The Effect of Co-Payments on the Take-Up of Prenatal Tests

Marina Di Giacomo Massimiliano Piacenza Luigi Siciliani Gilberto Turati*

29 October 2021

Abstract

Noninvasive prenatal screening tests help identify genetic disorders in a fetus, but their take-up remains low in several countries. Using a regression discontinuity design, we test the causal effect of a policy that eliminated co-payments for noninvasive screening tests in Italy. We identify the treatment effects by a discontinuity in women's eligibility for a free test based on their conception date. We find that the policy increases the probability of women's undergoing noninvasive screening tests by 5.5 percentage points, and the effect varies by socioeconomic status. We do not find evidence of substitution effects with more expensive and riskier invasive diagnostic tests. In addition, the increase in take-up does not affect pregnancy termination or newborn health. We find some evidence of positive effects on mothers' health behaviors during pregnancy as measured by reductions in mothers' weight gain and hospital admissions during pregnancy, but these are statistically significant only at the 10 percent level.

Keywords: Prenatal tests, Co-payments, Maternal and newborn health, Regression Discontinuity Design.

JEL Classification: I18, I12, I14.

* Marina Di Giacomo: University of Torino, Department of Economics, Social Sciences, Applied Mathematics and Statistics (ESOMAS), (e-mail: marina.digiacomo@unito.it);

Massimiliano Piacenza (Corresponding author): University of Piemonte Orientale, Department of Economics and Business (DISEI), Novara, Italy (e-mail: massimiliano.piacenza@uniupo.it);

Luigi Siciliani: University of York, Department of Economics and Related Studies, York, United Kingdom (e-mail: luigi.siciliani@york.ac.uk);

Gilberto Turati: Università Cattolica del Sacro Cuore, Department of Economics and Finance, Rome, Italy (e-mail: gilberto.turati@unicatt.it).

1 Introduction

The medical literature has long recognized that appropriate prenatal care, such as recommended doctor visits and tests in the first stage of pregnancy, can decrease the likelihood of adverse outcomes for both the mother and the child (e.g., Almond and Currie 2011). Prenatal tests are offered in the first trimester of pregnancy in many high-income countries to identify possible genetic disorders, for example, Down syndrome and Edwards' syndrome, in the fetus (Boyd et al. 2008; Palomaki et al. 2006; Grimes and Schulz 2002). Two different types of tests are typically available. Noninvasive *screening* tests, such as combined or integrated prenatal screening tests, are generally inexpensive and allow doctors to estimate the risk of genetic disorders. Invasive *diagnostic* tests are more expensive but permit doctors to diagnose the presence of chromosomal anomalies with greater accuracy through a genetic map of the fetus.

When combined, these tests can detect congenital anomalies that affect around 0.3 to 0.5 percent of newborns. Congenital anomalies are a leading cause of infant mortality, morbidity, and long-term disability (EUROCAT 2010). Accuracy, safety, and the mother's peace of mind are often cited as the benefits of the prenatal tests (e.g., Dormandy et al. 2005). Types of genetic screening programs, as well as medical recommendations and guidelines, are heterogeneous across countries. At the beginning of the 2000s, many countries implemented reforms to harmonize policies at the national and international levels, but the process remains ongoing (World Health Organization 2016).

The literature (e.g., Crombag et al. 2014) and international health organizations (EUROCAT 2010) highlight the variability in take-up rates for prenatal tests across and within countries as a critical issue. Many factors explain such variability: differences in service delivery and health systems, socioeconomic traits and cultural components of the population, and market failures (e.g., underestimation of benefits and positive externalities associated with the use of prenatal care). Policymakers justify introducing free screening programs to the underserved by pointing to the perceived suboptimal take-up of screening tests within some institutional settings (Currie 2006; Shurtz et al. 2016).

In this study, we apply quasi-experimental methods to investigate whether eliminating co-payments for noninvasive *screening* tests is an effective policy to increase the take-up of prenatal testing. We use a regression discontinuity design (RDD) to quantify the effect of a government policy in a large region in Italy (Piedmont) that eliminated co-payments for noninvasive prenatal screening tests. Identification of the treatment effects is determined by a discontinuity in women's eligibility for obtaining free prenatal screening tests before and after the policy cutoff date in October 2009.

We, therefore, use time as our running variable and compare women over a narrow time window before and after the policy is introduced.

This paper addresses three research questions. First, we test the extent to which eliminating co-payments for noninvasive tests increases their utilization and whether, in turn, this reduces the use of costlier and riskier invasive tests (a substitution effect). Second, we investigate heterogeneity in the effect across different groups, particularly women from low socioeconomic status. Although some noninvasive prenatal screening tests are inexpensive, even a relatively small co-payment can be a barrier and discourage women from disadvantaged groups from undergoing the tests. We, therefore, investigate whether the elimination of the co-payment not only increases use but also whether it reduces inequalities in access to prenatal care. Third, the lack of prenatal care is associated with poor birth outcomes, such as low birth weight, preterm birth, and infant mortality (Woodhouse et al. 2014; Currie and Rossin-Slater 2015; Corman et al. 2018), as well as worse health of mothers (Conway and Kutinova, 2006). Thus, we investigate whether an increase in prenatal tests affects mothers' health behaviors and newborn health as a result of additional visits, contacts, and interactions with midwives and doctors.

We provide evidence on these questions by considering the interesting case of Italy. Regions in Italy have a high degree of autonomy in legislating healthcare policies, including prenatal and maternal care. Prenatal tests for the diagnosis of Down syndrome and other chromosomal disorders have been available since 2001 for all Italian women, including those covered by the Regional Health Service of the Piedmont Region.[1] The regional co-payment scheme required women to contribute to the cost of these tests. The co-payment for noninvasive screening tests was in the range of €27–€54, depending on the specific test provided (the combined test, the triple test, or the integrated test). The co-payment was in the €160–200 range for invasive diagnostic tests, depending on the specific test provided (chorionic villus sampling or amniocentesis) and several maternal characteristics.[2]

In 2009, the regional government of Piedmont eliminated the co-payment for noninvasive screening tests, making the noninvasive tests available to all women free of charge. There was no change to the co-payment for invasive diagnostic prenatal tests. The aim of the policy was twofold.

