	0.0001 (0.885)	-0.0001 (0.843)	-0.0003 (0.722)	-0.0003 (0.710)
Plan type: EPO	-0.0004 (0.843)	-0.0023 (0.061)	-0.0020 (0.068)	-0.0011 (0.214)	-0.0008 (0.364)	-0.0008 (0.323)	-0.0017 (0.170)
Plan type: HMO	0.0056 (0.729)	0.0101 (0.369)	0.0057 (0.546)	0.0089 (0.294)	0.0123 (0.122)	0.0122 (0.108)	0.0060 (0.574)
Plan type: POS	-0.0007 (0.971)	0.0082 (0.533)	0.0065 (0.557)	0.0052 (0.601)	-0.0003 (0.978)	0.0018 (0.836)	0.0049 (0.706)
Plan type: PPO	0.0094 (0.524)	-0.0021 (0.841)	-0.0080 (0.364)	-0.0071 (0.365)	-0.0032 (0.660)	-0.0016 (0.820)	0.0112 (0.357)
Plan type: POS with capitation	0.0012 (0.471)	0.0008 (0.472)	-0.0004 (0.681)	0.0001 (0.945)	-0.0001 (0.908)	-0.0003 (0.697)	0.0004 (0.708)
Plan type: CDHP	-0.0163 (0.093)	-0.0069 (0.304)	-0.0055 (0.328)	-0.0095 (0.057)	-0.0108 (0.022)	-0.0123 (0.006)	-0.0073 (0.224)
Plan type: HDHP	0.0008 (0.928)	-0.0065 (0.272)	0.0041 (0.406)	0.0042 (0.337)	0.0032 (0.441)	0.0019 (0.629)	-0.0044 (0.500)
Urbanicity: Metro	-0.0038 (0.720)	-0.0105 (0.159)	-0.0084 (0.172)	-0.0075 (0.173)	-0.0045 (0.383)	-0.0036 (0.476)	-0.0118 (0.095)
Urbanicity: Missing	0.0003 (0.977)	0.0063 (0.329)	0.0047 (0.371)	0.0059 (0.219)	0.0042 (0.349)	0.0038 (0.375)	0.0068 (0.259)
Urbanicity: Non-metro	0.0035 (0.532)	0.0042 (0.293)	0.0037 (0.275)	0.0017 (0.575)	0.0003 (0.902)	-0.0003 (0.926)	0.0049 (0.215)
Health-plan-reporting flag: record supplied by health plan	0.0034 (0.834)	0.0122 (0.283)	0.0011 (0.908)	0.0057 (0.506)	0.0075 (0.351)	0.0115 (0.135)	0.0046 (0.657)
Health-plan-reporting flag: missing	0.0009 (0.366)	-0.0007 (0.402)	-0.0002 (0.792)	-0.0002 (0.727)	-0.0001 (0.812)	-0.0004 (0.550)	-0.0003 (0.645)
Maternal age (years)	0.046 (0.823)	-0.010 (0.945)	0.011 (0.926)	-0.011 (0.919)	-0.026 (0.799)	-0.034 (0.731)	0.025 (0.867)
Calendar year of glucose test	0.020 (0.861)	0.011 (0.887)	-0.009 (0.897)	0.020 (0.741)	-0.031 (0.584)	-0.018 (0.744)	-0.001 (0.994)

continued on next page

(continued from previous page)

Covariate: level	±5	±10	±15	±20	±25	±30	Optimal
Pre-test pregnancy spending (total payment, $; pregnancy start to day before test)	-646.8 (0.169)	-274.7 (0.339)	45.3 (0.854)	85.3 (0.736)	118.7 (0.611)	219.2 (0.323)	-335.6 (0.237)
Additional covariates from the covariate-balance analysis (local linear only)							
Metropolitan-area codes: estimable / significant at 5% / expected by chance							329 / 9 / 16.4
Prior 50-gram screen on record (within 30 days before the test)	–	0.0196 (0.127)	–	–	–	–	–
Days from nearest prior 50-gram screen to test (women with a prior screen)	–	-0.262 (0.330)	–	–	–	–	–

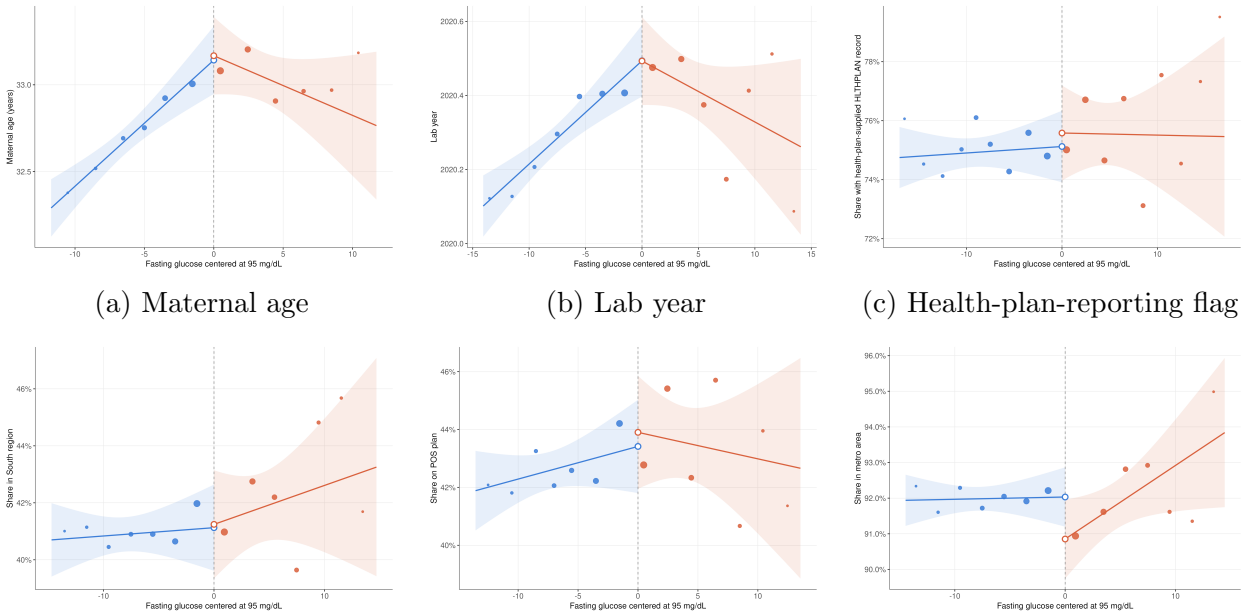
Each row of Table I.6 is a separate regression discontinuity estimate with the covariate, or an indicator for one of its levels, as the outcome. The 23 main-covariate rows cover 7 covariates; because the level indicators for region (6 levels), plan type (9), and urbanicity (3) each sum to one, the rows are not independent tests, and plan type has no row for basic/major medical coverage, which does not occur in this sample. Across these 23 covariates, the 6 bandwidths, and the optimal-bandwidth estimator (161 specifications), 2 are significant at the 5 percent level (8.1 expected by chance under the null) and 8 at the 10 percent level (16.1 expected), both below the number expected by chance. The only covariate significant at the 5 percent level in more than one specification is plan type: CDHP, negative at every bandwidth and significant at ±25 and ±30 mg/dL (and at the 10 percent level at ±5 and ±20 mg/dL as well), but not at the fixed ±10 mg/dL bandwidth or under the optimal-bandwidth estimator. Columns headed $\pm h$ report local-linear estimates at a bandwidth of h mg/dL; the final column reports the optimal-bandwidth estimate. Figures in parentheses are p-values, based on standard errors clustered by woman. Local-linear estimates weight observations equally within the bandwidth; optimal-bandwidth estimates use a triangular kernel and a mean-squared-error-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014) and are the conventional estimates at that bandwidth, without bias correction. The local-linear estimation sample is 26,937 at ±10 mg/dL and varies with bandwidth, and the optimal-bandwidth estimate uses all 85,611 fasting glucose testing episodes, for every row except maternal age (26,907 at ±10 mg/dL; 85,500 under the optimal-bandwidth estimator) and pre-test pregnancy spending, whose sample is the 68,928-testing-episode subsample described above.

Sex is not tested because all fasting glucose testing episodes are from women. The health-plan-reporting flag (Section 3) describes the data source rather than the patient and is omitted from Table 3 for that reason; it is included here because it is the strongest predictor of laboratory-result capture, so its row tests whether the composition of the estimation sample is continuous at the cutoff (selection into that sample is a separate question). Pre-test pregnancy spending is a placebo outcome: spending before the glucose test cannot be caused by crossing the threshold, so a discontinuity would contradict the identifying assumption; it is estimated only on the 68,928 of 85,611 testing episodes continuously enrolled from the start of the pregnancy episode, which is why its estimation sample is smaller, and entry into this subsample is tested separately for continuity at the cutoff (Appendix J).

The table's last three rows come from the covariate-balance analysis of the running variable, which estimates only at the fixed ±10 mg/dL bandwidth; the other bandwidth columns and the optimal-bandwidth estimator are shown as – for these three rows. They use the same local-linear estimator, the same sample, and standard errors clustered by woman. Metropolitan area has too many categories (371 codes) to tabulate individually, so one row reports the number of codes whose indicator could be estimated (329), the number of those with a significant estimate at the 5 percent level (9), and the number expected by chance under the null (0.05 times the number estimable, 16.4); no formal multiplicity-adjusted test is reported for metropolitan area. A prior 50-gram screen is any 50-gram glucose screen recorded in the

30 days before the test; the days row is defined only for women with such a screen, so its estimation sample is smaller.

Figure I.4: Covariate continuity at the 95 mg/dL cutoff



(a) Maternal age

(b) Lab year

(c) Health-plan-reporting flag

(d) Region: South (modal)

(e) Plan type: point-of-service (modal)

(f) Urbanicity: Metro (modal)

Notes: Points are bin means of each covariate against the fasting glucose value, centered at 95 mg/dL. Lines are local-polynomial fits on either side of the threshold with shaded pointwise 95% confidence bands, estimated with the optimal-bandwidth estimator (triangular kernel, each panel's own MSE-optimal bandwidth: (a) ±11.72; (b) ±14.07; (c) ±17.02; (d) ±14.69; (e) ±13.55; (f) ±14.46 mg/dL). None reject continuity (p = 0.867, 0.994, 0.657, 0.925, 0.706, 0.095 for panels (a)-(f) respectively), matching each covariate's optimal-bandwidth row in Table I.6. The estimation samples, N = 85,500 for maternal age and N = 85,611 for the other five panels, are larger than the fixed ±10 mg/dL local-linear samples (26,907 for maternal age, 26,937 for the other five). Panels (a)-(b) are numeric covariates on their own scale; panels (c)-(f) are the share of the sample at the modal (or, for the health-plan-reporting flag, substantively relevant) level of a categorical covariate. Table I.6 reports every level tested, including the health-plan-reporting flag discussed in Section 3. Author's calculations from MarketScan claims data.

Appendix J: Sample-Restriction Smoothness Detail

This appendix reports the sample-restriction smoothness tests summarized in Section 5.1.

Several downstream outcome families require more than the sample used for the diagnosis first stage. Some need an observed delivery or a definable pregnancy spending window (Section 4). Others are defined only within the delivery-observed subsample: the infant outcomes need a linked infant claim, and the delivery-payment outcomes need an inpatient admission record dated inside the delivery window, the only claim type that carries an admission-level payment (Section 5.5). These restrictions differ from calendar-time attrition in the sense

of Section V.C of Alalouf, Miller and Wherry (2024), where sample sizes decline because patients leave the insurer over the follow-up years; they determine whether an outcome can be *defined* at all for a given fasting glucose testing episode. I test whether the probability of entering each outcome-specific subsample changes at 95 mg/dL. A discontinuity raises a selection concern. A statistically insignificant estimate does not establish that sample composition is comparable on the two sides of the cutoff.

A fifth restriction is of a different kind. Pre-test spending, the total payment on claims from the start of the pregnancy episode through the day before the glucose test, serves as a placebo covariate (Table I.6). It is defined only for the 68,928 of the 85,611 fasting glucose testing episodes that are continuously enrolled over that pre-test window. Enrollment over this window precedes the test, so a diagnosis is unlikely to affect it. A discontinuity in entry into this subsample would therefore more likely point to imbalance or to a poorly fitting specification than to selection induced by the diagnosis.

Table J.1: Entry into outcome-specific subsamples at the 95 mg/dL cutoff

Restriction	±5	±10	±15	±20	±25	±30	Optimal bandwidth
Delivery observed within 294 days of test	-0.0188 (p=0.279)	-0.0097 (p=0.424)	-0.0115 (p=0.262)	-0.0118 (p=0.195)	-0.0190 (p=0.025)	-0.0250 (p=0.002)	-0.0100 (p=0.430)
Pregnancy spending window observable	-0.0024 (p=0.856)	-0.0063 (p=0.495)	0.0028 (p=0.720)	0.0031 (p=0.657)	0.0059 (p=0.367)	0.0048 (p=0.446)	-0.0084 (p=0.406)
Pre-test spending window observable	0.0026 (p=0.874)	0.0114 (p=0.323)	0.0245 (p=0.011)	0.0263 (p=0.002)	0.0161 (p=0.044)	0.0085 (p=0.265)	0.0123 (p=0.353)
Delivery admission payment observed	0.0281 (p=0.207)	0.0116 (p=0.459)	0.0145 (p=0.268)	0.0264 (p=0.024)	0.0233 (p=0.033)	0.0157 (p=0.134)	0.0154 (p=0.338)
Infant record linked (among observed deliveries)	0.0364 (p=0.108)	0.0362 (p=0.022)	0.0301 (p=0.024)	0.0145 (p=0.222)	0.0049 (p=0.656)	-0.0036 (p=0.735)	0.0353 (p=0.033)

Notes: Each row is a regression discontinuity estimate in which the outcome is an indicator for entering the stated subsample. Columns headed $\pm h$ report local-linear estimates at a bandwidth of h mg/dL; the final column reports the optimal-bandwidth estimate. Local-linear estimates weight observations equally within the bandwidth; optimal-bandwidth estimates use a triangular kernel and a mean-squared-error-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014); these are the conventional estimates at that bandwidth, without bias correction. p-values, in parentheses, are based on standard errors clustered by woman. Estimation samples: delivery observation, spending-window observability, and pre-test spending observability (the full cohort, $N = 85{,}611$); delivery-payment observability and infant linkage (the delivery-observed subsample, $N = 63{,}143$). Infant linkage is defined only when a delivery is observed. Estimation N varies by bandwidth and is not shown. Because the spending window ends at delivery for most women, the delivery and spending-window indicators are not independent. Delivery observation is significant at ±25 and ±30 mg/dL (negative); it is not significant at ±5, ±10, ±15, and ±20 mg/dL and under the optimal-bandwidth estimator. Spending-window observability shows no significant discontinuity at any bandwidth or under the optimal-bandwidth estimator. Pre-test spending observability is significant at ±15, ±20, and ±25 mg/dL (positive); it is not significant at ±5, ±10, and ±30 mg/dL and under the optimal-bandwidth estimator. Delivery-payment observability is significant at ±20 and ±25 mg/dL (positive); it is not significant at ±5, ±10, ±15, and ±30 mg/dL and under the optimal-bandwidth estimator. Infant linkage is significant at ±10 and ±15 mg/dL (positive) and under the optimal-bandwidth estimator (positive); it is not significant at ±5, ±20, ±25, and ±30 mg/dL. Appendix J discusses the implications, including for the infant outcomes, which are estimated at ±10 and ±15 mg/dL and with the optimal-bandwidth estimator. Author's calculations from MarketScan claims data.

The five restrictions yield different results (Table J.1). I find no statistically significant discontinuity in spending window observability at any of the reported bandwidths (p from 0.37 to 0.86). Delivery observation shows no discontinuity at the fixed ±10 mg/dL bandwidth or under the optimal-bandwidth estimator ($p = 0.424/0.430$) and remains null through ±20 mg/dL, but the estimate turns negative and significant at ±25 mg/dL (−1.9 points, $p = 0.025$) and ±30 mg/dL (−2.5 points, $p = 0.002$). The base rate of delivery observation is roughly 74%, and the significant estimates appear only at bandwidths wider than 20 mg/dL.