[1] Piedmont is a large region located in the northwest of Italy. In 2010, it had around 4.5 million inhabitants, and the number of annual births was around 35,000. The Piedmont annual GDP per capita was about €30,000 in purchasing power standards, in 2016 (Eurostat). Piedmont has roughly the same population as Ireland, or a medium-sized US state, such as Kentucky or Louisiana. The crude birth rate (number of live births per 1,000 inhabitants) is around 8, which is slightly lower than the Italian average birth rate (8.5), and much lower than the EU (10), and the US birth rates (12.5).

[2] The Regional Health Service reimbursed hospitals €98 for a noninvasive screening test (e.g., the integrated test) and €592 for an invasive diagnostic test (e.g., amniocentesis).

First, public authorities wanted to improve the *efficiency* of prenatal testing by increasing the take-up of noninvasive screening tests (and, as a result, reducing the inappropriate use of invasive diagnostic tests). According to medical guidelines, noninvasive screening tests allow women with a high (low) risk of delivering an infant with congenital anomalies to be identified. Women identified as "low risk" do not require further testing; this reduces the need for invasive diagnostic tests and the number of miscarriages related to these invasive tests. On the contrary, women identified as "high risk" (i.e., those with a positive screening test) can obtain the invasive diagnostic tests for free. Increasing take-up as a result of the policy is expected to improve detection of congenital anomalies up to about 90 percent since noninvasive tests are more effective than a simple alternative screening method based on a fixed threshold rule identifying all pregnant women aged 35 years or over as "high risk." [3] In addition, recent developments make noninvasive tests almost as reliable as invasive tests.[4] Increasing take-up will bring about a relatively minor increase in direct costs (i.e., those associated with the provision of prenatal tests by the national health system) and a reduction in indirect costs (i.e., those associated with invasive tests, like miscarriages) since noninvasive tests are less costly and less risky than the more-invasive diagnostic tests.

Second, public authorities aimed to improve *equity* in prenatal testing by mitigating disparities in access to prenatal tests. Observed take-up rates were meager among more deprived areas within the region and lower socioeconomic groups.

The eligibility rules justify our RDD approach to identifying the effects of the co-payment reform. First, after the policy became effective on October 1, 2009, all pregnant women were automatically eligible. Second, the initial general policy was announced in May 2008. Still, the detailed rules about the policy were made public only in August 2009, which almost coincided with the start of the policy on October 1, 2009. The timing of the regulatory acts prevents any strategic behavior of women around the policy change date. Third, we can reasonably rule out any manipulation of the pregnancy decision as a strategic reaction to eliminating the co-payment because the price of prenatal tests represents a negligible portion of the costs associated with having a child. The decision to have a child involves many other factors, including the costs of raising a child, and eliminating the co-payment should not affect the number of women becoming pregnant after the policy implementation.

[3] The detection rate based on maternal age only (a maternal age at conception of 35 or over) is around 45 percent, while the detection rate for an integrated test, the most common noninvasive screening test, is around 90 percent (Wald et al. 2003).

[4] Boyd et al. (2008) and Palomaki et al. (2013), for instance, point out that many improvements in prenatal screening methods are responsible for the increase in birth defects detection rates.

The policy elements generate a discontinuity in the treatment assignment between women who are tested before and after the policy cutoff date. We exploit this discontinuity in eligibility to obtain an exogenous variation in participation. We, therefore, use time as our running variable, an approach that many recent papers adopted in different empirical settings, e.g., housing (Fetter 2013, Moulton et al. 2018), energy consumption (Ito 2015), health (Halla et al. 2016, Aguilara et al. 2021), pollution abatement (Li et al. 2020), drug consumption (Hansen et al. 2020), and domestic violence (Carr and Packham 2021). This approach exploits the timing of the policy change to causally estimate its effects, under the key identifying assumption that there are no changes in the outcome variable other than those caused by the policy itself. Finally, we use a relatively narrow window of 52 weeks before and after the cutoff date to avoid concerns of secular trends in prenatal testing. We also check that our results are robust to an even narrower (or broader) window. We finally provide several falsification tests using as alternative cutoff dates both the same month but in different years, before and after 2009, and different months in 2009. We show no evidence of a change in testing behavior in these alternative dates, ruling out any potential anticipation effects.

Our key results are as follows. First, in our baseline specification, we find that eliminating the co-payment increases the probability that a woman undergoes any prenatal tests by 6.6 percentage points. Women's increased use in the screening test, by 5.5 percentage points, mainly drives the results. The average pre-policy rate of women taking only the noninvasive screening test was 63 percent; the policy increased the take-up rate for screening tests by 8.7 percent. However, we do not observe any substitution effects because eliminating the co-payment did not reduce the take-up rate of the costlier invasive diagnostic tests. This result is driven mainly by the behavior of women older than 35 years at conception. Before the elimination of the co-payment, these women could already access free diagnostic tests based on a legislative provision that considered them to be at high risk of congenital disorders. After the co-payments were eliminated for all women, women older than 35 could still obtain screening tests free of charge. However, these women have not changed their behavior after the policy change; they are still likely to skip the screening tests and undergo invasive diagnostic tests. Hence, the reduction in the proportion of women who do not undergo any test is due entirely to the increase in the proportion of women who undergo noninvasive screening tests. If it takes time for the effects of the reform to entirely accrue, our RDD approach provides a conservative estimate of the full effect of the policy.

Second, we find heterogeneity in the impact of the reform across women with different socioeconomic statuses. The elimination of the co-payment increased the take-up rate, especially for women 25–34 years old, those born abroad, those who reside in metropolitan areas, and those with an intermediate level of education.

Third, accounting for self-selection into prenatal care, we assess whether screening tests affect the health outcomes of mothers and newborns. Having a prenatal test early in the pregnancy increases the number of visits and contacts with midwives and doctors, which in turn increases the opportunity for clinicians to advise on lifestyle and potentially harmful effects of unhealthy behaviors that may impact newborns' health outcomes (Raatikainen et al. 2007, Harris et al. 2012, Metcalfe et al. 2013). After ruling out any increase in the number of pregnancy terminations following the policy, we use the discontinuity in eligibility of free prenatal screening tests triggered by the policy as an instrumental variable for utilization. However, the IV results show a reduction in mothers' weight gain and hospital admissions during pregnancy triggered by prenatal testing, which are only marginally statistically significant. We do not find any effect of prenatal screening tests on the health status of newborns at delivery, which is consistent with the results on maternal behavior.

Prenatal tests have been the focus of economic studies. Fajnzylber et al. (2010) develop and calibrate a model of amniocentesis choice and find that the amniocentesis take-up rate should decrease (instead of increase) with age, once all the risk factors are considered (e.g., the risk of an affected child, the risk of miscarriage associated with amniocentesis, as well as the risk of a decline in fertility with age). Seror (2008) uses an interview-based survey to assess different models of amniocentesis choice. Gajdos et al. (2016) exploit French amniocentesis regulations to offer a measure of the disutility associated with a child with Down syndrome.

The studies by Garrouste et al. (2011) and Shurtz et al. (2016) are the closest to our work. They investigate the effect of public policies aimed at subsidizing diagnostic prenatal tests (amniocentesis) in France and Israel, respectively. Both studies find that utilization rates rise sharply when women are eligible for full reimbursements.

Our contribution differs from previous studies in three key dimensions. First, our main focus is on noninvasive prenatal screening tests. To our knowledge, no other economics paper has studied this type of prenatal test. Screening tests are not conclusive, but they are less expensive and widely used across different medical specialties (e.g., cholesterol measurement tests, Pap tests, fecal occult blood tests, mammograms) to identify subpopulations that may require additional care.[5] Our study contributes to identifying the effects of the subsidization of prenatal screening tests on take-up rates and, in turn, on mother and newborn health. Second, we test for substitution effects between screening

[5] Within the literature on screening programs, Bitler and Carpenter (2016, 2017) find that more-generous insurance coverage induces a large and significant increase in the utilization rates of breast cancer and cervical cancer screening programs (mammograms and Pap tests, respectively) in the US. Cohen et al. (2015) find that subsidizing a rapid diagnostic malaria test doubled the test take-up rate in Kenya.

and diagnostic tests, which none of the preceding papers has investigated. Third, the institutional setting is different, which allows us to compare findings across health systems.

We also contribute to the literature on prenatal care and its short-run effects on the health of mothers and children. Most of the existing literature on prenatal care focuses on children's health (see Corman et al. 2018 for a recent survey). Only a few papers explore, within a causal setting, the beneficial effects of prenatal care on health during pregnancy (e.g., Conway and Kutinova 2006; Yan 2017). Our contribution closes this gap in the existing literature by estimating the causal effects of prenatal screening tests on the health of both mothers and children.

The remainder of the paper proceeds as follows. Section 2 presents the institutional background. A conceptual framework to inform the empirical analysis is provided in Section 3. Section 4 describes the data, and Section 5 describes the empirical strategy. In Section 6 we present the results, and Section 7 draws policy implications and conclusions.

2 The Reform of the Co-Payment Scheme for Prenatal Tests

2.1 Prenatal tests

Prenatal tests for Down syndrome (trisomy 21) and other genetic diseases (trisomy 18 or Edwards syndrome, and spina bifida) belong to two broad classes: diagnostic and screening tests. Diagnostic tests are more invasive and mainly consist of removing a sample of fluid from the amniotic sac. Given their invasive nature, there is a risk of miscarriage (around 1 percent). The two most common diagnostic tests are amniocentesis and chorionic villus sampling (CVS), the only prenatal genetic tests for Down syndrome until the mid-1980s. More recently, some new prenatal screening tests have been developed. These tests are noninvasive, and they mostly consist of an ultrasound scan (nuchal translucency) and blood tests.[6] The screening tests do not diagnose the presence of a chromosomal disorder, but they estimate the likelihood that the fetus carries certain genetic diseases. These tests are safe for the mother and the fetus, but they lack precision, as false positives or false negatives are

[6] The most common noninvasive screening test in Piedmont is the integrated test. It consists of two phases. The first phase (in the first trimester) involves an ultrasound scan (nuchal translucency) and blood tests (the concentration of a chemical called pregnancy-associated plasma protein A, PAPP-A). The second phase involves a blood test in which the levels of three markers are measured (alpha-fetoprotein, free beta-hCG, unconjugated estriol). Results from the two phases, along with maternal age, are "integrated" to compute the risk for chromosomal anomalies for that baby. The last generation of noninvasive screening tests (NIPT, also called Non-Invasive Prenatal Screening—NIPS, or cell free DNA test—cfDNA) have become increasingly available since 2015. These tests are based on an analysis of small fragments of fetus DNA circulating in the mother's blood. In the case of a high probability outcome, a diagnostic invasive follow-up test is still necessary. Our empirical analysis is based on data preceding the introduction of these new screening tests.

possible. If the screening test is positive, the mother must decide whether to seek a definitive (but invasive and risky) diagnostic test.

Most high-income countries offer prenatal genetic screening programs. However, there is significant heterogeneity across countries in policies, recommendations, guidelines, and program take-up results (Javaher et al. 2010). Moreover, variability in take-up rates for prenatal tests is explained by differences in service delivery across health systems, as well as socioeconomic traits and culture (Javaher et al. 2010; Palomaki et al. 2013; Vassy et al. 2014; Crombag et al. 2014).

Beginning in the early 2000s, many countries implemented new rules to harmonize national and international policies,[7] but the process is still under way (World Health Organization 2016). The Netherlands and Sweden show the lowest prenatal screening take-up rates (below 30 percent), while Denmark, France, Belgium, and Iceland have the highest (above 80 percent). The United States, the United Kingdom, Finland, and Italy are in between (between 60 and 70 percent), with significant regional differences *within* countries (Palomaki et al. 2013; Vassy et al. 2014; Crombag 2016). The picture for invasive diagnostic tests is even more fragmented because eligibility varies according to age or risk assessment of genetic disorders. For example, amniocentesis is routinely offered to women older than 35 in the United States, Germany, Sweden, Italy, and Spain, women older than 38 in France and Norway, and women older than 39 in Finland. In the United Kingdom, only women with a high risk of genetic disorder as identified in a screening test have access to invasive diagnostic tests.