Pre-test spending observability and delivery-payment observability are smooth at the fixed ±10 mg/dL bandwidth and under the optimal-bandwidth estimator, but not at every wider bandwidth. Entry into the pre-test spending subsample shows no discontinuity at ±5 mg/dL (+0.26 points, $p = 0.874$), at the fixed ±10 mg/dL bandwidth (+1.14 points, $p = 0.323$), or under the optimal-bandwidth estimator (+1.23 points, $p = 0.353$). It is positive and significant at ±15 (+2.45 points, $p = 0.011$), ±20 (+2.63 points, $p = 0.002$), and ±25 mg/dL (+1.61 points, $p = 0.044$), and not significant again at ±30 mg/dL ($p = 0.265$). Because pre-test enrollment precedes the test, this wider-bandwidth pattern is more consistent with a steep gradient below the cutoff, which a linear fit captures poorly at wider bandwidths, than with selection induced by the diagnosis. The observable share among fasting glucose testing episodes below the cutoff rises from 74.2 percent within ±5 mg/dL to 81.5 percent within ±30 mg/dL, while the share above the cutoff stays between 74.2 and 74.7 percent. Pre-test spending itself shows no discontinuity at any bandwidth from ±5 to ±30 mg/dL (p from 0.17 to 0.85), including the bandwidths at which entry moves. At the primary bandwidth it is −$275 ($p = 0.339$), and under the optimal-bandwidth estimator −$336 ($p = 0.237$) (Table I.6). Because entry is smooth there, these two estimates are the cleaner version of the placebo test.

Among delivery-observed pregnancies, the delivery payment is observed for 30,906 of 63,143 (48.9 percent). Whether it is observed does not change at the cutoff at the fixed ±10 mg/dL bandwidth (+1.16 points, $p = 0.459$), under the optimal-bandwidth estimator (+1.54 points, $p = 0.338$), or at ±5, ±15, and ±30 mg/dL. It rises by 2.64 points at ±20 mg/dL ($p = 0.024$) and 2.33 points at ±25 mg/dL ($p = 0.033$). The delivery-payment estimates in Section 5.5 use the fixed ±10 mg/dL bandwidth and the optimal-bandwidth estimator, where entry is smooth. At ±20 and ±25 mg/dL, however, the payment estimates in Figure M.2 come from a subsample whose entry moves at the cutoff. Smooth entry at the fixed ±10 mg/dL bandwidth also does not make this subsample representative of all deliveries; how it differs from the other deliveries is described in Section 5.5.

Infant-claim linkage is the only restriction whose entry moves at the fixed ±10 mg/dL bandwidth and under the optimal-bandwidth estimator. Among delivery-observed women, crossing the cutoff raises the probability of a linked infant claim by 3.6 points at the fixed ±10 mg/dL bandwidth ($p = 0.022$), 3.0 points at ±15 ($p = 0.024$), and 3.5 points under the optimal-bandwidth estimator ($p = 0.033$), so the discontinuity is significant at the fixed ±10 mg/dL bandwidth. At ±5 mg/dL the estimate is similar in magnitude (3.64 points) but less precise ($p = 0.108$); it is smaller at ±20 mg/dL and beyond.

To examine this discontinuity, I decompose the linkage outcome into four mutually exclusive categories for each delivery-observed mother: a unique single match; a match resolved by nearest enrollment-start date among several candidates; an unresolved multi-candidate case; and no linkage candidate. Under the optimal-bandwidth estimator, three of the four categories move at the cutoff: the single-match share rises 3.28 points ($p = 0.047$) and the no-candidate share falls 3.47 points ($p = 0.036$), nearly mirror images of each other, while the share resolved by nearest enrollment-start date moves by much less (0.28 points, $p = 0.049$). Only the unresolved multi-candidate category ($p = 0.55$) and the rate of missing family-identifier data (the mechanical prerequisite for any linkage attempt, which is constant in this cohort and therefore cannot be tested) are flat. Splitting the sample by whether delivery falls within 45 days of December 31 does not support the most obvious mechanical explanation, namely that late-in-year deliveries link at different rates by GDM status because enrollment-year processing lags. Within 45 days of the year boundary (12.0 percent of the sample, $n = 7{,}549$), both categories are null under the optimal-bandwidth estimator (single match +0.89 points, $p = 0.761$; no candidate −0.75 points, $p = 0.801$); away from the boundary ($n = 55{,}594$), both are significant and, if anything, larger (single match +3.77 points, $p = 0.035$; no candidate −3.99 points, $p = 0.027$). A calendar-processing artifact would instead concentrate the discontinuities near the boundary. This finding does not identify the cause of the discontinuity. A behavioral response to diagnosis is one possibility, but the data do not distinguish it from other differences in enrollment or record linkage. Selection into the linked infant sample therefore remains a concern, and the direction of any resulting bias in the infant estimates is not established.

The infant outcomes rest on two sequential restrictions that are tested separately. The more fundamental one, whether a pregnancy results in an observed delivery at all, is smooth at the cutoff (Table J.1). A separate check shows that the analogous continuous-enrollment requirements for the mother at the 24–60 month horizons in Section 5.6 are likewise smooth at conventional levels, with one borderline exception at 60 months ($p = 0.050$ for the local-linear estimate, not confirmed by the optimal-bandwidth estimate, $p = 0.105$), the horizon already noted there as too sparse to support a confident reading in either direction. These results do not detect a change in whether a pregnancy enters the delivery-anchored analysis, although they cannot rule one out. Infant-claim linkage is a second, conditional restriction, and it is the one that moves at the cutoff. Outcomes estimated on the full delivery-observed subsample, including the long-run maternal results in Section 5.6, are not subject to this particular restriction. Only outcomes estimated on the infant-linked subsample – the infant-diagnosis results in Section 5.4 and the offspring result in Section 5.6 – carry the caveat that crossing the cutoff predicts inclusion in that subsample, so those estimates may partly reflect selection into observability rather than a pure treatment effect.

Composition of the delivery-payment subsample

Table J.2 compares deliveries with an observed admission payment with the rest of the delivery-observed subsample. As reported above, entry into the payment subsample is smooth at the fixed ±10 mg/dL bandwidth and under the optimal-bandwidth estimator. Its composition nonetheless differs from that of other deliveries. Among deliveries with an observed payment, 82.2 percent carry an obstetric anesthesia claim and 36.7 percent a cesarean-delivery claim; among the others, 3.5 and 1.1 percent do (standardized differences 2.62 and 1.02; within ±10 mg/dL the shares are 83.3 and 42.1 percent against 3.9 and 1.0 percent, as in the table). Over the whole delivery-observed subsample, the payment-observed deliveries are also more often linked to an infant record (72.4 against 46.8 percent) and more often preceded by a 50-gram screening test (51.0 against 34.4 percent). Within ±10 mg/dL of the cutoff, they are more often diagnosed with GDM (34.5 against 22.3 percent) and had lower pre-test spending ($2,425 against $4,407). Differences in region, plan type, urbanicity, the health-plan-reporting flag, and maternal age are small (standardized differences of at most 0.11 in absolute value).

The diagnosis first stage is larger in the payment subsample. At ±10 mg/dL it is 0.161 among deliveries with an observed payment (N = 8,330), against 0.092 among the others (N = 10,735) and 0.120 among all deliveries (N = 19,065). The difference of 0.069 is significant ($p = 0.022$), and it widens at ±20 mg/dL (0.214 against 0.112; difference 0.103, $p < 0.001$). Within the payment subsample, 2 of the 23 covariate rows show a discontinuity at the 5 percent level at the fixed ±10 mg/dL bandwidth, against 0 of 23 among all deliveries and about 1.2 expected by chance. Both are plan types: consumer-directed health plans (−2.57 points, $p = 0.048$) and high-deductible health plans (−2.47 points, $p = 0.028$). Restricting the payment estimate to deliveries whose delivery window is fully enrollment-observable changes it from $554 (SE $1,027, $p = 0.590$, N = 8,330) to −$36 (SE $762, $p = 0.962$, N = 8,254; optimal bandwidth −$271, $p = 0.706$). The delivery-payment estimates therefore provide no evidence of higher payment at the threshold, but they describe a selected set of deliveries and are imprecise.

Table J.2: Delivery-payment observability at the 95 mg/dL cutoff

Panel A. Discontinuity in observability, delivery-observed pregnancies			
	Payment observed	Delivery window within enrollment	*N*
Local linear, ±5 mg/dL	0.0281 (p=0.207)	-0.0051 (p=0.102)	8,513
Local linear, ±10 mg/dL	0.0116 (p=0.459)	-0.0041 (p=0.082)	19,065
Local linear, ±15 mg/dL	0.0145 (p=0.268)	-0.0011 (p=0.613)	33,041
Local linear, ±20 mg/dL	0.0264 (p=0.024)	-0.0014 (p=0.450)	48,019
Local linear, ±25 mg/dL	0.0233 (p=0.033)	-0.0013 (p=0.475)	57,815
Local linear, ±30 mg/dL	0.0157 (p=0.134)	-0.0000 (p=0.986)	61,363
Optimal bandwidth	0.0154 (p=0.338)	-0.0036 (p=0.131)	63,143
Panel B. First stage: discontinuity in any GDM diagnosis			
	±10 mg/dL	±20 mg/dL	*N* (±10 / ±20)
All delivery-observed pregnancies	0.1198 (p<0.001)	0.1528 (p<0.001)	19,065 / 48,019
Delivery payment observed	0.1608 (p<0.001)	0.2145 (p<0.001)	8,330 / 22,871
Delivery payment not observed	0.0921 (p<0.001)	0.1115 (p<0.001)	10,735 / 25,148
Difference, observed minus not observed	0.0687 (p=0.022)	0.1030 (p<0.001)	19,065 / 48,019
Panel C. Characteristics by payment observability, within ±10 mg/dL			
	Payment observed	Payment not observed	Standardized difference
Obstetric anesthesia	83.3%	3.9%	2.67
Cesarean delivery	42.1%	1.0%	1.15
Days from test to end of delivery episode	90.4	55.3	0.64
Infant record linked	69.6%	41.2%	0.60
Prior 50-g screening test	53.1%	27.5%	0.54
Diagnosed with GDM	34.5%	22.3%	0.27
Panel D. Delivery admission total payment ($)			
	Local linear, ±10 mg/dL	Optimal bandwidth	*N*
All deliveries with payment observed	$553.92 (p=0.590)	$284.66 (p=0.784)	8,330
Delivery window within enrollment	-$36.34 (p=0.962)	-$270.94 (p=0.706)	8,254

Notes: Payment information on the delivery admission is available for 30,906 of the 63,143 delivery-observed pregnancies. Panel A: the outcome is an indicator that the delivery admission has payment information (first column) or that the delivery window lies within enrollment (second column), estimated on all delivery-observed pregnancies; the optimal-bandwidth row uses all of them. The discontinuity in payment observability is significant at the 5% level at ±20 and ±25 mg/dL; it is not significant at ±5, ±10, ±15, and ±30 mg/dL or under the optimal-bandwidth estimator. Panel B: the discontinuity in the probability of any GDM diagnosis, estimated separately in each group; the last row is the difference between the payment-observed and payment-unobserved discontinuities from a pooled local-linear regression in which the discontinuity is allowed to differ by payment observability, and its *p*-value tests that the two are equal. Panel C: group means for deliveries with and without payment information and the standardized difference (difference in means divided by the square root of the average of the two group variances). Means are computed within ±10 mg/dL of the cutoff over 8,330 deliveries with and 10,735 without payment information (obstetric anesthesia: 8,254 and 10,685; cesarean delivery: 8,254 and 10,685, the deliveries on which these indicators are defined). Panel D: the discontinuity in total payment (plan plus patient, nominal dollars) on the delivery admission, which reproduces the payment row of Table 12; the second row keeps only deliveries whose delivery window lies within enrollment, which removes 76 of the 8,330 payment-observed deliveries within ±10 mg/dL. Local-linear estimates weight observations equally within the bandwidth; optimal-bandwidth estimates use a triangular kernel and a mean-squared-error-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014); these are the conventional estimates at that bandwidth, without bias correction. *p*-values, in parentheses, are based on standard errors clustered by woman. In Panels A, B, and D, *N* is the local-linear estimation sample, except in the optimal-bandwidth row of Panel A. Author's calculations from MarketScan claims data.

J.1 Selection into the infant-linked sample

Infant outcomes are observed only for pregnancies with a linked infant enrollment record, and crossing the cutoff raises the probability of linkage by 3.6 percentage points (Table J.1). This subsection describes four analyses of the resulting selection concern.

Covariate balance. I repeat the covariate continuity tests (Table I.6) among infant-linked women, among all delivery-observed women, and among women without a linked infant, using the same local linear estimator at ±10 mg/dL with standard errors clustered by woman. For each covariate I also estimate a regression of the covariate on the above-cutoff indicator, the running variable, an indicator for a linked infant, and their interactions; the coefficient on the interaction of the above-cutoff indicator and the linked indicator is the difference between the covariate discontinuity among linked and unlinked women (Table J.3).

Reweighting. I estimate the probability that a pregnancy has a linked infant with a logit, fitted separately on each side of the cutoff within ±30 mg/dL, using maternal age, calendar year and month of the test, region, plan type, urbanicity, the health-plan-reporting flag, pre-test pregnancy spending (with an indicator for missing values), and a cubic in the running variable. Continuous covariates are centered and scaled, and collinear indicators are dropped. Linked infants are weighted by the inverse of their fitted probability, truncated at the 1st and 99th percentiles of the weights (1.09 and 6.29); the model classifies linkage with an area under the receiver operating characteristic curve of 0.76. The weighted local linear regressions use the same bandwidths, with standard errors clustered by woman that treat the weights as known. Reweighting can adjust only for differences in observed characteristics and cannot address selection on unobserved characteristics (Table J.4).

Trimming bounds. Suppose crossing the cutoff never lowers the probability that an infant is linked. Then infants who are linked just below the cutoff would also be linked just above it, and the linked infants above the cutoff include some who would not have been linked below (Lee, 2009). I estimate the linked share at the cutoff on each side as the intercept of a linear regression of the linked indicator on the running variable within the bandwidth, and set the trimming share to $p = (q_{above} - q_{below})/q_{above}$, where q is the linked share; at ±10 mg/dL this is 7.0 percent, or 154 of the 2,220 linked infants above the cutoff. I drop the p share of above-cutoff linked infants with the highest outcome values, and separately the same number with the lowest, and re-estimate the local linear discontinuity on each trimmed sample; the smaller and larger estimates are the lower and upper bounds. For a binary outcome, dropping the highest values removes events first, so the lower bound moves the above-cutoff rate toward zero, and the upper bound removes infants without the event; because these outcomes are rare (1.3 percent above the cutoff), the upper bound moves the estimate little. Ties among infants with equal outcomes are broken by a fixed ordering. The confidence intervals come from 500 bootstrap samples that resample women, so that all of a woman's pregnancies

enter or leave together, with the trimming share, both trimmed estimates and the bounds recomputed in every draw; the interval is the Imbens-Manski interval for a partially identified parameter (Imbens and Manski, 2004), computed from the bootstrap standard errors of the two bounds. The bounds are approximate because the trimming share is estimated at the cutoff and applied throughout the bandwidth window. They concern the reduced-form effect of crossing the cutoff for infants linked on either side; they are not bounds on the effect of a diagnosis and do not validate the per-complier estimates directly (Table J.5).

Claims visibility. An outcome code can be recorded only for an infant with a claim in the window from one day before to 30 days after birth. Among linked infants, the share with any claim is 32.2 percent below and 27.3 percent above the cutoff (-0.6 points, $p = 0.76$). Restricting to linked infants with a claim raises the outcome rates, so the estimate in levels is expected to be about $1/s$ times the unconditional one, where s is the claim share (about 3.1); I therefore compare relative effects, the estimate divided by the below-cutoff mean (Table J.6).