2.2 Institutional setting and the policy reform

The Italian National Health Service (NHS) is a tax-funded system providing universal coverage to all citizens. While the central government provides funding of the NHS and the framework legislation, the management and provision of day-to-day health services are performed on the regional level (Turati 2013). Within limits set by national legislation, the twenty regional governments organize their hospital network, take decisions over staff in public hospitals, and set the diagnosis-related group (DRG) rates and the co-payment rules for resident citizens.

Piedmont's reorganization of perinatal healthcare services started in 2001 to comply with the national maternal and child healthcare services regulations.[8] The process completes in 2007, with the approval of the regional healthcare plan that outlines the political, organizational, and administrative

[7] Among others, Boyd et al. (2008), EUROCAT (2010), Javaher et al. (2010), Crombag et al. (2014) and Vassy et al. (2014) survey some recent European national programs. Plachinski (2017) reviews US policies over the last twenty years.
[8] Italian Department for Health, Ministerial Decree, September 10, 1998, followed by the *Progetto obiettivo Materno Infantile* (Project for the Promotion of Maternal, Infant, and Child Health, Decree, April 24, 2000), and the Decree on Essential Levels of Assistance, February 14, 2001.

guidelines for the protection and promotion of health in Piedmont, incorporating all the national regulations.[9]

Prenatal tests have been routinely offered to all pregnant women since 2001. Since then, a co-payment set at the regional level has been in place, which means that women must contribute to the cost of these tests. In 2009, the regional government of Piedmont introduced a reform of its co-payment scheme for prenatal screening tests.[10] No other reforms at the regional or national level affected the health system at the same time.

The new policy made the noninvasive screening tests free for women residing within the administrative borders of the region but maintained the co-payment for diagnostic tests. Before the 2009 reform, the co-payment was in the range of €27–€54 for noninvasive screening tests, while for invasive tests, the co-payment was €160–€200 depending on the specific test provided (the combined test, the triple test, or the integrated test, and the chorionic villus sampling or the amniocentesis). Some exemptions from the co-payment requirement were available for the invasive diagnostic tests based on the risk of the pregnancy: i.e., if the mother was age 35 or older at the time of conception, there were other cases of chromosomal disorders among members of the family, or the mother had a positive screening test. After the 2009 reform, the co-payment for the noninvasive screening tests was eliminated for all women. Nothing changed in the co-payment scheme for the invasive diagnostic tests: all women had to pay the co-payment, except for women with high-risk pregnancies.

Pregnant women can access prenatal tests with a medical prescription from a midwife, a general practitioner (GP), or a gynecologist. The physician must obtain the pregnant woman's informed consent to the genetic screening. The information provided when obtaining consent must include the characteristics of the test (reliability), the methods of execution, and a detailed description of the test. If a genetic problem is diagnosed, the physician gives information and advice about possible behavioral changes during pregnancy, care for the infant after birth, and ending the pregnancy. A woman's participation in a prenatal program is voluntary and is based on her personal preferences, culture, social background, and medical information (Santalahti et al. 1998, Markens et al. 1999, Hall et al. 2006). The 2009 co-payment reform did not introduce any new information campaigns or guidelines for obstetricians and gynecologists. Public hospitals' medical and administrative staff received an administrative circular concerning the elimination of co-payment and the details about the exemptions for medical prescriptions after October 1, 2009 (e.g., the application forms, the prescription codes, and the procedures to adopt when prescribing the prenatal tests).[11] Moreover, there

[9] The *Piano Socio-Sanitario Regionale* (the Regional Healthcare Plan), in Regional Law (L.R.) n. 18, August 6, 2007.
[10] Resolutions of the Regional Government (Delibera della Giunta Regionale) n. 38-11960, August 4, 2009.
[11] Administrative circular by the Piedmont Regional Health Department, dated September 30, 2009.

were no advertising campaigns about the newly established free tests for the entire population. Thus, pregnant women know they will be offered a free prenatal screening test at their first prenatal visits.

Within our sample, only public hospitals offer prenatal tests because they are the only accredited institutions that satisfy the required international quality standards in the region (AReSS 2008). As only public hospitals can supply the tests, this rules out any monetary incentives for physicians to prescribe prenatal tests. The co-payment flows directly to the hospital, partly paid by the regional government according to a predetermined rate. Physicians working outside hospitals are paid by capitation, while those working within the hospital are salaried public employees.

Policymakers justified the reform on two grounds.[12] The first rationale was efficiency. Noninvasive screening tests are cost-effective in detecting pregnancies at high risk of a genetic disorder (Ohno and Caughey, 2013), and their take-up should increase.[13] On the one hand, public health authorities claimed that the take-up rate for noninvasive tests was low because risk assessment was often based on a simple alternative screening method based on maternal age only, which is largely imprecise as a screening method. On the other hand, spurred by recent medical evidence (Wald et al. 2003; Malone et al. 2005; Palomaki et al. 2006; Loane et al. 2013), public authorities recognized that noninvasive screening tests should substitute for costlier and riskier diagnostic tests as the first test to be taken by pregnant women to assess the risk of congenital anomalies. The final aim of the policy was to identify the greatest number of cases while using the best clinical practices that limit the use of invasive diagnosis methods, which are riskier for the mother's and infant's health and are also more expensive for the public budget.