Table J.3: Balance of pre-test characteristics among women with a linked infant record

Sample	Covariates tested	Significant at 5%	Significant at 10%	Expected at 5% by chance	Expected at 10% by chance
Panel A. Number of pre-test characteristics with a significant jump at the cutoff, ±10 mg/dL					
Women with a linked infant	23	0	0	1.15	2.30
All women with an observed delivery	23	0	0	1.15	2.30
Women without a linked infant	23	0	3	1.15	2.30
Panel B. Jump at the cutoff in the probability that a pregnancy has a linked infant					
Bandwidth	Jump (pp)	SE (pp)	*p*-value	*N*	
±5 mg/dL	3.64	2.26	0.108	8,513	
±10 mg/dL	3.62	1.58	0.022	19,065	
±15 mg/dL	3.01	1.33	0.024	33,041	
±20 mg/dL	1.45	1.19	0.222	48,019	
Panel C. Difference in covariate jump, linked minus unlinked: 1 of 23 significant at 5%					
Covariate	Difference	SE	*p*-value	*N*	
Pre-test pregnancy spending	1,283.26	551.50	0.020	13,528	

Notes: Infant outcomes can be studied only for pregnancies with a linked infant record, and crossing the 95 mg/dL cutoff raises the probability of a linked record (Panel B), so the linked women just above the cutoff could differ from those just below. Panel A repeats the covariate-continuity test for the 23 pre-test characteristics (maternal age, calendar year of the glucose test, region, plan type, urbanicity, the health-plan reporting flag, and pre-test pregnancy spending; categories of the same characteristic are separate rows) in each sample. Each characteristic is a separate regression discontinuity estimate from a uniform-kernel local-linear regression within ±10 mg/dL of the cutoff, with standard errors clustered by woman; the columns count the characteristics with a jump significant at the 5% and 10% level. The expected counts are 0.05 and 0.10 times the number tested, the number of significant results if no characteristic jumped; the tests are not independent, because categories of one characteristic are mechanically related. Panel B reports the jump in the probability of a linked infant record among all women with an observed delivery, from the same local-linear regression, in percentage points (pp). Panel C reports, for each characteristic, the difference between linked and unlinked women in the jump at the cutoff, from a regression within ±10 mg/dL of the characteristic on an indicator for being at or above the cutoff, an indicator for a linked infant, their interaction, and all interactions with the running variable, among all women with an observed delivery, with standard errors clustered by woman; the difference is the coefficient on the interaction of the above-cutoff indicator with the linked indicator, in the characteristic's own units (dollars for pre-test spending), and only characteristics with a significant difference at the 5% level are listed. With 23 tests, about 1.15 significant differences are expected by chance. Balance on observed characteristics does not rule out differences in unobserved characteristics. Author's calculations from MarketScan claims data.

Table J.4: Infant outcomes reweighted for the probability of a linked infant record

Weighting	±5 mg/dL	±10 mg/dL	±15 mg/dL	±20 mg/dL	Optimal bandwidth
Macrosomia (P08.0 or P08.1)					
					−1.36 (0.49)
	−1.66 (0.90)	−1.29 (0.56)	−1.37 (0.46)	−1.42 (0.41)	[0.006]
Unweighted	[0.066]	[0.021]	[0.003]	[<0.001]	$h = 16.0$
	−1.58 (0.97)	−1.37 (0.62)	−1.61 (0.53)	−1.57 (0.47)	−1.38 (0.88)
Inverse-probability weighted	[0.103]	[0.026]	[0.002]	[<0.001]	[0.119]
Heavy for gestational age (P08.1)					
					−1.45 (0.44)
	−1.82 (0.89)	−1.51 (0.54)	−1.46 (0.44)	−1.50 (0.39)	[<0.001]
Unweighted	[0.040]	[0.006]	[<0.001]	[<0.001]	$h = 18.3$
	−1.71 (0.96)	−1.47 (0.59)	−1.63 (0.50)	−1.59 (0.46)	−1.53 (0.88)
Inverse-probability weighted	[0.075]	[0.013]	[0.001]	[<0.001]	[0.081]
NICU admission (revenue-code indicator)					
					−1.33 (0.52)
	−1.84 (0.87)	−1.43 (0.57)	−0.99 (0.46)	−1.19 (0.40)	[0.011]
Unweighted	[0.035]	[0.012]	[0.033]	[0.003]	$h = 14.6$
	−1.16 (0.83)	−1.43 (0.58)	−1.08 (0.50)	−1.27 (0.44)	−1.09 (0.68)
Inverse-probability weighted	[0.166]	[0.014]	[0.031]	[0.004]	[0.110]
Pregnancies in the estimation sample (N)	4,373	10,217	18,619	27,980	37,464 / 36,500

Notes: NICU: neonatal intensive care unit. Each cell is a separate regression discontinuity estimate of the effect of crossing the 95 mg/dL cutoff on the outcome among pregnancies with a linked infant record, in percentage points; standard errors, clustered by woman, are in parentheses and p-values are in brackets. The unweighted rows are the estimates in the main infant results. In the inverse-probability-weighted rows each pregnancy is weighted by the inverse of the estimated probability that the mother's infant is linked. The probability comes from a logit fit separately on each side of the cutoff among pregnancies with an observed delivery within ±30 mg/dL of the cutoff, using characteristics measured before the test: age, calendar year and month, region, plan type, urbanicity, the health-plan reporting flag, pre-test pregnancy spending, and a cubic in the running variable. Weights are truncated at the 1st and 99th percentiles of the weights of linked pregnancies within the fitting window, so that they range from 1.09 to 6.29 in the ±10 mg/dL sample; the area under the receiver operating characteristic curve of the linkage model is 0.758, and the effective sample size of the weighted ±10 mg/dL sample is 7,886 of 10,217 linked pregnancies. Reweighting can adjust only for observed characteristics, and the standard errors treat the weights as known, so they are approximate. Local-linear estimates use a uniform kernel within the stated bandwidth of the cutoff. The optimal bandwidth estimate uses a triangular kernel and a mean-squared-error-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014); it is the conventional point estimate at that bandwidth. In that column h is the selected bandwidth in mg/dL (shown for the unweighted estimate only; none is reported for the weighted estimate), and N is the number of pregnancies entering the estimator, unweighted / weighted: the weights are defined only within ±30 mg/dL of the cutoff, so the weighted estimator uses that window. Author's calculations from MarketScan claims data.

Table J.5: Bounds on the infant effects that allow for selection into infant linkage

Outcome and bandwidth	Trim share (%)	Trimmed (n)	Untrimmed estimate (pp)	p-value	Lower bound (pp)	Upper bound (pp)	95% confidence interval (pp)
Panel A. Linked infants							
Macrosomia (P08.0 or P08.1)							
±5 mg/dL	6.89	111	−1.66	0.066	−2.76	−1.59	[−4.10, 0.12]
±10 mg/dL	6.97	154	−1.29	0.021	−2.31	−1.23	[−3.10, −0.25]
±15 mg/dL	5.83	149	−1.37	0.003	−2.38	−1.31	[−2.98, −0.48]
±20 mg/dL	2.82	78	−1.42	<0.001	−2.47	−1.39	[−3.24, −0.37]
Heavy for gestational age (P08.1)							
±5 mg/dL	6.89	111	−1.82	0.040	−2.77	−1.76	[−4.12, −0.05]
±10 mg/dL	6.97	154	−1.51	0.006	−2.31	−1.46	[−3.08, −0.48]
±15 mg/dL	5.83	149	−1.46	<0.001	−2.31	−1.42	[−2.88, −0.61]
±20 mg/dL	2.82	78	−1.50	<0.001	−2.38	−1.48	[−3.07, −0.48]
NICU admission (revenue-code indicator)							
±5 mg/dL	6.89	111	−1.84	0.035	−2.99	−1.77	[−4.41, −0.01]
±10 mg/dL	6.97	154	−1.43	0.012	−2.59	−1.36	[−3.33, −0.39]
±15 mg/dL	5.83	149	−0.99	0.033	−2.15	−0.93	[−2.76, −0.12]
±20 mg/dL	2.82	78	−1.19	0.003	−2.43	−1.16	[−3.29, −0.23]
Panel B. Linked infants, excluding a test value of exactly 95 mg/dL, ±10 mg/dL							
Macrosomia (P08.0 or P08.1)	4.79	88	−1.72	0.004	−2.31	−1.71	[−3.17, −0.38]
Heavy for gestational age (P08.1)	4.79	88	−1.89	0.001	−2.31	−1.89	[−3.16, −0.56]
NICU admission (revenue-code indicator)	4.79	88	−1.19	0.075	−2.59	−1.14	[−3.67, 0.35]

continued on next page

(continued from previous page)

Outcome and bandwidth	Trim share (%)	Trimmed (n)	Untrimmed estimate (pp)	p-value	Lower bound (pp)	Upper bound (pp)	95% confidence interval (pp)
Panel C. Linked infants with any infant claim, ±10 mg/dL							
Macrosomia (P08.0 or P08.1)	2.14	55	−4.48	0.020	−4.70	−1.74	[−8.48, 4.47]
Heavy for gestational age (P08.1)	2.14	55	−5.23	0.005	−5.45	−2.35	[−8.95, 3.85]
NICU admission (revenue-code indicator)	2.14	55	−4.86	0.012	−5.13	−1.07	[−9.11, 5.44]

Notes to Table J.5: NICU is neonatal intensive care unit. Trim share is the share of linked infants removed from the side of the cutoff with the larger linked share (above the cutoff in Panels A and B, below the cutoff in Panel C), and trimmed is the number of infants removed. The untrimmed estimate is the reduced-form effect of crossing the cutoff on the outcome in the untrimmed sample, from a uniform-kernel local linear regression within the stated bandwidth, in percentage points (pp). The lower and upper bounds are the two trimmed estimates, and the 95 percent confidence interval is the Imbens–Manski interval from 500 bootstrap samples of women. Panel B repeats the ±10 mg/dL calculation after dropping linked pregnancies whose test value is exactly 95 mg/dL. Panel C applies the same procedure to claims visibility: the sample is linked infants with any infant claim in the window from one day before through 30 days after delivery, and the selection share is the share with a claim. Outcome rates in that sample are higher, because a claim is necessary to record any outcome. Author's calculations from MarketScan claims data.

Table J.6: Infant outcomes conditional on any infant claim: absolute and relative effects

Outcome and sample	N	Mean below cutoff (%)	Estimate (SE), pp	p-value	Relative effect (%)	Ratio of estimates
Macrosomia (P08.0 or P08.1)						
All linked infants	10,217	2.3	−1.29 (0.56)	0.021	−57.2	–
Linked infants with any infant claim	3,184	7.0	−4.48 (1.93)	0.020	−63.8	3.46
Heavy for gestational age (P08.1)						
All linked infants	10,217	2.1	−1.51 (0.54)	0.006	−70.4	–
Linked infants with any infant claim	3,184	6.6	−5.23 (1.86)	0.005	−78.8	3.47
NICU admission (revenue-code indicator)						
All linked infants	10,217	2.4	−1.43 (0.57)	0.012	−60.4	–
Linked infants with any infant claim	3,184	7.4	−4.86 (1.93)	0.012	−65.9	3.39
Normal-newborn nursery admission (revenue-code indicator)						
All linked infants	10,217	14.1	−2.42 (1.37)	0.079	−17.1	–
Linked infants with any infant claim	3,184	43.8	−7.79 (4.02)	0.053	−17.8	3.23
Claims visibility among linked infants						
Share with any infant claim in the window (27.3% at or above the cutoff)	10,217	32.2	−0.61 (1.99)	0.758	−1.9	–
Expected ratio of the two estimates without a visibility effect, $1/s$	–	–	–	–	–	3.10

Notes: NICU: neonatal intensive care unit. Infant outcomes are diagnosis and revenue codes on infant claims from one day before through 30 days after delivery, so an outcome can be recorded only for an infant with at least one claim in that window. Each row is a separate regression discontinuity estimate of the effect of crossing the 95 mg/dL cutoff, from a uniform-kernel local-linear regression within ±10 mg/dL, in percentage points (pp), with standard errors clustered by woman in parentheses. *All linked infants* is every pregnancy with a linked infant record; *Linked infants with any infant claim* keeps those with a claim in the window. Mean below cutoff is the outcome mean among infants within 10 mg/dL below the cutoff, in percent, and the relative effect is the estimate divided by that mean, in percent. The ratio of estimates is the conditional estimate divided by the estimate for all linked infants. With no change in claims visibility at the cutoff, the conditional estimate is expected to be about $1/s$ times the unconditional one, where s is the share of linked infants with any infant claim (32.2% below the cutoff, 27.3% above it; $1/s = 3.10$), because the conditioning removes infants who could not have been flagged. The relative estimates are similar in the all-linked and claim-observed samples. This comparison does not rule out selection into claim observation, and restricting to infants with a claim changes the population described by the estimates. The last panel reports the jump in the share of linked infants with any infant claim, which is small and not significant. Author's calculations from MarketScan claims data.

Appendix K: Secondary Care-Response Estimates

Table 7 reports diabetes medication initiation in the 90 days after the diagnostic lab, and Table K.1 reports provider-contact breadth through 180 days, both summarized in Section 5.2.

Table K.1: Provider contact by specialty

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Endocrinology (mean below cutoff: 5.4%)							
Reduced form	−0.26	−0.26	−0.18	0.11	0.51	0.82	−0.08
	(p=0.782)	(p=0.702)	(p=0.745)	(p=0.824)	(p=0.272)	(p=0.065)	(p=0.900)
Per complier	−2.75	−2.21	−1.41	0.73	2.87	4.17	−1.15
	(p=0.782)	(p=0.702)	(p=0.745)	(p=0.824)	(p=0.273)	(p=0.067)	(p=0.856)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							13.6
Dietitian (mean below cutoff: 2.0%)							
Reduced form	0.10	0.64	0.65	1.09	1.23	1.31	0.51
	(p=0.883)	(p=0.161)	(p=0.087)	(p=0.001)	(p<0.001)	(p<0.001)	(p=0.288)
Per complier	1.02	5.51	5.03	7.17	6.93	6.64	3.73
	(p=0.882)	(p=0.156)	(p=0.084)	(p=0.001)	(p<0.001)	(p<0.001)	(p=0.413)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							12.3
Obstetrics/gynecology (mean below cutoff: 69.2%)							
Reduced form	1.40	0.16	1.03	0.23	−0.09	−0.89	0.80
	(p=0.420)	(p=0.896)	(p=0.314)	(p=0.804)	(p=0.915)	(p=0.274)	(p=0.515)
Per complier	14.58	1.37	7.94	1.50	−0.51	−4.52	8.37
	(p=0.426)	(p=0.896)	(p=0.315)	(p=0.804)	(p=0.915)	(p=0.275)	(p=0.526)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							12.3
Midwife (mean below cutoff: 4.5%)							
Reduced form	0.48	0.12	−0.09	0.10	0.17	0.16	0.19
	(p=0.540)	(p=0.823)	(p=0.843)	(p=0.805)	(p=0.651)	(p=0.657)	(p=0.715)
Per complier	5.00	1.05	−0.71	0.67	0.98	0.83	3.25
	(p=0.539)	(p=0.823)	(p=0.843)	(p=0.804)	(p=0.650)	(p=0.657)	(p=0.577)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							14.5

continued on next page

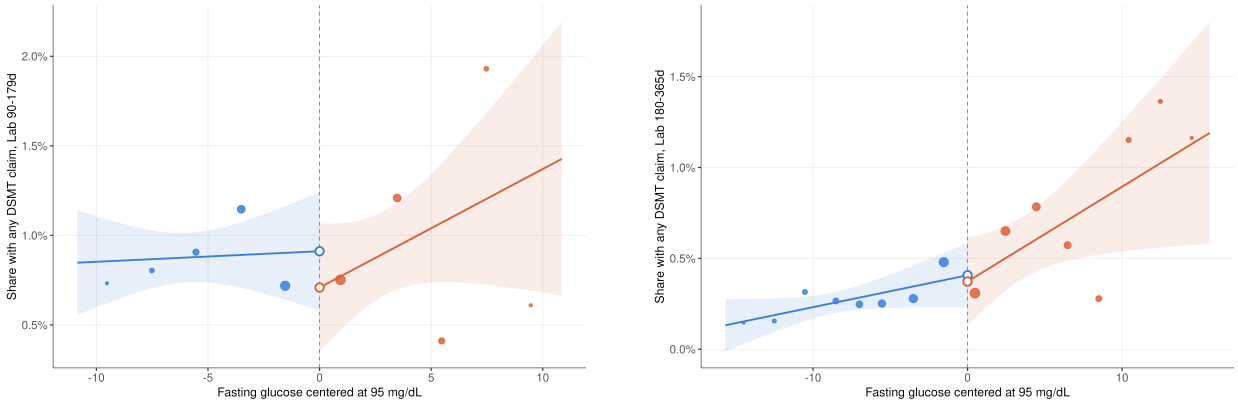
(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Primary-care physician (mean below cutoff: 23.7%)							
Reduced form	−0.53	−0.51	−1.97	−1.44	−0.74	−0.05	−0.43
	(p=0.746)	(p=0.657)	(p=0.041)	(p=0.095)	(p=0.354)	(p=0.948)	(p=0.739)
Per complier	−5.55	−4.39	−15.15	−9.48	−4.20	−0.25	−8.37
	(p=0.746)	(p=0.656)	(p=0.041)	(p=0.094)	(p=0.353)	(p=0.948)	(p=0.559)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							9.9
Advanced-practice primary care (mean below cutoff: 13.4%)							
Reduced form	−1.59	−0.82	−0.89	−0.71	−0.62	−0.06	−0.87
	(p=0.231)	(p=0.377)	(p=0.253)	(p=0.309)	(p=0.338)	(p=0.928)	(p=0.359)
Per complier	−16.60	−7.05	−6.82	−4.67	−3.51	−0.29	−8.16
	(p=0.246)	(p=0.380)	(p=0.257)	(p=0.311)	(p=0.339)	(p=0.928)	(p=0.403)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							12.3
Pediatrics/neonatal (mean below cutoff: 5.7%)							
Reduced form	−0.09	0.52	0.34	0.67	0.95	0.99	0.51
	(p=0.924)	(p=0.412)	(p=0.533)	(p=0.163)	(p=0.034)	(p=0.023)	(p=0.406)
Per complier	−0.91	4.50	2.59	4.43	5.39	5.02	6.07
	(p=0.924)	(p=0.410)	(p=0.531)	(p=0.161)	(p=0.033)	(p=0.022)	(p=0.344)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							14.0
Other/unclassified (mean below cutoff: 61.9%)							
Reduced form	−3.02	−1.36	−0.59	−0.53	0.27	0.29	−1.65
	(p=0.096)	(p=0.286)	(p=0.583)	(p=0.577)	(p=0.760)	(p=0.733)	(p=0.227)
Per complier	−31.40	−11.73	−4.54	−3.52	1.54	1.48	−16.25
	(p=0.113)	(p=0.290)	(p=0.583)	(p=0.577)	(p=0.760)	(p=0.733)	(p=0.228)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							11.0

Notes: Each outcome is an indicator for any claim from a rendering provider of the stated specialty between the test and the earlier of 180 days after the test or the end of the pregnancy window. The diagnosis changes which services a woman receives, not which type of provider she sees. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations

within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Figure K.1: DSMT claims, 90 days and later after the diagnostic lab

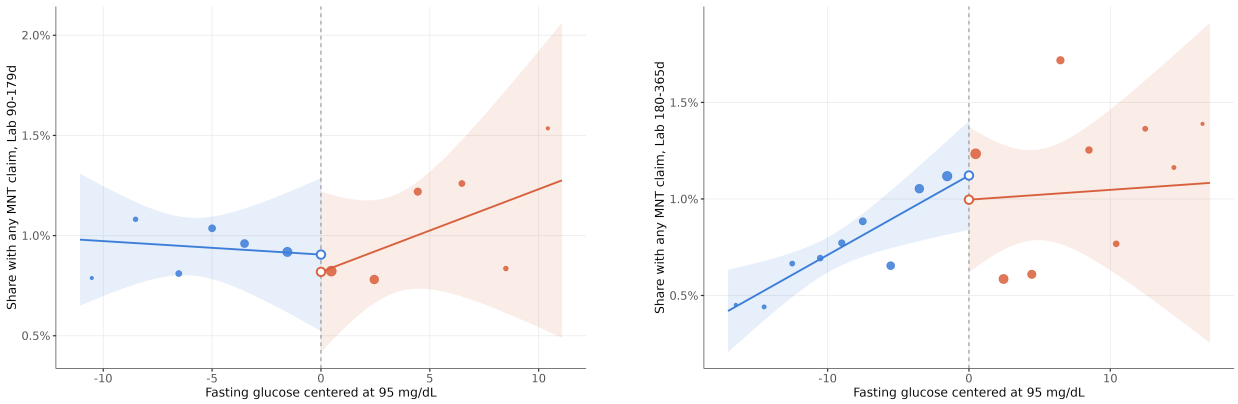


(a) DSMT, 90–179d

(b) DSMT, 180–365d

Notes: Points are bin means of DSMT claim probability against the fasting glucose value, centered at 95 mg/dL, restricted to claims falling in the labeled event-time window. Lines are local-polynomial fits with shaded pointwise 95% confidence bands, estimated with the optimal-bandwidth estimator (triangular kernel, own MSE-optimal bandwidth per panel: (a) ±10.9; (b) ±15.7 mg/dL). RD = −0.002 (p = 0.414), −0.0003 (p = 0.821) for panels (a)–(b) respectively, matching each window's optimal-bandwidth estimate in Table 5. The eligible sample is 85,611 fasting glucose testing episodes for both panels, and each estimate uses the testing episodes inside its own bandwidth. The 0–29 day jump shown in Figure 4 fades to null by these later windows. Author's calculations from MarketScan claims data.

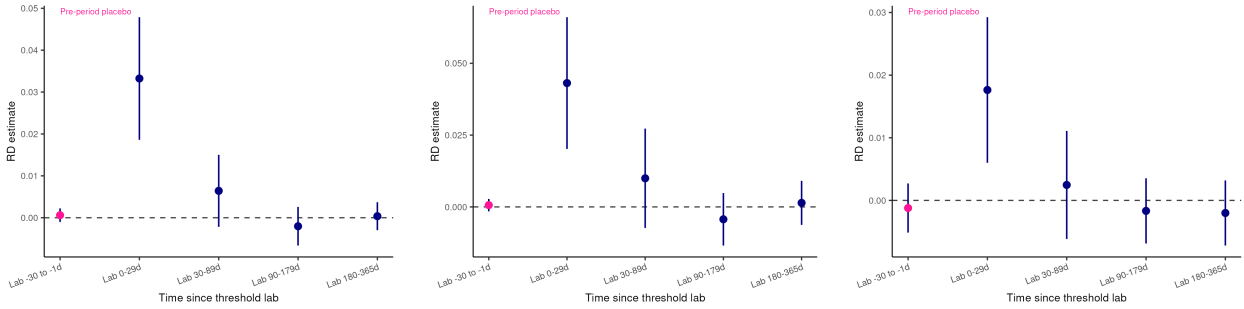
Figure K.2: MNT claims, 90 days and later after the diagnostic lab



(a) MNT, 90–179d (b) MNT, 180–365d

Notes: Points are bin means of MNT claim probability against the fasting glucose value, centered at 95 mg/dL, restricted to claims falling in the labeled event-time window. Lines are local-polynomial fits with shaded pointwise 95% confidence bands, estimated with the optimal-bandwidth estimator (triangular kernel, own MSE-optimal bandwidth per panel: (a) ±11.1; (b) ±17.1 mg/dL). RD = −0.0009 ($p = 0.760$), −0.0013 ($p = 0.601$) for panels (a)–(b) respectively, matching each window's optimal-bandwidth estimate in Table 6. The eligible sample is 85,611 fasting glucose testing episodes for both panels, and each estimate uses the testing episodes inside its own bandwidth. The 0–29 day jump shown in Figure 5 fades to null by these later windows. Author's calculations from MarketScan claims data.

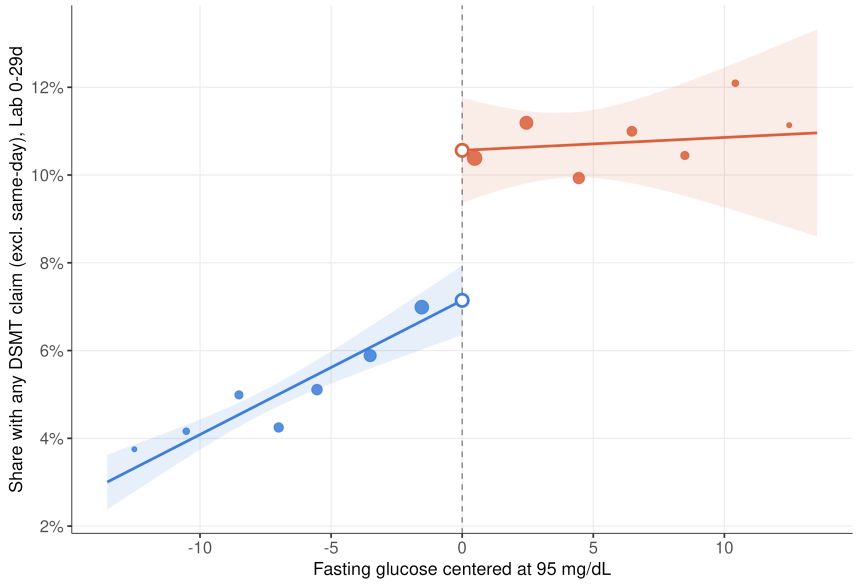
Figure K.3: DSMT/MNT claim response by event time since the diagnostic lab



(a) DSMT, any claim (b) DSMT, claim-day count (c) MNT, any claim

Notes: Local-linear reduced-form estimate (point) and 95% confidence interval (bar) by event-time bucket relative to the diagnostic fasting lab, fixed ±10 mg/dL bandwidth. The "Lab −30 to −1d" point is a pre-diagnosis placebo window: it is null in all three series, while the jump is concentrated in the 0–29 day window and fades by 90 days. See also Figure 4 and Table 5. Author's calculations from MarketScan claims data.

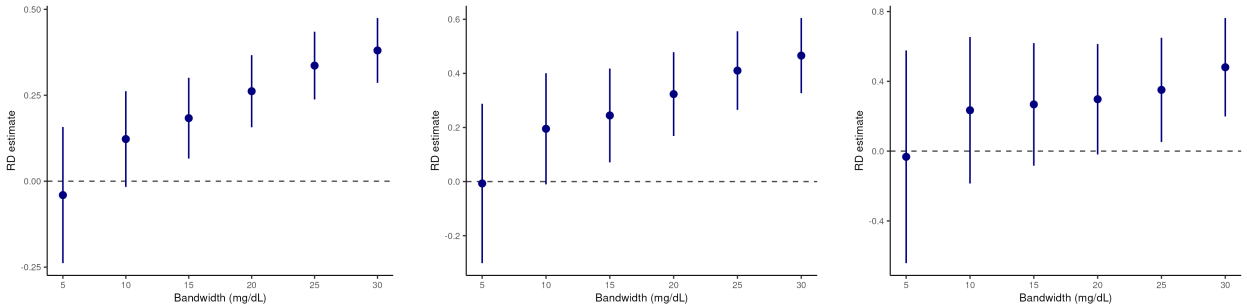
Figure K.4: DSMT claims, 0–29 days post-lab, excluding same-day claims



Notes: Points are bin means against the fasting glucose value, centered at 95 mg/dL, excluding claims dated exactly the same day as the diagnostic lab. Line is a local-polynomial fit with a shaded pointwise 95% confidence band, estimated with the optimal-bandwidth estimator (triangular kernel, MSE-optimal bandwidth, ±13.5 mg/dL). RD = +0.034, $p < 0.001$, matching the optimal-bandwidth row in Table 5. The eligible sample (N = 85,611) is larger than the number of episodes within ±10 mg/dL (26,937). Author's calculations from MarketScan claims data.

Provider-contact breadth shows no discontinuity for any of the eight provider types at the optimal bandwidth (Table K.1). The result is not uniform across bandwidths: at wider bandwidths, dietitian contact becomes significant from ±20 to ±30 mg/dL (positive, and increasing in magnitude with bandwidth) and pediatric/neonatal contact from ±25 to ±30 mg/dL (also positive), while primary-care-physician contact is significant only at ±15 mg/dL (negative, and not significant at any other bandwidth from ±5 to ±30 mg/dL). The provider-contact result in Section 5.2 refers to the optimal bandwidth.

Figure K.5: General care intensity: bandwidth stability, three representative outcomes



(a) Drug-claim days, 90 days (b) Drug-claim days, 180 days (c) Encounters, 180 days

Notes: Local-linear reduced-form estimate (point) and 95% confidence interval (bar) against bandwidth. Both drug-claim-day estimates become significant at conventional levels from ±15 mg/dL onward, and the encounter-day estimate from ±25 mg/dL, the same pattern of estimates that grow with bandwidth reported in the main text for any insulin fill, non-DSMT/MNT spending, and the delivery episode span. All seven outcomes in Table 9 (Figure 8) are null at the fixed ±10 mg/dL bandwidth and under the optimal-bandwidth estimator. Author's calculations from MarketScan claims data.

Appendix L: Spending Bandwidth-Stability Detail

Table 10 reports the full bandwidth sweep for both spending series summarized in Section 5.3: DSMT/MNT spending is positive and significant at every bandwidth tested, while non-DSMT/MNT spending (level scale) is not sign-stable.

Table L.1 reports the reduced-form estimates for each spending series on the $\log(1 + y)$ and inverse hyperbolic sine scales.