The second rationale for the reform was related to equity. There was regional evidence, collected by a multidisciplinary working group advising the regional government (see footnote 12) of inequalities in the use of prenatal tests based on socioeconomic status (SES) and cultural traits. This evidence is also supported by the medical literature (Dormandy et al. 2005; Fransen et al. 2010), which reports low take-up rates for some disadvantaged socioeconomic groups. Thus, the new Piedmont policy aimed to achieve a higher take-up rate of prenatal screening tests and to mitigate

[12] The 'birth and pregnancy care plan' *Percorso Nascita* (literally, birth pathway) defined by the regional resolution (Delibera della Giunta Regionale) n. 34-8769, May 12, 2008 contains the justification of the policy. The plan is a summary report of the regulatory framework for pregnancy care. The plan is prepared by a multidisciplinary working group, and its objective is twofold. On the one hand it summarizes the national and regional legislation in force, within the framework of international guidelines on the perinatal care by Italian and International Scientific Societies. On the other hand, it identifies some critical areas, suggesting possible solutions. The plan recognizes prenatal screening and diagnostic tests as a critical area.

[13] In a technical report, the Regional Health Agency (AReSS, 2008) stated that a 90 percent take-up rate for prenatal screening tests was a viable, though ambitious, target in Piedmont.

inequalities in access to tests. There is no evidence of low SES pregnant women raising concerns about unequal access to prenatal testing.

3 Conceptual Framework: Demand for Screening and Diagnostic Tests

To understand the possible effects of the reform on take-up rates, we consider a simple conceptual framework. Define $P^n(c^s, c^d)$ as the proportion of women who do not undergo any test; $P^s(c^s, c^d)$ as the proportion of women who undergo the noninvasive screening test; and $P^d(c^s, c^d)$ as the proportion of women who undergo *only* the invasive diagnostic test, where c^s is the co-payment for the noninvasive screening test, and c^d is the co-payment for the invasive diagnostic test.

Considering standard medical practice, we assume that women who undergo the invasive diagnostic test first never also undergo the noninvasive screening test, since the invasive diagnostic test provides more accurate information than the noninvasive screening one. We also assume that out of $P^s(c^s, c^d)$ who undergo the noninvasive screening test, a proportion Z undergoes only the noninvasive test, while a proportion $(1 - Z)$ also undergoes the invasive diagnostic test. The latter could include women who had a positive test result (conveying "bad news") from the screening test who decide to follow up with the more invasive diagnostic test.

The whole population can therefore be split into four groups, which could be interpreted as related to the demand for each of the tests, or for both tests:

- $P^n(c^s, c^d)$ do not undergo any test,
- $P^d(c^s, c^d)$ undergo *only* the invasive diagnostic test,
- $Z \times P^s(c^s, c^d)$ undergo *only* the noninvasive screening test,
- $(1 - Z) \times P^s(c^s, c^d)$ undergo *both* tests,

where $P^n + P^d + P^s Z + P^s(1 - Z) = 1$, so that $P^n = 1 - P^d - P^s$.

We assume that $\partial P^s / \partial c^s < 0$, $\partial P^s / \partial c^d \geq 0$: an increase in the co-payment for the noninvasive screening test reduces the proportion of women who undergo the noninvasive screening test, and an increase in the co-payment for the invasive diagnostic test increases the proportion of women undergoing the noninvasive screening test if the tests are substitutes.

Similarly, we assume that $\partial P^d / \partial c^d < 0$, $\partial P^d / \partial c^s \geq 0$: an increase in the co-payment for the invasive diagnostic test reduces the proportion of women who undergo only the invasive test, and an increase in the co-payment for the noninvasive screening test increases the proportion of women who undergo the invasive diagnostic test if the tests are substitutes.

The hypotheses, which we test in the following sections that relate to the elimination of the co-payment for the noninvasive screening test on four groups of women are as follows:

1) $\partial(ZP^s)/\partial c^s < 0$. The elimination of the co-payment for the noninvasive screening test *increases* the proportion of women who undergo *only* the noninvasive screening test.
2) $\partial P^d/\partial c^s \geq 0$. The elimination of the co-payment for the noninvasive screening test weakly *reduces* the proportion of women who undergo *only* the invasive diagnostic test.
3) $\partial((1-Z)P^s)/\partial c^s \leq 0$. The elimination of the co-payment for the noninvasive screening test weakly *increases* the proportion of women who undergo *both* tests.
4) $\partial P^n/\partial c^s = \partial(1-P^s-P^d)/\partial c^s = -\partial P^s/\partial c^s - \partial P^d/\partial c^s$. The elimination of the co-payment for the noninvasive screening test weakly *decreases* the proportion of women who do not undergo any test if the increase in the proportion of women who undergo the noninvasive screening test is higher than the decrease in the proportion of women who undergo the invasive diagnostic test.

Finally, note that the elimination of the co-payment for the noninvasive screening tests always increases the proportion of the population undergoing the noninvasive screening test, P^s, that is, the sum of those who undergo only the noninvasive test or both tests. The effect on the proportion undergoing the invasive diagnostic test, $(P^d + (1-Z)P^s)$, that is the sum of those who undergo only the invasive test or both, is instead indeterminate and given by $\partial P^d/\partial c^s + (1-Z)\,\partial P^s/\partial c^s \gtrless 0$. On one hand, the elimination of the co-payment for the noninvasive screening test encourages a reduction in the proportion of those who undergo only the invasive diagnostic test due to the substitution effect (the first term above) but also increases the proportion of those who undergo both tests (the second term above), due to a higher proportion of positive test results (conveying "bad news"), which in turn induces more women to undergo the invasive test. These results, therefore, highlight the usefulness of splitting the whole population into four groups (i.e., no test, only noninvasive screening test, only invasive diagnostic test, both tests), rather than three groups (i.e., no test, noninvasive test, and invasive test). In the empirical section, we test these four hypotheses.

4 Data

4.1 Sources and sample definition

We exploit the administrative archive of the CEDAP (CErtificato Di Assistenza al Parto, literally "Delivery Certificate"). The Ministry of Health introduced the Delivery Certificate in 2001.[14] The midwife or the doctor who attended the birth is responsible for filling the form within ten days of the

[14] Ministerial Decree n. 249, July 16, 2001.

delivery. Despite the national regulation, data collection is administered and managed at the regional level, and there are differences in the information collected across regions. There are some essential data that all regions have to collect: a set of variables about sociodemographic characteristics of the parent(s), the course of the pregnancy, labor, childbirth, and the health status of the newborn. However, each regional administration can collect additional information for their purposes. The Piedmont region has collected additional data on prenatal testing since 2002: whether the mother underwent a prenatal test and, if so, which type of test. Most of the other regions only collect data on diagnostic (invasive) prenatal tests, and few regions do not collect information on prenatal tests.