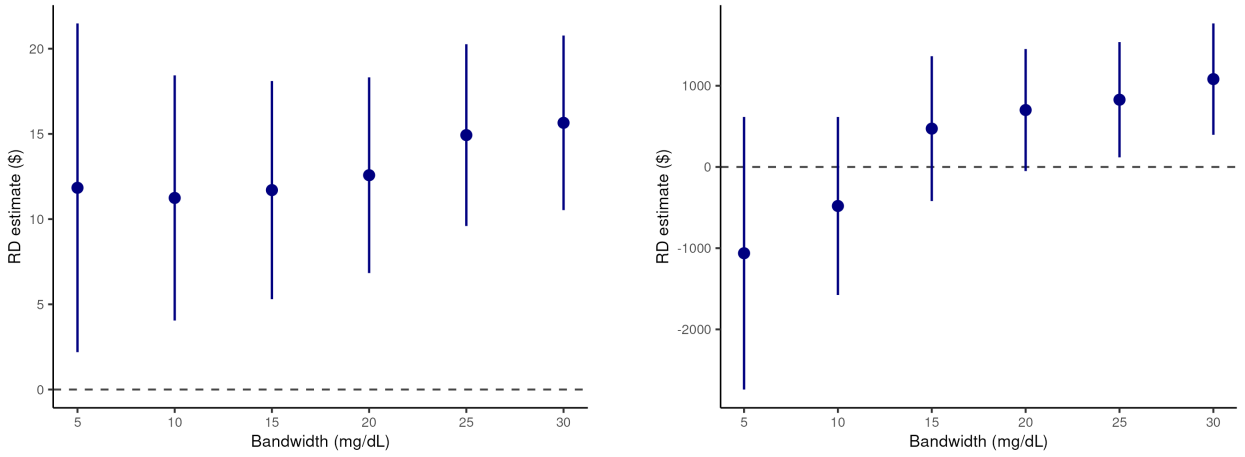
Table L.1: Spending from the test through delivery: log and inverse hyperbolic sine transformations

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Non-DSMT/MNT, $\log(1+y)$ (mean below cutoff: 6.19)							
Reduced form	−0.022	−0.008	0.007	0.009	0.037	0.074	0.005
	(p=0.859)	(p=0.923)	(p=0.921)	(p=0.896)	(p=0.544)	(p=0.208)	(p=0.954)
N	10,421	22,897	38,886	55,621	66,512	70,458	72,683
Optimal h							12.1
Non-DSMT/MNT, IHS (mean below cutoff: 6.82)							
Reduced form	−0.024	−0.013	0.004	0.004	0.033	0.071	0.002
	(p=0.856)	(p=0.884)	(p=0.963)	(p=0.958)	(p=0.604)	(p=0.247)	(p=0.987)
N	10,421	22,897	38,886	55,621	66,512	70,458	72,683
Optimal h							12.2
Total, including pharmacy, $\log(1+y)$ (mean below cutoff: 6.46)							
Reduced form	0.000	0.039	0.037	0.052	0.080	0.119	0.044
	(p=1.000)	(p=0.631)	(p=0.595)	(p=0.400)	(p=0.161)	(p=0.030)	(p=0.599)
N	10,421	22,897	38,886	55,621	66,512	70,458	72,683
Optimal h							11.9
Total, including pharmacy, IHS (mean below cutoff: 7.10)							
Reduced form	−0.003	0.036	0.034	0.049	0.078	0.118	0.041
	(p=0.981)	(p=0.672)	(p=0.638)	(p=0.441)	(p=0.190)	(p=0.039)	(p=0.636)
N	10,421	22,897	38,886	55,621	66,512	70,458	72,683
Optimal h							12.0
Total, medical claims only, $\log(1+y)$ (mean below cutoff: 6.20)							
Reduced form	−0.017	−0.006	0.011	0.014	0.043	0.082	0.010
	(p=0.889)	(p=0.942)	(p=0.884)	(p=0.834)	(p=0.481)	(p=0.165)	(p=0.915)
N	10,421	22,897	38,886	55,621	66,512	70,458	72,683
Optimal h							12.1
Total, medical claims only, IHS (mean below cutoff: 6.83)							
Reduced form	−0.018	−0.011	0.007	0.009	0.039	0.079	0.006
	(p=0.888)	(p=0.901)	(p=0.927)	(p=0.897)	(p=0.540)	(p=0.199)	(p=0.946)
N	10,421	22,897	38,886	55,621	66,512	70,458	72,683
Optimal h							12.1
Pharmacy claims, IHS (mean below cutoff: 2.94)							
Reduced form	0.065	0.254	0.242	0.278	0.358	0.406	0.204
	(p=0.609)	(p=0.004)	(p=0.001)	(p<0.001)	(p<0.001)	(p<0.001)	(p=0.021)
N	10,421	22,897	38,886	55,621	66,514	70,460	72,685
Optimal h							12.9

Notes: Spending is the total payment (plan plus patient) on claims after the glucose test and before the pregnancy endpoint, in nominal dollars; the medical-only total excludes pharmacy claims. Outcomes marked $\log(1+y)$ and IHS apply the log of one plus spending and the inverse hyperbolic sine (IHS) transformation,

respectively, which limit the influence of very large values. The estimates for non-DSMT/MNT payments and the medical-only total are not statistically significant on either transformed scale at the reported bandwidths. These tests do not establish that a large change in dollar spending is absent, and the coefficients describe changes in the transformed outcomes. All outcomes use the same sample. Pharmacy spending is set to zero for an episode with a matched enrollment record that carries the drug-coverage field and no pharmacy claim in the spending window; this rule does not verify affirmative pharmacy coverage. Each cell is the effect of crossing 95 mg/dL on the transformed outcome, with the p-value in parentheses. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Standard errors are clustered by woman. N is the number of episodes within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Figure L.1: Payments after the glucose test and before the pregnancy endpoint: bandwidth stability



(a) DSMT/MNT

(b) Total minus DSMT/MNT

Notes: Local-linear reduced-form estimate (point) and 95% confidence interval (bar) against bandwidth, in untransformed dollars (plan-paid plus out-of-pocket). Panel (a) is positive and significant throughout. Panel (b) crosses zero between ±10 and ±15 mg/dL and becomes significant only at ±25–±30 mg/dL, so this estimate is better described as sign-unstable than as a stable null (Section 5.3). Author's calculations from MarketScan claims data.

Appendix M: Secondary Maternal-Health Estimates

Table M.1 reports antepartum maternal comorbidity diagnosis codes, summarized in Section 5.5.

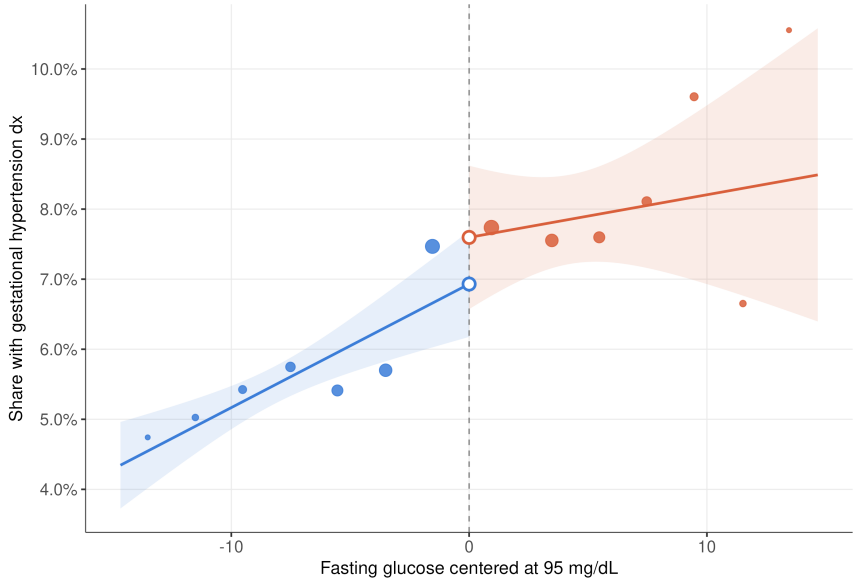
Table M.1: Maternal pregnancy complications

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Gestational hypertension (O13/O16) (mean below cutoff: 5.8%)							
Reduced form	−0.56	0.54	0.55	0.51	0.98	1.16	0.67
	(p=0.569)	(p=0.435)	(p=0.344)	(p=0.329)	(p=0.043)	(p=0.012)	(p=0.305)
Per complier	−5.82	4.64	4.21	3.36	5.51	5.91	4.97
	(p=0.576)	(p=0.432)	(p=0.341)	(p=0.326)	(p=0.042)	(p=0.011)	(p=0.419)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							14.7
Preeclampsia (O14) (mean below cutoff: 3.4%)							
Reduced form	0.17	0.68	0.15	0.33	0.69	0.96	0.61
	(p=0.814)	(p=0.193)	(p=0.740)	(p=0.414)	(p=0.063)	(p=0.007)	(p=0.270)
Per complier	1.80	5.84	1.15	2.17	3.92	4.86	4.63
	(p=0.814)	(p=0.194)	(p=0.740)	(p=0.413)	(p=0.062)	(p=0.007)	(p=0.417)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							11.0
Severe preeclampsia or HELLP syndrome (O14.1/O14.2) (mean below cutoff: 1.6%)							
Reduced form	0.15	0.32	0.06	0.06	0.31	0.47	0.29
	(p=0.777)	(p=0.387)	(p=0.854)	(p=0.835)	(p=0.239)	(p=0.058)	(p=0.462)
Per complier	1.52	2.71	0.44	0.39	1.74	2.41	2.56
	(p=0.777)	(p=0.386)	(p=0.854)	(p=0.835)	(p=0.238)	(p=0.057)	(p=0.486)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							11.2
Polyhydramnios (O40) (mean below cutoff: 1.7%)							
Reduced form	−0.76	−0.33	−0.36	−0.26	−0.17	−0.15	−0.54
	(p=0.128)	(p=0.335)	(p=0.198)	(p=0.310)	(p=0.470)	(p=0.518)	(p=0.120)
Per complier	−7.90	−2.82	−2.80	−1.69	−0.96	−0.74	−5.25
	(p=0.151)	(p=0.342)	(p=0.204)	(p=0.313)	(p=0.472)	(p=0.519)	(p=0.119)
N	12,195	26,937	45,817	65,570	78,361	83,048	85,611
Optimal h							11.5

Notes: Outcomes are indicators for any ICD-10-CM diagnosis code in the stated category from the test through the end of the pregnancy window. More frequent contact with clinicians after a diagnosis could raise recorded diagnosis rates regardless of underlying disease, the mechanism Alalouf, Miller and Wherry (2024) propose for the rise in diagnosed complications they document. Although DSMT/MNT care does increase here, recorded complication rates do not. HELLP: hemolysis, elevated liver enzymes, and low platelets. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth

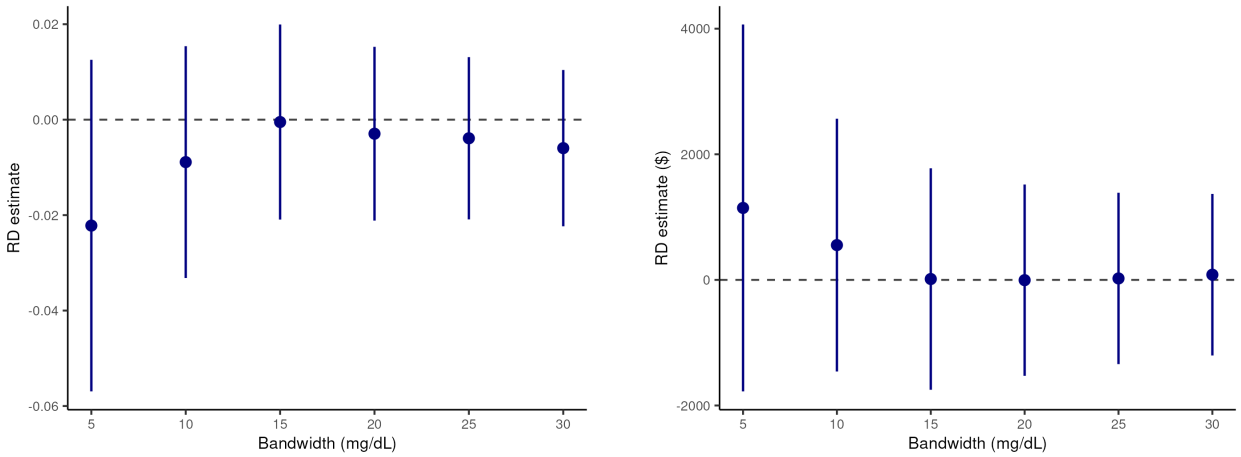
for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Figure M.1: Gestational hypertension diagnosis at the fasting threshold



Notes: Points are bin means against the fasting glucose value, centered at 95 mg/dL. Line is a local-polynomial fit with a shaded pointwise 95% confidence band, estimated with the optimal-bandwidth estimator (triangular kernel, MSE-optimal bandwidth, ±14.66 mg/dL). RD = +0.007, $p = 0.305$, matching the optimal-bandwidth row in Table M.1. The eligible sample (N = 85,611) is larger than the number of episodes within ±10 mg/dL (26,937). Author's calculations from MarketScan claims data.

Figure M.2: Delivery management: bandwidth stability



(a) Cesarean delivery

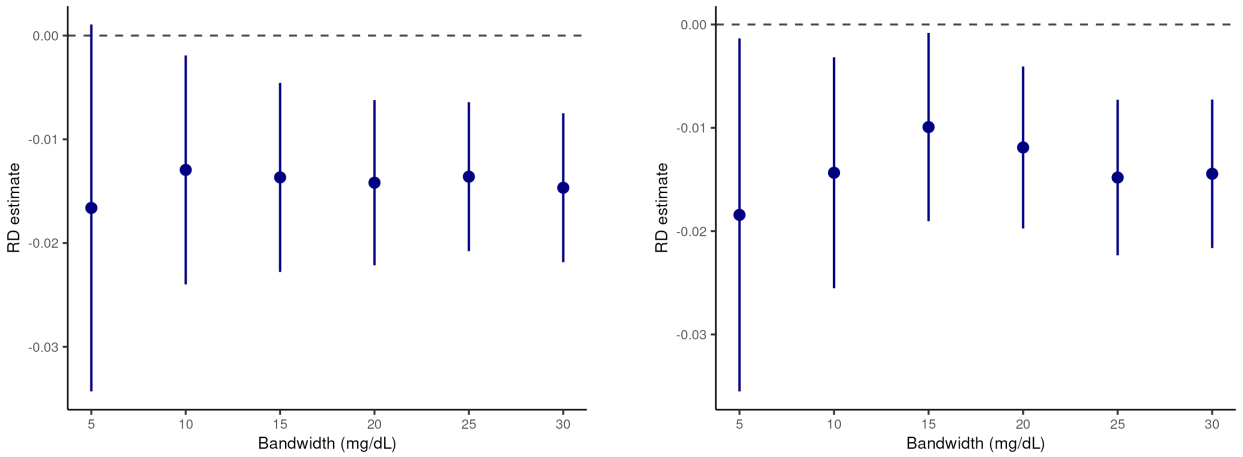
(b) Total delivery payment

Notes: Local-linear reduced-form estimate (point) and 95% confidence interval (bar) against bandwidth. Neither outcome in Table 12 (Figure 13) shows a significant discontinuity at any bandwidth shown. Panel (b) is estimated on deliveries with an observed admission payment. Entry into that subsample is smooth at ±10 mg/dL but rises at the cutoff at ±20 and ±25 mg/dL (Appendix J). Author's calculations from MarketScan claims data.

Appendix N: Secondary Infant-Health Estimates

Table 11 reports the full bandwidth sweep behind the macrosomia and NICU results in Section 5.4: both estimates are negative at every bandwidth from ±5 to ±30 mg/dL and under the optimal-bandwidth estimator, and significant from ±10 mg/dL onward.

Figure N.1: Infant health: bandwidth stability of the macrosomia/NICU result



(a) Macrosomia (ICD-10-CM P08.0 and P08.1, combined)

(b) NICU admission (revenue-code indicator)

Notes: Local-linear reduced-form estimate (point) and 95% confidence interval (bar) against bandwidth, infant-linked subsample. Both series are sign-stable throughout, unlike the non-DSMT/MNT-spending series in Figure L.1. Author's calculations from MarketScan claims data.

Table N.1 reports preterm, jaundice, the NICU claim-day count, and other infant diagnoses in the infant-linked subsample; Section 5.4 summarizes them alongside the main macrosomia and NICU admission results. The NICU claim-day count is a near-duplicate of the NICU admission indicator (about one claim-day per NICU infant in the 31-day window) and is not an independent length-of-stay measure.

Table N.1: Other infant outcomes

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Preterm birth or extreme immaturity (P07.2/P07.3) (mean below cutoff: 2.4%)							
Reduced form	−1.51	−1.15	−1.16	−0.76	−0.80	−0.71	−1.32
	(p=0.112)	(p=0.062)	(p=0.023)	(p=0.086)	(p=0.058)	(p=0.075)	(p=0.038)
Per complier	−14.58	−9.26	−8.28	−4.55	−3.86	−3.00	−10.93
	(p=0.150)	(p=0.075)	(p=0.028)	(p=0.090)	(p=0.060)	(p=0.077)	(p=0.065)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							12.1
Low birth weight (P07.0/P07.1) (mean below cutoff: 0.9%)							
Reduced form	0.37	−0.20	−0.04	0.10	−0.06	−0.03	−0.24
	(p=0.417)	(p=0.601)	(p=0.898)	(p=0.737)	(p=0.832)	(p=0.915)	(p=0.506)
Per complier	3.59	−1.60	−0.29	0.60	−0.29	−0.12	−2.05
	(p=0.429)	(p=0.603)	(p=0.898)	(p=0.737)	(p=0.831)	(p=0.915)	(p=0.507)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							12.8

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Neonatal hypoglycemia or other glucose disorder (P70.2–P70.9) (mean below cutoff: 1.1%)							
Reduced form	−0.54	−0.38	−0.41	−0.36	−0.38	−0.39	−0.41
	(p=0.386)	(p=0.354)	(p=0.222)	(p=0.210)	(p=0.139)	(p=0.123)	(p=0.254)
Per complier	−5.27	−3.04	−2.95	−2.12	−1.85	−1.65	−3.79
	(p=0.400)	(p=0.359)	(p=0.226)	(p=0.212)	(p=0.141)	(p=0.124)	(p=0.260)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							16.4
Neonatal respiratory distress (P22) (mean below cutoff: 1.9%)							
Reduced form	−0.26	−0.59	−0.67	−0.65	−0.82	−0.73	−0.64
	(p=0.732)	(p=0.284)	(p=0.148)	(p=0.104)	(p=0.029)	(p=0.042)	(p=0.203)
Per complier	−2.50	−4.78	−4.78	−3.87	−3.98	−3.06	−5.41
	(p=0.732)	(p=0.288)	(p=0.151)	(p=0.105)	(p=0.029)	(p=0.043)	(p=0.218)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							14.4
Birth injury (P10–P15) (mean below cutoff: 0.9%)							
Reduced form	−0.27	−0.42	−0.09	−0.30	−0.29	−0.35	−0.29
	(p=0.598)	(p=0.260)	(p=0.761)	(p=0.246)	(p=0.237)	(p=0.143)	(p=0.403)
Per complier	−2.66	−3.33	−0.65	−1.78	−1.40	−1.48	−2.60
	(p=0.602)	(p=0.269)	(p=0.761)	(p=0.249)	(p=0.239)	(p=0.145)	(p=0.382)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							13.3
Neonatal jaundice (P59) (mean below cutoff: 9.4%)							
Reduced form	−2.68	−2.05	−2.44	−2.12	−2.32	−2.67	−2.10
	(p=0.102)	(p=0.070)	(p=0.009)	(p=0.011)	(p=0.003)	(p<0.001)	(p=0.059)
Per complier	−25.95	−16.49	−17.41	−12.62	−11.25	−11.21	−16.94
	(p=0.122)	(p=0.074)	(p=0.011)	(p=0.011)	(p=0.003)	(p<0.001)	(p=0.075)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							12.7
NICU claim-day count (mean below cutoff: 0.02)							
Reduced form	−0.019	−0.014	−0.009	−0.011	−0.014	−0.013	−0.013
	(p=0.040)	(p=0.026)	(p=0.082)	(p=0.013)	(p=0.001)	(p=0.001)	(p=0.023)
Per complier	−0.182	−0.112	−0.064	−0.066	−0.068	−0.057	−0.120
	(p=0.078)	(p=0.037)	(p=0.089)	(p=0.016)	(p=0.001)	(p=0.001)	(p=0.023)
N	4,373	10,217	18,619	27,980	34,238	36,500	37,464
Optimal h							14.9

Notes: Estimated on the infant-linked subsample; diagnoses are ICD-10-CM codes on the infant's claims from the day before delivery to 30 days after. These are exploratory secondary outcomes measured in the same linked infant sample. The NICU claim-day count is the number of days with a NICU claim; with about one claim day per infant with a NICU stay, it nearly duplicates the NICU-stay indicator and is not

an independent measure of length of stay. Low birth weight, neonatal hypoglycemia, respiratory distress, and birth injury show no discontinuity. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Donut-hole and placebo-cutoff checks for the infant results

Table N.2 re-estimates the macrosomia and NICU results on the infant-linked subsample (37,464 infants) after excluding infants whose mothers' fasting glucose values lie within 0, 1, 2, 3, or 5 mg/dL of the cutoff. Radius 0 already drops the cell at exactly 95 mg/dL. At the fixed ±10 mg/dL bandwidth this removes 372 infants (9,845 remain of 10,217), so the radius-0 estimates differ from the headline estimates in Table 11. The exact-cutoff cell moves the two outcomes in opposite directions.

The macrosomia estimate keeps its magnitude under donut exclusion but loses precision as the excluded band widens. At the ±20 mg/dL reference bandwidth it is −1.66, −1.27, −1.43, −1.53, and −1.53 percentage points at radii 0, 1, 2, 3, and 5 ($p < 0.001$, 0.021, 0.029, 0.059, and 0.216). At the fixed ±10 mg/dL bandwidth, dropping the exact-cutoff cell strengthens it, from −1.29 points ($p = 0.021$) to −1.72 points ($p = 0.004$). At wider radii it stays between −0.95 and −1.91 points but is not significant (p from 0.21 to 0.51), and at radius 5 only 603 infants remain above the cutoff. Under the optimal-bandwidth estimator it is significant only at radius 0 (−1.84 points, $p = 0.001$).

The NICU estimate depends more on infants at and next to the cutoff. At the primary bandwidth, dropping the exact-cutoff cell moves it from −1.43 points ($p = 0.012$) to −1.19 points ($p = 0.075$; optimal bandwidth −1.21, $p = 0.070$), and it is not significant at any larger radius. At the ±20 mg/dL reference bandwidth it is −1.01, −1.04, and −1.16 points at radii 0, 1, and 2 ($p = 0.029$, 0.043, and 0.050), then −0.86 points at radius 3 ($p = 0.215$) and −0.45 points at radius 5 ($p = 0.673$).

Table N.3 reports macrosomia, its P08.1 component, and NICU admission at placebo cutoffs 15 and 25 mg/dL above and below the true cutoff, at ±5 and ±10 mg/dL and under the optimal-bandwidth estimator. Of the 36 placebo estimates, 25 can be estimated; the rest have too few events or a bandwidth that would span the true cutoff. None of the NICU placebo estimates is significant. Two placebo estimates are significant, and they are the same estimate: at 25 mg/dL above the cutoff with a ±10 mg/dL bandwidth, macrosomia

and its P08.1 component coincide (+4.49 points, $p = 0.042$, $N = 560$). That estimate is opposite in sign to the true-cutoff effect.

Table N.2: Infant outcomes: donut-hole estimates

Donut radius	Local linear, ±10 mg/dL	N	Local linear, ±20 mg/dL	N	Optimal bandwidth
Panel A. Macrosomia (P08.0 or P08.1)					
0 mg/dL	-0.0172 (p=0.004)	9,845	-0.0166 (p<0.001)	27,608	-0.0184 (p=0.001)
1 mg/dL	-0.0095 (p=0.240)	9,142	-0.0127 (p=0.021)	26,905	-0.0065 (p=0.472)
2 mg/dL	-0.0129 (p=0.212)	8,349	-0.0143 (p=0.029)	26,112	-0.0112 (p=0.336)
3 mg/dL	-0.0169 (p=0.226)	7,528	-0.0153 (p=0.059)	25,291	-0.0181 (p=0.195)
5 mg/dL	-0.0191 (p=0.514)	5,844	-0.0153 (p=0.216)	23,607	-0.0125 (p=0.434)
Panel B. Heavy for gestational age, 4,000–4,499 g (P08.1)					
0 mg/dL	-0.0189 (p=0.001)	9,845	-0.0170 (p<0.001)	27,608	-0.0184 (p<0.001)
1 mg/dL	-0.0124 (p=0.109)	9,142	-0.0137 (p=0.009)	26,905	-0.0096 (p=0.274)
2 mg/dL	-0.0185 (p=0.051)	8,349	-0.0162 (p=0.008)	26,112	-0.0186 (p=0.082)
3 mg/dL	-0.0212 (p=0.101)	7,528	-0.0160 (p=0.038)	25,291	-0.0243 (p=0.058)
5 mg/dL	-0.0168 (p=0.566)	5,844	-0.0139 (p=0.260)	23,607	-0.0075 (p=0.629)
Panel C. NICU admission (revenue-code indicator)					
0 mg/dL	-0.0119 (p=0.075)	9,845	-0.0101 (p=0.029)	27,608	-0.0121 (p=0.070)
1 mg/dL	-0.0125 (p=0.113)	9,142	-0.0104 (p=0.043)	26,905	-0.0119 (p=0.105)
2 mg/dL	-0.0176 (p=0.076)	8,349	-0.0116 (p=0.050)	26,112	-0.0232 (p=0.037)
3 mg/dL	-0.0090 (p=0.461)	7,528	-0.0086 (p=0.215)	25,291	-0.0188 (p=0.234)
5 mg/dL	0.0064 (p=0.786)	5,844	-0.0045 (p=0.673)	23,607	-0.0331 (p=0.305)

Notes: Estimated on the 37,464 pregnancies with a linked infant record, the sample of Table 11. Each row excludes pregnancies whose fasting glucose lies within the stated radius of 95 mg/dL; radius 0 excludes only the 372 pregnancies with a value of exactly 95 mg/dL. At a radius of 5 mg/dL and a bandwidth of ±10 mg/dL, 603 infants remain above the cutoff. Number of the 5 radii at which the estimate is significant at the 5% level (local linear at ±10 mg/dL, local linear at ±20 mg/dL, and optimal bandwidth, respectively): macrosomia, 1, 3, and 1; P08.1, 1, 4, and 1; NICU admission, 0, 3, and 1. Local-linear estimates weight observations equally within the bandwidth; optimal-bandwidth estimates use a triangular kernel and a mean-squared-error-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014); these are the conventional estimates at that bandwidth, without bias correction. p-values, in parentheses, are based on standard errors clustered by woman. N is the local-linear estimation sample. Author's calculations from MarketScan claims data.

Table N.3: Infant outcomes: placebo cutoffs

Cutoff offset (mg/dL)	Local linear, ±5 mg/dL	N	Local linear, ±10 mg/dL	N	Optimal bandwidth
Panel A. Macrosomia (P08.0 or P08.1)					
-25	-0.0049 (p=0.427)	10,035	-0.0028 (p=0.564)	19,560	-0.0060 (p=0.330)
-15	0.0012 (p=0.766)	18,516	-0.0006 (p=0.850)	29,262	0.0002 (p=0.953)
0 (true cutoff)	-0.0166 (p=0.066)	4,373	-0.0129 (p=0.021)	10,217	-0.0136 (p=0.006)
15	-0.0117 (p=0.667)	636	0.0068 (p=0.679)	1,433	–
25	–	–	0.0449 (p=0.042)	560	–
Panel B. Heavy for gestational age, 4,000–4,499 g (P08.1)					
-25	-0.0061 (p=0.310)	10,035	-0.0031 (p=0.518)	19,560	-0.0070 (p=0.263)
-15	0.0010 (p=0.790)	18,516	-0.0004 (p=0.900)	29,262	0.0005 (p=0.890)
0 (true cutoff)	-0.0182 (p=0.040)	4,373	-0.0151 (p=0.006)	10,217	-0.0145 (p<0.001)
15	–	–	0.0106 (p=0.513)	1,433	–
25	–	–	0.0449 (p=0.042)	560	–
Panel C. NICU admission (revenue-code indicator)					
-25	0.0056 (p=0.477)	10,035	-0.0003 (p=0.956)	19,560	0.0042 (p=0.592)
-15	-0.0023 (p=0.634)	18,516	0.0012 (p=0.728)	29,262	0.0002 (p=0.957)
0 (true cutoff)	-0.0184 (p=0.035)	4,373	-0.0143 (p=0.012)	10,217	-0.0133 (p=0.011)
15	–	–	-0.0091 (p=0.519)	1,433	–
25	–	–	-0.0047 (p=0.889)	560	–

Notes: Estimated on the 37,464 pregnancies with a linked infant record, the sample of Table 11. Each row estimates the discontinuity at a cutoff shifted by the stated offset from 95 mg/dL; offset 0 is the true cutoff. Of the 25 placebo-cutoff specifications that could be estimated, 2 are significant at the 5% level: macrosomia at offset 25 mg/dL (±10 mg/dL, $N = 560$) and P08.1 at offset 25 mg/dL (±10 mg/dL, $N = 560$), each opposite in sign to the estimate at the true cutoff. 11 of the 36 placebo-cutoff specifications are not estimated (–): 6 because the optimal bandwidth would extend across the true cutoff and 5 because fewer than the minimum number of infants with the outcome on one side of the cutoff within the bandwidth. At the true cutoff, 8 of the 9 estimates are significant at the 5% level. Local-linear estimates weight observations equally within the bandwidth; optimal-bandwidth estimates use a triangular kernel and a mean-squared-error-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014); these are the conventional estimates at that bandwidth, without bias correction. p-values, in parentheses, are based on standard errors clustered by woman. N is the local-linear estimation sample. Author's calculations from MarketScan claims data.

Appendix O: Long-Run Outcomes Detail

Table O.1 reports the postpartum spending estimates summarized in Section 5.6.

Table O.1: Postpartum spending: first and second postpartum years

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Panel A. First postpartum year (through 365 days after delivery)							
Total payments (medical and pharmacy claims) (mean below cutoff: $12,394)							
Reduced form	−$1,225	−$754.63	−$1,113	−$1,373	−$1,497	−$1,541	−$935.28
	(p=0.256)	(p=0.330)	(p=0.087)	(p=0.021)	(p=0.007)	(p=0.009)	(p=0.233)
Per complier	–	−$7,482	−$10,086	−$10,598	−$10,089	−$8,936	−$10,632
		(p=0.315)	(p=0.078)	(p=0.017)	(p=0.006)	(p=0.007)	(p=0.202)
N	4,179	9,344	16,197	23,433	28,200	29,932	30,758
Optimal h							12.0
Non-DSMT/MNT payments (medical claims other than DSMT/MNT) (mean below cutoff: $11,487)							
Reduced form	−$994.45	−$852.14	−$1,069	−$1,441	−$1,581	−$1,634	−$908.72
	(p=0.310)	(p=0.226)	(p=0.069)	(p=0.007)	(p=0.001)	(p=0.002)	(p=0.204)
Per complier	–	−$8,448	−$9,692	−$11,127	−$10,656	−$9,477	−$10,012
		(p=0.212)	(p=0.061)	(p=0.006)	(p=0.001)	(p=0.002)	(p=0.187)
N	4,179	9,344	16,197	23,433	28,200	29,932	30,758
Optimal h							11.6
DSMT/MNT claims, total payment (mean below cutoff: $9.41)							
Reduced form	−$3.17	−$0.41	−$0.96	−$1.15	−$1.22	−$0.58	−$0.66
	(p=0.589)	(p=0.917)	(p=0.793)	(p=0.726)	(p=0.691)	(p=0.843)	(p=0.861)
Per complier	–	−$4.13	−$8.69	−$8.85	−$8.22	−$3.36	−$8.32
		(p=0.917)	(p=0.793)	(p=0.725)	(p=0.690)	(p=0.843)	(p=0.846)
N	4,180	9,347	16,201	23,437	28,210	29,944	30,771
Optimal h							14.8
Total payments, inverse hyperbolic sine (mean below cutoff: 8.85)							
Reduced form	−0.116	−0.021	−0.036	−0.094	−0.115	−0.125	−0.032
	(p=0.406)	(p=0.829)	(p=0.652)	(p=0.196)	(p=0.091)	(p=0.058)	(p=0.735)
Per complier	–	−0.209	−0.330	−0.727	−0.777	−0.725	−0.630
		(p=0.827)	(p=0.647)	(p=0.182)	(p=0.082)	(p=0.050)	(p=0.556)
N	4,179	9,344	16,197	23,433	28,200	29,932	30,758
Optimal h							13.9