In general, CEDAP data are disclosed on an aggregate annual basis by each regional administration. The Ministry of Health publishes a yearly CEDAP report providing descriptive evidence on maternal and child health. We obtained access to microdata for all women who give birth within the administrative borders of the Piedmont region, whose information is in the delivery certificate CEDAP database.[15] The data cover around 98 percent of all births within the Region. Home births (representing 0.2 percent of total births) and deliveries in private hospitals (1.8 percent of total births) are unavailable. All deliveries in our sample occurred in one of the 32 public hospitals within the administrative area of Piedmont.

The certificate reports a large set of women characteristics and health behaviors during pregnancy (e.g., age, education level, mother's and father's employment status, marital status, twin pregnancy, previous miscarriages and abortions, nationality, smoking and alcohol consumption, weight gain, folic acid supplements, and hospital admissions) in addition to the type of prenatal test. Unfortunately, it does not report the exact date of the prenatal test nor the result of the test. The certificate also records the health status of the infant immediately after birth. The sources of all data in the certificate are medical records and official personal data. However, the socioeconomic information (marital status, education level, and employment status) and some health behaviors during pregnancy (smoking and alcohol consumption, folic acid supplement intakes) are self-reported.

We define our final sample according to the following criteria. First, we consider all women who became pregnant one year before or one year after the elimination of the co-payment, which occurred on October 1, 2009. We are primarily interested in the conception dates. In our sample, the calculation of the conception date is based on the gestational age of the fetus at birth, which, in turn,

[15] We do not have information on women terminating their pregnancy before delivery occurs. Pregnancies mainly end before delivery because of spontaneous abortions or because of induced (or voluntary) abortions. In Piedmont, the former represents around 11 percent of all pregnancies, while the latter about 15 percent of all pregnancies. The source of information is the Regione Piemonte, Assessorato Sanità.

is computed from the first day of the last normal menstrual period (LMP) to the birth date in weeks (Quintana-Domeque and Ródenas-Serrano 2017, Lautharte 2021). If the menstrual period is regular (28 days), the conception usually takes place two weeks (14 days) after the LMP. Therefore, we recover the conception week by subtracting the gestational age at birth, measured in weeks and reduced by two, from the date of birth.[16] The gestational age is a fundamental parameter for a safe and healthy pregnancy (Verloove et al. 1986). In the baseline specifications, we use the "week" as the observational unit for our forcing variable (i.e., the number of weeks from conception to/from the elimination of the co-payment). To check the robustness of our results, we also consider the presumed conception day (obtained by multiplying the weeks of gestation by seven), instead of the presumed conception week, as the observational unit for the forcing variable. We also explore issues related to the actual date of the prenatal tests in Section 6.4.

The total number of observations at this step is 70,960 women conceiving 52 weeks before and 52 weeks after the policy change, October 1, 2009. Then, we restrict our sample to nulliparous women who did not exhibit any pathological disorders during pregnancy (high blood pressure, diabetes, placenta praevia, psychological disease, the familial occurrence of congenital malformation, etc.). Following previous literature, excluding women who experienced past pregnancies (we excluded 32,954 observations here) allows us to avoid some confounding effects of the policy, as experience may have a prominent role in the choice of prenatal tests (and prenatal care behavior, in general). The reason for excluding women with any pathological disorders (we exclude 5,383 observations) is to remove all women who, in the pre-2009 period, had access to free prenatal tests due to their high-risk pregnancies.[17]

After applying our inclusion criteria, the final sample consists of 32,623 women: 51.5 percent became pregnant during the pre-reform period (16,788 women), while the remaining 48.5 percent became pregnant in the post-reform period (15,835 women).

4.2 Variables definitions and summary statistics

Our first set of dependent variables includes the prenatal tests take-up rates, as reported in the delivery certificate. Consistent with our conceptual framework, we consider the four following utilization rates:

[16] The presumed conception date is only partially based on a woman's report. The definitive estimated conception date is defined by the physician on the basis of a number of fetus measures at the first ultrasound exam. Usually, this estimation is accurate, and ultrasound exams have an intrinsic error of up to five to seven days. Its manipulation would be considered a serious medical negligence.

[17] We also experiment with including these women (multiparous women and/or women with pathological conditions) in our estimated models, and the results do not qualitatively change.

- No test (P^n): a binary indicator equal to one if the woman had no prenatal tests during her pregnancy, and zero otherwise;
- Screening (noninvasive) test only (ZP^s): a binary indicator equal to one if the woman had only a noninvasive prenatal screening test, and zero otherwise;
- Diagnostic (invasive) test only (P^d): a binary indicator equal to one if the woman had only an invasive diagnostic prenatal test, and zero otherwise;
- Both tests ($(1-Z)P^s$): a binary indicator equal to one if the woman had a prenatal screening test followed by a diagnostic test, and zero otherwise.

Table 1 shows summary statistics for the take-up rate of prenatal tests in the whole sample and the pre- and post-policy periods. Comparing the pre-reform period to the post-reform period, the percentage of women taking up only the noninvasive screening tests increases by around 8 percentage points (from 63.2 percent in the pre-reform period to 71.3 percent in the post-reform period), while the rate of invasive diagnostic tests increases by 0.6 percentage points (from 6.8 percent to 7.4 percent). On the other hand, the percentage undergoing no tests decreases by 8.8 percentage points: from 27 percent to 18.2 percent after the policy change. The take-up rate for those undergoing both tests only marginally increases (0.2 percentage points).

The second set of dependent variables includes women's and newborn health status. For the health status of mothers, we exploit information about lifestyle during pregnancy: weight gain during pregnancy, smoking, alcohol consumption, the use of folic acid, and hospital admissions. For the health status of the newborns, we use the following indicators, all measured at birth: newborn weight, length, and head circumference, the Apgar scores one minute and five minutes after birth, a resuscitation binary indicator, preterm birth, stillbirth, and sex. Table 2 reports the summary statistics for the health outcomes of women (Panel A) and of newborns (Panel B).