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Panel B. First two postpartum years, cumulative (through 730 days after delivery)							
Total payments (medical and pharmacy claims) (mean below cutoff: $18,385)							
Reduced form	−$2,299	$45.51	−$1,242	−$1,608	−$1,198	−$2,869	−$755.28
	(p=0.272)	(p=0.976)	(p=0.321)	(p=0.172)	(p=0.263)	(p=0.159)	(p=0.652)
Per complier	–	–	–	–	–	–	–
N	2,368	5,334	9,227	13,326	16,049	17,028	17,460
Optimal h							9.5
Non-DSMT/MNT payments (medical claims other than DSMT/MNT) (mean below cutoff: $16,449)							
Reduced form	−$1,396	−$9.84	−$968.37	−$1,513	−$1,158	−$2,657	−$283.83
	(p=0.422)	(p=0.994)	(p=0.361)	(p=0.134)	(p=0.212)	(p=0.149)	(p=0.838)
Per complier	–	–	–	–	–	–	–
N	2,368	5,334	9,227	13,326	16,049	17,028	17,460
Optimal h							9.1
DSMT/MNT claims, total payment (mean below cutoff: $16.32)							
Reduced form	$9.38	$14.36	$4.93	$2.75	−$0.49	−$2.72	$13.69
	(p=0.394)	(p=0.067)	(p=0.460)	(p=0.631)	(p=0.930)	(p=0.630)	(p=0.137)
Per complier	–	–	–	–	–	–	–
N	2,368	5,335	9,228	13,327	16,051	17,030	17,463
Optimal h							10.5
Total payments, inverse hyperbolic sine (mean below cutoff: 9.65)							
Reduced form	−0.133	0.037	−0.064	−0.081	−0.062	−0.066	−0.004
	(p=0.334)	(p=0.709)	(p=0.430)	(p=0.275)	(p=0.379)	(p=0.334)	(p=0.968)
Per complier	–	–	–	–	–	–	–
N	2,368	5,334	9,227	13,326	16,049	17,028	17,460
Optimal h							11.2

Notes: DSMT is diabetes self-management training, and MNT is medical nutrition therapy. Payments are total payments (plan plus patient) in nominal dollars, summed over a window that starts the day after the delivery episode ends and runs through 365 days (Panel A) or 730 days (Panel B, which includes the first year) after the delivery date. A fasting glucose testing episode's total below zero is set to zero. Total payments include medical (outpatient, facility, and inpatient) and pharmacy claims. Non-DSMT/MNT payments are total payments on medical claims other than DSMT/MNT claims and exclude pharmacy claims. DSMT/MNT payments are total payments on claims carrying a DSMT or MNT procedure code. The last outcome in each panel is the inverse hyperbolic sine of total payments, so its estimates are approximately proportional (log-point) changes and are shown with three decimals. The sample is pregnancies with an observed delivery and enrollment covering the whole window, so N is smaller in Panel B. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last

column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Table O.2 reports the full grid of 24 outcome-by-horizon estimates behind the secondary maternal-disease null in Section 5.6: hypertension, a cardiovascular-disease composite, dyslipidemia, metabolic syndrome, depression, and chronic kidney disease, each at 12, 24, 36, and 60 months.

Table O.2: Long-run maternal disease: secondary outcomes

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Hypertension (I10–I13)							
Hypertension (I10–I13), 12 months (mean below cutoff: 4.9%)							
Reduced form	−0.23	−1.59	−1.25	−1.14	−1.11	−0.97	−1.44
	(p=0.877)	(p=0.116)	(p=0.142)	(p=0.123)	(p=0.112)	(p=0.152)	(p=0.136)
Per complier	–	−15.81	−11.33	−8.84	−7.47	−5.61	−15.03
		(p=0.139)	(p=0.155)	(p=0.132)	(p=0.118)	(p=0.157)	(p=0.168)
N	4,180	9,347	16,201	23,437	28,210	29,944	30,771
Optimal h							14.2
Hypertension (I10–I13), 24 months (mean below cutoff: 6.9%)							
Reduced form	1.60	−0.04	−1.02	−1.51	−1.28	−1.11	0.45
	(p=0.492)	(p=0.979)	(p=0.449)	(p=0.202)	(p=0.248)	(p=0.297)	(p=0.802)
Per complier	–	–	–	–	–	–	–
N	2,368	5,335	9,228	13,327	16,051	17,030	17,463
Optimal h							10.7
Hypertension (I10–I13), 36 months (mean below cutoff: 9.3%)							
Reduced form	0.44	−1.26	−1.08	−1.85	−1.66	−1.38	−1.64
	(p=0.900)	(p=0.584)	(p=0.577)	(p=0.283)	(p=0.307)	(p=0.370)	(p=0.411)
Per complier	–	–	–	–	–	–	–
N	1,441	3,278	5,741	8,272	9,980	10,585	10,843
Optimal h							18.3

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Hypertension (I10–I13), 60 months (mean below cutoff: 13.8%)							
Reduced form	10.51	5.97	2.17	0.80	1.27	2.53	7.28
	(p=0.101)	(p=0.173)	(p=0.558)	(p=0.805)	(p=0.685)	(p=0.390)	(p=0.163)
Per complier	–	–	–	–	–	–	–
N	645	1,435	2,495	3,579	4,342	4,609	4,721
Optimal h							10.1
Cardiovascular disease composite							
Cardiovascular disease composite, 12 months (mean below cutoff: 0.5%)							
Reduced form	−0.10	−0.08	0.14	0.28	0.32	0.16	−0.12
	(p=0.857)	(p=0.837)	(p=0.661)	(p=0.317)	(p=0.229)	(p=0.583)	(p=0.748)
Per complier	–	−0.76	1.27	2.19	2.19	0.91	−1.23
		(p=0.837)	(p=0.661)	(p=0.322)	(p=0.233)	(p=0.584)	(p=0.750)
N	4,180	9,347	16,201	23,437	28,210	29,944	30,771
Optimal h							12.6
Cardiovascular disease composite, 24 months (mean below cutoff: 0.7%)							
Reduced form	−0.34	0.32	0.81	0.63	0.54	0.25	0.51
	(p=0.731)	(p=0.626)	(p=0.120)	(p=0.182)	(p=0.236)	(p=0.607)	(p=0.366)
Per complier	–	–	–	–	–	–	–
N	2,368	5,335	9,228	13,327	16,051	17,030	17,463
Optimal h							18.6
Cardiovascular disease composite, 36 months (mean below cutoff: 1.1%)							
Reduced form	1.11	0.87	0.96	0.81	0.87	0.54	0.75
	(p=0.427)	(p=0.379)	(p=0.243)	(p=0.277)	(p=0.222)	(p=0.484)	(p=0.464)
Per complier	–	–	–	–	–	–	–
N	1,441	3,278	5,741	8,272	9,980	10,585	10,843
Optimal h							12.0
Cardiovascular disease composite, 60 months (mean below cutoff: 2.2%)							
Reduced form	–	0.07	−0.20	−0.04	−0.11	−0.68	0.58
		(p=0.966)	(p=0.898)	(p=0.976)	(p=0.928)	(p=0.602)	(p=0.737)
Per complier	–	–	–	–	–	–	–
N	–	1,435	2,495	3,579	4,342	4,609	4,721
Optimal h							10.8

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Dyslipidemia (E78.0–E78.5)							
Dyslipidemia (E78.0–E78.5), 12 months (mean below cutoff: 5.1%)							
Reduced form	−0.11	2.31	1.90	1.15	0.95	0.54	1.90
	(p=0.947)	(p=0.036)	(p=0.039)	(p=0.165)	(p=0.218)	(p=0.482)	(p=0.081)
Per complier	–	23.01	17.26	8.86	6.44	3.10	18.66
		(p=0.051)	(p=0.047)	(p=0.168)	(p=0.220)	(p=0.481)	(p=0.123)
N	4,180	9,347	16,201	23,437	28,210	29,944	30,771
Optimal h							14.0
Dyslipidemia (E78.0–E78.5), 24 months (mean below cutoff: 9.4%)							
Reduced form	−1.79	2.54	1.46	−0.88	−0.63	−0.70	1.33
	(p=0.503)	(p=0.169)	(p=0.354)	(p=0.538)	(p=0.637)	(p=0.588)	(p=0.487)
Per complier	–	–	–	–	–	–	–
N	2,368	5,335	9,228	13,327	16,051	17,030	17,463
Optimal h							12.5
Dyslipidemia (E78.0–E78.5), 36 months (mean below cutoff: 14.8%)							
Reduced form	−3.23	0.46	−0.19	−0.99	−0.46	0.06	−1.79
	(p=0.431)	(p=0.870)	(p=0.938)	(p=0.643)	(p=0.817)	(p=0.975)	(p=0.550)
Per complier	–	–	–	–	–	–	–
N	1,441	3,278	5,741	8,272	9,980	10,585	10,843
Optimal h							11.7
Dyslipidemia (E78.0–E78.5), 60 months (mean below cutoff: 23.7%)							
Reduced form	−7.81	−0.90	−2.13	−2.28	−1.22	−1.23	−5.31
	(p=0.275)	(p=0.857)	(p=0.614)	(p=0.541)	(p=0.730)	(p=0.714)	(p=0.343)
Per complier	–	–	–	–	–	–	–
N	645	1,435	2,495	3,579	4,342	4,609	4,721
Optimal h							10.6
Metabolic syndrome (E88.81)							
Metabolic syndrome (E88.81), 12 months (mean below cutoff: 0.5%)							
Reduced form	−0.31	0.22	0.29	0.46	0.37	0.30	0.19
	(p=0.485)	(p=0.509)	(p=0.343)	(p=0.112)	(p=0.196)	(p=0.281)	(p=0.529)
Per complier	–	2.23	2.65	3.54	2.47	1.75	0.81
		(p=0.514)	(p=0.350)	(p=0.120)	(p=0.201)	(p=0.283)	(p=0.810)
N	4,180	9,347	16,201	23,437	28,210	29,944	30,771
Optimal h							16.3

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Metabolic syndrome (E88.81), 24 months (mean below cutoff: 0.9%)							
Reduced form	0.89	0.88	0.91	0.60	0.51	0.43	0.92
	(p=0.193)	(p=0.183)	(p=0.118)	(p=0.296)	(p=0.351)	(p=0.413)	(p=0.145)
Per complier	–	–	–	–	–	–	–
N	2,368	5,335	9,228	13,327	16,051	17,030	17,463
Optimal *h*							13.5
Metabolic syndrome (E88.81), 36 months (mean below cutoff: 1.6%)							
Reduced form	−0.66	0.38	1.12	0.98	0.90	0.84	0.43
	(p=0.697)	(p=0.748)	(p=0.247)	(p=0.276)	(p=0.302)	(p=0.308)	(p=0.718)
Per complier	–	–	–	–	–	–	–
N	1,441	3,278	5,741	8,272	9,980	10,585	10,843
Optimal *h*							13.7
Metabolic syndrome (E88.81), 60 months (mean below cutoff: 3.0%)							
Reduced form	1.49	1.36	0.92	0.25	0.22	0.15	1.65
	(p=0.547)	(p=0.442)	(p=0.557)	(p=0.856)	(p=0.870)	(p=0.910)	(p=0.435)
Per complier	–	–	–	–	–	–	–
N	645	1,435	2,495	3,579	4,342	4,609	4,721
Optimal *h*							11.1
Depression (F32/F33)							
Depression (F32/F33), 12 months (mean below cutoff: 5.7%)							
Reduced form	−0.14	1.04	0.50	0.09	0.46	0.33	0.76
	(p=0.924)	(p=0.330)	(p=0.576)	(p=0.914)	(p=0.543)	(p=0.648)	(p=0.471)
Per complier	–	10.39	4.54	0.67	3.11	1.93	7.76
		(p=0.340)	(p=0.578)	(p=0.914)	(p=0.544)	(p=0.648)	(p=0.500)
N	4,180	9,347	16,201	23,437	28,210	29,944	30,771
Optimal *h*							13.3
Depression (F32/F33), 24 months (mean below cutoff: 8.8%)							
Reduced form	−0.88	0.70	0.17	−1.21	−0.90	−0.98	0.53
	(p=0.721)	(p=0.685)	(p=0.907)	(p=0.349)	(p=0.457)	(p=0.397)	(p=0.766)
Per complier	–	–	–	–	–	–	–
N	2,368	5,335	9,228	13,327	16,051	17,030	17,463
Optimal *h*							12.2
Depression (F32/F33), 36 months (mean below cutoff: 12.0%)							
Reduced form	2.93	2.83	1.90	0.02	0.70	1.00	3.62
	(p=0.423)	(p=0.270)	(p=0.385)	(p=0.990)	(p=0.700)	(p=0.563)	(p=0.195)
Per complier	–	–	–	–	–	–	–
N	1,441	3,278	5,741	8,272	9,980	10,585	10,843
Optimal *h*							10.8

continued on next page

(continued from previous page)

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Depression (F32/F33), 60 months (mean below cutoff: 19.3%)							
Reduced form	0.25	1.63	−1.56	−2.84	−1.04	1.08	1.50
	(p=0.970)	(p=0.722)	(p=0.688)	(p=0.401)	(p=0.744)	(p=0.719)	(p=0.775)
Per complier	–	–	–	–	–	–	–
N	645	1,435	2,495	3,579	4,342	4,609	4,721
Optimal h							10.3
Chronic kidney disease (N18)							
Chronic kidney disease (N18), 12 months (mean below cutoff: 0.3%)							
Reduced form	0.43	0.38	0.26	0.07	0.01	−0.04	0.38
	(p=0.356)	(p=0.214)	(p=0.323)	(p=0.748)	(p=0.970)	(p=0.875)	(p=0.242)
Per complier	–	3.79	2.32	0.55	0.06	−0.21	3.93
		(p=0.236)	(p=0.334)	(p=0.749)	(p=0.970)	(p=0.875)	(p=0.252)
N	4,180	9,347	16,201	23,437	28,210	29,944	30,771
Optimal h							12.3
Chronic kidney disease (N18), 24 months (mean below cutoff: 0.3%)							
Reduced form	0.58	0.55	0.49	0.34	0.24	0.19	0.57
	(p=0.419)	(p=0.254)	(p=0.219)	(p=0.363)	(p=0.515)	(p=0.592)	(p=0.250)
Per complier	–	–	–	–	–	–	–
N	2,368	5,335	9,228	13,327	16,051	17,030	17,463
Optimal h							14.7
Chronic kidney disease (N18), 36 months (mean below cutoff: 0.6%)							
Reduced form	0.36	0.66	0.73	0.65	0.50	0.38	0.62
	(p=0.746)	(p=0.385)	(p=0.250)	(p=0.243)	(p=0.370)	(p=0.498)	(p=0.418)
Per complier	–	–	–	–	–	–	–
N	1,441	3,278	5,741	8,272	9,980	10,585	10,843
Optimal h							14.1
Chronic kidney disease (N18), 60 months (mean below cutoff: 1.4%)							
Reduced form	3.39	3.03	2.22	1.82	1.57	0.81	2.70
	(p=0.223)	(p=0.116)	(p=0.162)	(p=0.183)	(p=0.217)	(p=0.524)	(p=0.165)
Per complier	–	–	–	–	–	–	–
N	645	1,435	2,495	3,579	4,342	4,609	4,721
Optimal h							15.6

Notes: Same subsample and outcome construction as the type 2 diabetes outcomes. The cardiovascular disease composite combines ischemic heart disease, heart failure, and stroke; each component is individually too rare to estimate reliably at most horizons and is not reported separately. The chronic kidney disease rate is 0.23 percent below the cutoff and 0.54 percent above it at 12 months, and 0.83 and 2.08 percent at 60 months. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico,

Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

Table O.3 reports the bandwidth sweep behind the offspring atopic-disease result in Section 5.6: allergic rhinitis is significant at most bandwidths tested, while asthma is significant only at the narrowest bandwidths and decays at wider ones, a pattern that, in a sample this small, cannot be distinguished from noise.

Table O.3: Offspring asthma and allergic rhinitis by 36 months

Outcome	±5	±10	±15	±20	±25	±30	Optimal
Asthma, by 36 months (mean below cutoff: 10.5%)							
Reduced form	18.99	12.86	5.69	1.83	1.35	1.13	13.10
	(p=0.017)	(p=0.021)	(p=0.248)	(p=0.685)	(p=0.730)	(p=0.765)	(p=0.046)
Per complier	–	46.52	23.55	7.80	5.36	4.03	39.56
		(p=0.062)	(p=0.273)	(p=0.688)	(p=0.732)	(p=0.766)	(p=0.096)
N	309	698	1,298	1,969	2,415	2,593	2,654
Optimal h							9.6
Allergic rhinitis, by 36 months (mean below cutoff: 13.2%)							
Reduced form	14.62	13.93	13.56	8.77	9.65	9.52	13.27
	(p=0.107)	(p=0.039)	(p=0.015)	(p=0.091)	(p=0.040)	(p=0.036)	(p=0.047)
Per complier	–	50.41	56.12	37.39	38.19	34.12	47.51
		(p=0.062)	(p=0.034)	(p=0.112)	(p=0.053)	(p=0.045)	(p=0.066)
N	309	698	1,298	1,969	2,415	2,593	2,654
Optimal h							12.8

Notes: The outcome is a diagnosis of the stated condition by 36 months of age, estimated on the infant-linked subsample with the child's own enrollment. Atopic dermatitis and urticaria show no discontinuity at any bandwidth and are not shown. Each cell is the effect of crossing 95 mg/dL, with the p-value in parentheses, in percentage points unless the outcome is a count or a dollar amount. Columns headed $\pm h$ are local linear estimates with a uniform kernel; the last column uses a triangular kernel and the MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014), and I report the conventional estimate. Per complier estimates divide the reduced form by the first stage effect on any GDM diagnosis in the same sample, and they are omitted at every bandwidth for an outcome whose first stage F-statistic is below 10 at ±10 mg/dL or at the fixed bandwidth nearest the optimal one, and at any bandwidth where that bandwidth's own F-statistic is below 10; in the last column they use the bandwidth selected for the fuzzy design. Standard errors are clustered by woman. N is the number of observations within the bandwidth (the full eligible sample in the last column), and h is the optimal bandwidth. Author's calculations from MarketScan claims data.

References

Alalouf, Mattan, Sarah Miller and Laura R. Wherry. 2024. "What Difference Does a Diagnosis Make? Evidence from Marginal Patients." *American Journal of Health Economics* 10(1):97–131. Earlier version: NBER Working Paper No. 26363.

Almond, Douglas. 2006. "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy* 114(4):672–712.

Almond, Douglas, Joseph J. Doyle, Amanda E. Kowalski and Heidi Williams. 2010. "Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns." *Quarterly Journal of Economics* 125(2):591–634.

American College of Obstetricians and Gynecologists. 2018. "ACOG Practice Bulletin No. 190: Gestational Diabetes Mellitus." *Obstetrics & Gynecology* 131(2):e49–e64.
URL: *https://doi.org/10.1097/AOG.0000000000002501*

American Diabetes Association. 2019*a*. "14. Management of Diabetes in Pregnancy: Standards of Medical Care in Diabetes—2019." *Diabetes Care* 42(Supplement 1):S165–S172.
URL: *https://doi.org/10.2337/dc19-S014*

American Diabetes Association. 2019*b*. "2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2019." *Diabetes Care* 42(Supplement 1):S13–S28.
URL: *https://doi.org/10.2337/dc19-S002*

American Diabetes Association Professional Practice Committee. 2026. "15. Management of Diabetes in Pregnancy: Standards of Care in Diabetes—2026." *Diabetes Care* 49(Supplement 1):S321–S338.

Barreca, Alan I., Melanie Guldi, Jason M. Lindo and Glen R. Waddell. 2011. "Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification." *Quarterly Journal of Economics* 126(4):2117–2123.

Bellamy, Leanne, Juan-Pablo Casas, Aroon D. Hingorani and David Williams. 2009. "Type 2 Diabetes Mellitus after Gestational Diabetes: A Systematic Review and Meta-Analysis." *Lancet* 373(9677):1773–1779.

Bharadwaj, Prashant, Katrine Vellesen Løken and Christopher Neilson. 2013. "Early Life Health Interventions and Academic Achievement." *American Economic Review* 103(5):1862–91.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.103.5.1862*

Black, Sandra E., Paul J. Devereux and Kjell G. Salvanes. 2007. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *Quarterly Journal of Economics* 122(1):409–439.

Blatt, Alan J., Jon M. Nakamoto and Harvey W. Kaufman. 2011. "Gaps in Diabetes Screening During Pregnancy and Postpartum." *Obstetrics & Gynecology* 117(1):61–68.

Bütikofer, Aline, Katrine V. Løken and Kjell G. Salvanes. 2019. "Infant Health Care and Long-Term Outcomes." *Review of Economics and Statistics* 101(2):341–354.

Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82(6):2295–2326.

Carpenter, Marshall W. and Donald R. Coustan. 1982. "Criteria for Screening Tests for Gestational Diabetes." *American Journal of Obstetrics and Gynecology* 144(7):768–773. **URL:** *https://doi.org/10.1016/0002-9378(82)90349-0*

Cattaneo, Matias D., Michael Jansson and Xinwei Ma. 2020. "Simple Local Polynomial Density Estimators." *Journal of the American Statistical Association* 115(531):1449–1455.

Centers for Disease Control and Prevention. 2023. "QuickStats: Percentage of Mothers with Gestational Diabetes, by Maternal Age — National Vital Statistics System, United States, 2016 and 2021." *Morbidity and Mortality Weekly Report* 72(1):16.

Chyn, Eric, Samantha Gold and Justine Hastings. 2021. "The returns to early-life interventions for very low birth weight children." *Journal of Health Economics* 75:102400. **URL:** *https://www.sciencedirect.com/science/article/pii/S0167629620310468*

Crowther, Caroline A., Diane Samuel, Lesley M. E. McCowan, Richard Edlin, Thach Tran and Christopher J. D. McKinlay. 2022. "Lower versus Higher Glycemic Criteria for Diagnosis of Gestational Diabetes." *New England Journal of Medicine* 387:587–598.

Crowther, Caroline A., Janet E. Hiller, John R. Moss, Andrew J. McPhee, William S. Jeffries and Jeffrey S. Robinson. 2005. "Effect of Treatment of Gestational Diabetes Mellitus on Pregnancy Outcomes." *New England Journal of Medicine* 352(24):2477–2486.

Currie, Janet and Jonathan Gruber. 1996. "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women." *Journal of Political Economy* 104(6):1263–1296.

Davis, Jody, Amy Hess Fischl, Joni Beck, Lillian Browning, Amy Carter et al. 2022. "2022 National Standards for Diabetes Self-Management Education and Support." *Diabetes Care* 45(2):484–494.

Duryea, Elaine L., Josiah S. Hawkins, Donald D. McIntire, Brian M. Casey and Kenneth J. Leveno. 2014. "A Revised Birth Weight Reference for the United States." *Obstetrics & Gynecology* 124(1):16–22.

Farrar, Diane, Lelia Duley, Therese Dowswell and Debbie A. Lawlor. 2016. "Different Strategies for Diagnosing Gestational Diabetes Mellitus to Improve Maternal and Infant Health." *Cochrane Database of Systematic Reviews* 2016(8):CD007122.

Farrar, Diane, Mark Simmonds, Maria Bryant, Trevor A. Sheldon, Derek Tuffnell, Su Golder, Fidelma Dunne and Debbie A. Lawlor. 2016. "Hyperglycaemia and Risk of Adverse Perinatal Outcomes: Systematic Review and Meta-Analysis." *BMJ* 354:i4694.

Farrar, Diane, Mark Simmonds, Maria Bryant et al. 2017. "Risk Factor Screening to Identify Women Requiring Oral Glucose Tolerance Testing to Diagnose Gestational Diabetes: A Systematic Review and Meta-analysis and Analysis of Two Pregnancy Cohorts." *PLoS ONE* 12(4):e0175288.

Ferrara, Assiamira, Monique M. Hedderson, Charles P. Quesenberry and Joseph V. Selby. 2002. "Prevalence of Gestational Diabetes Mellitus Detected by the National Diabetes Data Group or the Carpenter and Coustan Plasma Glucose Thresholds." *Diabetes Care* 25(9):1625–1630.

Frandsen, Brigham R. 2017. Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete. In *Regression Discontinuity Designs: Theory and Applications*, ed. Matias D. Cattaneo and Juan Carlos Escanciano. Vol. 38 of *Advances in Econometrics* Emerald Publishing pp. 281–315.

Gaggero, Alessio, Joan Gil, Dolores Jiménez-Rubio and Eugenio Zucchelli. 2022. "Does Health Information Affect Lifestyle Behaviours? The Impact of a Diabetes Diagnosis." *Social Science & Medicine* 314:115420.

Gaggero, Alessio, Joan Gil, Dolores Jiménez-Rubio and Eugenio Zucchelli. 2023. "Sick and Depressed? The Causal Impact of a Diabetes Diagnosis on Depression." *Health Economics Review* 13:34.

HAPO Study Cooperative Research Group. 2008. "Hyperglycemia and Adverse Pregnancy Outcomes." *New England Journal of Medicine* 358(19):1991–2002.
URL: *https://doi.org/10.1056/NEJMoa0707943*

Harper, Lorie M., Lisa Mele, Mark B. Landon, Marshall W. Carpenter, Susan M. Ramin, Uma M. Reddy et al. 2016. "Carpenter-Coustan Compared with National Diabetes Data Group Criteria for Diagnosing Gestational Diabetes." *Obstetrics & Gynecology* 127(5):893–898.

Heckman, James J. 2008. "Schools, Skills, and Synapses." *Economic Inquiry* 46(3):289–324.
URL: *https://doi.org/10.1111/j.1465-7295.2008.00163.x*

Hillier, Teresa A., Kathryn L. Pedula, Keith K. Ogasawara, Kimberly K. Vesco, Caryn E. Oshiro, Sylvia L. Lubarsky and Jan Van Marter. 2021. "A Pragmatic, Randomized Clinical Trial of Gestational Diabetes Screening." *New England Journal of Medicine* 384:895–904.

Hoynes, Hilary, Marianne Page and Ann Huff Stevens. 2011. "Can Targeted Transfers Improve Birth Outcomes? Evidence from the Introduction of the WIC Program." *Journal of Public Economics* 95(7):813–827.

Imbens, Guido W. and Charles F. Manski. 2004. "Confidence Intervals for Partially Identified Parameters." *Econometrica* 72(6):1845–1857.

International Association of Diabetes and Pregnancy Study Groups Consensus Panel. 2010. "International Association of Diabetes and Pregnancy Study Groups Recommendations on the Diagnosis and Classification of Hyperglycemia in Pregnancy." *Diabetes Care* 33(3):676–682.
URL: *https://doi.org/10.2337/dc09-1848*

Kolesar, Michal and Christoph Rothe. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *American Economic Review* 108(8):2277–2304.

Landon, Mark B., Catherine Y. Spong, Elizabeth Thom, Marshall W. Carpenter, Susan M. Ramin, Brian Casey et al. 2009. "A Multicenter, Randomized Trial of Treatment for Mild Gestational Diabetes." *New England Journal of Medicine* 361(14):1339–1348.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76(3):1071–1102.

Lee, David S. and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48(2):281–355.

Mullainathan, Sendhil and Ziad Obermeyer. 2022. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care." *The Quarterly Journal of Economics* 137(2):679–727.
URL: *https://doi.org/10.1093/qje/qjab046*

National Academies of Sciences, Engineering, and Medicine. 2023. *Inclusion of Pregnant and Lactating Persons in Clinical Trials: Proceedings of a Workshop*. Washington, DC: The National Academies Press.

National Academies of Sciences, Engineering, and Medicine. 2024. *Advancing Clinical Research with Pregnant and Lactating Populations: Overcoming Real and Perceived Liability Risks*. Washington, DC: The National Academies Press.

National Diabetes Data Group. 1979. "Classification and Diagnosis of Diabetes Mellitus and Other Categories of Glucose Intolerance." *Diabetes* 28(12):1039–1057.
URL: *https://doi.org/10.2337/diab.28.12.1039*

National Institute for Health and Care Excellence. 2015. Diabetes in Pregnancy: Management from Preconception to the Postnatal Period. NICE guideline NG3 National Institute for Health and Care Excellence London: . Published 25 February 2015; updated 16 December 2020.
URL: *https://www.nice.org.uk/guidance/ng3*

National Institutes of Health Consensus Development Conference Panel. 2013. "National Institutes of Health Consensus Development Conference Statement: Diagnosing Gestational Diabetes Mellitus, March 4–6, 2013." *Obstetrics & Gynecology* 122(2 Pt 1):358–369.
URL: *https://doi.org/10.1097/AOG.0b013e31829c3e64*

Nethery, Elizabeth, Jennifer A. Hutcheon, Jerry Lee, Patricia A. Janssen and Laura Schummers. 2025. "Effects of a Province-Wide Change in Gestational Diabetes Mellitus Screening Policy on Treatment and Newborn Birth Weight." *Diabetes Care* 48(8):1410–1417.

Oster, Emily. 2014. *Expecting Better: Why the Conventional Pregnancy Wisdom Is Wrong—and What You Really Need to Know*. Revised ed. Penguin.
URL: *https://books.google.com/books?id=80ECDAAAQBAJ*

O'Sullivan, J. B. and C. M. Mahan. 1964. "Criteria for the Oral Glucose Tolerance Test in Pregnancy." *Diabetes* 13:278–285.
URL: *https://pubmed.ncbi.nlm.nih.gov/14166677/*

Riukula, Krista. 2023. "The Effects of Screening for Gestational Diabetes." *Empirical Economics* 65:1931–1964.

Shankar, Mridula, Alya Hazfiarini, Thiago Melo Santos, Joshua P. Vogel, Rosalind McDougall, Annie McDougall, Sara Rushwan, Ahmet Metin Gülmezoglu and Meghan A. Bohren. 2026. "Bridging disciplinary siloes: a scoping review on the inclusion and exclusion of pregnant and lactating populations in clinical research." *BMJ Public Health* 4(1):e003423.

Smirnakis, Kelly V., Lisa Chasan-Taber, Myles Wolf, Glenn Markenson, Jeffrey L. Ecker and Ravi Thadhani. 2005. "Postpartum Diabetes Screening in Women with a History of Gestational Diabetes." *Obstetrics & Gynecology* 106(6):1297–1303.

Song, Dandan, John C. Hurley and Maryam Lia. 2020. "Estimated Treatment Effects of Tight Glycaemic Targets in Mild Gestational Diabetes Mellitus: A Multiple Cut-Off Regression Discontinuity Study." *International Journal of Environmental Research and Public Health* 17(21):7725.

Szmuilowicz, Emily D., Jami L. Josefson and Boyd E. Metzger. 2019. "Gestational Diabetes Mellitus." *Endocrinology and Metabolism Clinics of North America* 48(3):479–493.

Tennant, Peter W. G., Emily Doxford-Hook, Lucy Flynn, Kate Kershaw, Jenny Goddard and Tomasina Stacey. 2022. "Fasting Plasma Glucose, Diagnosis of Gestational Diabetes and the Risk of Large for Gestational Age: A Regression Discontinuity Analysis of Routine Data." *BJOG: An International Journal of Obstetrics & Gynaecology* 129(1):82–89.

US Preventive Services Task Force, Karina W. Davidson, Michael J. Barry, Carol M. Mangione et al. 2021. "Screening for Gestational Diabetes: US Preventive Services Task Force Recommendation Statement." *JAMA* 326(6):531–538.

Vounzoulaki, Elpida, Kamlesh Khunti, Sophia C. Abner, Bee K. Tan, Melanie J. Davies and Clare L. Gillies. 2020. "Progression to Type 2 Diabetes in Women with a Known History of Gestational Diabetes: Systematic Review and Meta-Analysis." *BMJ* 369:m1361.

Yan, Ji. 2020. "Healthy Babies." *American Journal of Health Economics* 6(2):199–215.