International guidelines set forth optimal weight gain during pregnancy (Institute of Medicine 2009), which should not exceed 11–13 kg (24–29 pounds) for a woman of normal weight. Excessive weight gain is associated with poor health outcomes such as hypertension, gestational diabetes, miscarriage, and delivery complications (Currie et al. 2010). In our analysis, we use two measures of weight gain. A continuous variable (Weight Gain in Pregnancy) expressed in kg, and a dichotomous variable of excessive weight gain (Weight Gain in Pregnancy>15 kg) equal to one if the weight gain during pregnancy is higher than 15 kg (see Currie et al. 2010). For a large number of women (about 5,000), we do not have data on weight gain; however, the balancing tests support the comparability of women around the cutoff. The average weight gain during pregnancy in our sample is 13 kg, with a lower average in the post-policy period (13 kg) relative to the pre-policy year (13.1 kg). Around

one quarter of women gained more than 15 kg; this proportion is slightly smaller in the post-policy period (23.3 percent versus 24 percent in the pre-policy period).

Smoking and alcohol consumption during pregnancy are associated with health problems for the fetus, including growth problems and damage to the nervous system. In our sample, we exploit two dummy variables for smoking and alcohol consumption. The variable "Smoke in Pregnancy" is equal to one if the woman declares during pregnancy that she smokes at least one cigarette per day. In contrast, the variable "Alcohol in Pregnancy" is equal to one if the woman acknowledges consuming at least 12 g of alcohol per day (about one glass of wine per day) during her pregnancy. Around 7.4 percent of the sample smoked during pregnancy (7.3 percent in the pre-policy period and 7.6 percent in the post-policy), and 4 percent consumed alcohol (5.4 percent in the pre-policy period and 2.8 percent in the post-policy period).

A deficiency of folic acid (a type of B vitamin) during pregnancy is related to fetal malformation. Therefore, folic acid supplements are recommended for all pregnant women and women who plan to become pregnant. We consider a binary variable "Folic Acid in Pregnancy" equal to one if the prospective mother used folic acid during her pregnancy: 84 percent of women, on average, used folic acid during pregnancy, with a larger share in the post-policy period (86 percent versus 81 percent in the pre-policy period).

A dichotomous variable for hospital admissions during pregnancy is defined as equal to one if the woman experienced one or more hospital admissions during pregnancy. Maternal hospitalization during pregnancy may be associated with several pregnancy complications and serious (and potentially deadly) conditions at childbirth (Kim et al. 2021). Around 4 percent of women were hospitalized at least once during the pregnancy. The average share is larger in the post-policy period (3.9 percent pre-policy versus 3.7 percent post-policy).

Birth weight is a widely accepted measure of newborn well-being. Low birth weight is associated with a higher probability of mortality and morbidity, both in the short and long run (e.g., Almond et al. 2005). We follow the World Health Organization's definition of "Low Weight" as a binary variable, equal to one if the birth weight was below 2.5 kg (roughly 5.5 pounds). On average, 5.6 percent of newborns had low birth weight. Similarly, the newborns' length and head circumference are related to their well-being. The average length was around 49.5 centimeters (19.5 inches), while the average head circumference was 34 centimeters (13.4 inches).

The Apgar score is a standard measure of the physical condition of a newborn. When measured one minute after birth, the Apgar score assesses how well the infant tolerated the birthing process. In contrast, the Apgar score measured five minutes after birth assesses how well the infant adapts to life

outside the womb. The score ranges from 0 (no vitality) to 10 (high vitality) and is measured by doctors and nurses by points for heart rate, respiratory effort, muscle tone, response to stimulation, and skin coloration. Apgar scores are predictive of health status, cognitive ability, and children's behavioral problems (Almond et al. 2005; Figlio et al. 2014). In our sample, we define two "Low Apgar Score" indicators, at one and five minutes, as binary variables equal to one if the Apgar score is below 9. In our sample, 20 percent and 5.5 percent of newborns had low Apgar scores after one minute and after five minutes, respectively.

The binary resuscitation indicator is equal to one if a resuscitation method (drugs, intubation, cardiac massage, or oxygen at birth) was necessary after birth. In our sample, a resuscitation method was used in 3.5 percent of deliveries.

The preterm binary indicator is equal to one if the birth occurred before the 37th gestational week. Premature birth is a strong predictor of mortality and morbidity and is associated with poor health (Borra et al. 2016). We observe preterm birth in 5.4 percent of newborns. Stillbirth (death of the infant immediately before or during delivery) occurs in 0.3 percent of deliveries.

Finally, we include the newborn sex among the outcomes. One potential indirect consequence of increasing the take-up of screening tests is the increase in the number of voluntary terminations of pregnancy (VTP) that may follow from increased diagnosis of genetic disorders. Unfortunately, while we are able to check for changes in VTP in the second trimester of pregnancy (see discussion in Section 6.6), we do not have information about VTP specifically following positive prenatal testing. However, there is some evidence that selective VTP may affect sex ratios at birth (Balhotra and Cochrane 2010; Almond and Edlund 2008; Almond et al. 2013); hence, we consider newborn sex as indirect evidence on the potential increase in VTP. In our sample, about 51.5 percent of all newborns are male. This share slightly increases in the post-policy period (51.3 percent in the pre-policy period versus 51.7 percent in the post-policy period).

In all specifications, we include the following information about the woman: age, education level, employment status, the employment status of her partner, marital status, whether she had a twin pregnancy, previous miscarriages and abortions, residence, and nationality.

Table A1 in the Appendix provides a detailed description of all of the characteristics of the women we consider, while Table A2 shows some summary statistics. Around one-third of women are between 30–34 years old when they conceive, 17 percent are younger than age 25, 28 percent are 25– 29, and the remaining 21 percent are older than 35. While 50 percent of women have a "medium" level of education (high school), 28 percent of the sample have a "low" level of education (compulsory education or less), and the remaining 22 percent have a "high" level of education

(tertiary education or more). In the post-policy period, we observe a higher rate of women with high education (23 percent in the post-policy period versus 21 percent in the pre-policy period) and a lower percentage of women with low education (27 percent in the post-policy period versus 29 percent in the pre-policy period). Around 71 percent of women are employed, and 60 percent are married. More than three-quarters of women are Italian natives. A working father is observed in 91 percent of cases, while twin pregnancies occur in 1 percent of the sampled women. The proportion of women who experienced past miscarriages and abortions is about 12 percent and 8 percent, respectively. Women living in metropolitan areas[18] represent around 40 percent of the sample. The proportion of women living in metropolitan areas slightly increases in the post-policy period (41 percent post-policy versus 38 percent pre-policy).

Concerns about inequalities in access to prenatal tests were one of the reasons for eliminating the co-payment. Table 3 shows descriptive evidence about the proportion of women who did not undergo any prenatal tests across several demographic and socioeconomic traits of mothers.

For all age groups, the proportion of women who do not undergo any test decreases by about 8 to 9 percentage points after the policy change. The ratio of prenatal tests increases with age, consistent with medical guidelines promoting these tests for older mothers, given their increased risk of giving birth to babies with congenital disabilities.

The proportion of women with low and medium education who do not undergo any test decreases by about 9 percentage points in the post-policy period. For women with high levels of education, the proportion decreases by 8 percentage points.

The proportions of native and non-native women not undergoing any test decrease by about 8 and 9 percentage points, respectively. Finally, the reduction is larger for women living in metropolitan areas relative to nonmetropolitan areas.

5 Estimation Strategy

Our estimation strategy is twofold. First, we test the effects of eliminating the co-payment on the take-up of prenatal tests and the extent to which these effects differ by socioeconomic status. Second, we investigate whether the variation in the utilization of prenatal tests affects mothers' health behaviors during pregnancy (weight gain, smoking, alcohol consumption, folic acid supplements, and

[18] Metropolitan areas are defined by the metropolitan area of Torino—the regional capital—and of the other seven largest towns of the region (provincial capitals). Each metropolitan area includes the municipality of a medium-large town and the neighboring municipalities, which depend on the main town for most services (healthcare, education, etc.). This may cause difficulties in accessing the services due to user congestion and possible deficiencies in the connections between the center and the periphery (e.g., poor public transit systems), especially for those citizens living in the suburbs.

hospitalization) and newborn health outcomes (birth weight, body length, head circumference, Apgar scores, resuscitation, preterm birth, stillbirth, and sex).

5.1 Prenatal tests take-up rates

To analyze the effect of eliminating the co-payment on prenatal tests utilization, we adopt an RDD approach. We use time, measured in weeks, as the forcing variable (see Hausman and Rapson, 2018, for a discussion of related estimators). In this study, the time threshold determines eligibility for the elimination of the co-payment for prenatal screening tests. The eligibility rule has three key features. First, women undergoing a prenatal screening test after the threshold date (the policy change) are automatically eligible, while those tested before the threshold date are ineligible. Second, the co-payment was eliminated shortly before the cutoff date. Some very general recommendations were introduced in the regional 'birth and pregnancy care plan' in May 2008. However, the detailed regulations were outlined in August 2009, which almost coincided with the new policy's start (i.e., co-payment removal starting from October 1, 2009). Thus, the timing of the regulatory provisions limits the strategic behavior of women close to the policy change. Third, we additionally rule out that women strategically delayed pregnancy to avoid co-payment. Having a child involves consideration of many other factors, and the price of prenatal tests (€27–€54) represents a negligible portion of the cost of raising a child. It follows that the pregnancy decision is exogenous to the threshold date that generates a discontinuity in the treatment assignment. We discuss issues related to the actual date of the prenatal tests in Section 6.4.

Our sample includes women at the date of conception. If the conception date falls in the post-policy period (after October 1, 2009), the woman is automatically enrolled in the program, and she can obtain a screening test for free. The time threshold determines which eligibility rule applies. Since we plausibly exclude strategic behavior (women do not delay/anticipate pregnancy as a consequence of the co-payment elimination), our identification is based on asymptotics in N, the number of cross-sectional units around the threshold, rather than asymptotics in T, the number of data points in time. This allows us to exploit the standard cross-sectional RD strategy (Hausman and Rapson 2018). We are dealing with a health policy targeted at the population level, so we lack a clean control group. Hence, in principle, both interrupted time series (ITS) and the event study approach would not be ill-suited. However, unlike an ITS, our main unit of analysis is the individual woman, observed within a cross-section, without a "classical" time frequency or panel structure. Thus, the main identification assumption is that women only differ in eligibility before and after the policy change. Moreover, to further verify the validity of the RD estimates, we follow the suggestions in Hausman and Rapson (2018) and perform several robustness checks. We conduct falsification tests that use placebo dates,

we estimate alternative local linear bandwidths, and test donut RD specifications (see also Barreca et al. 2011).

There are two main strategies for specifying the functional form to estimate the magnitude of the discontinuity in the outcome at the cutoff point within a classical RD setting: the parametric approach and the nonparametric approach (Imbens and Lemieux 2008; Lee and Lemieux 2010). The main difference is how data around the cutoff are used. While the parametric approach focuses on the optimal functional form to fit the full set of data, most nonparametric approaches search for the optimal data bandwidth where a linear regression function can produce a consistent estimate.

We first apply a parametric approach, and we fit flexible parametric functions to data within a one-year interval around the cutoff point (conception dates 52 weeks before and 52 weeks after the policy implementation). As a robustness check, we also implement some local linear and polynomial nonparametric specifications (Calonico et al. 2014, 2015, 2018).

The parametric specification is as follows:

$$(1) \qquad Y_i = \alpha_0 + \delta A_i + \sum_{k=1}^{K} \alpha_k (C_i - C^*)^k + \sum_{k=1}^{K} \lambda_k A_i \times (C_i - C^*)^k + X_i'\theta + \varepsilon_i,$$

where K can take the following values $K = 1, 2, 3$. We define the dependent variable Y_i as a binary indicator identifying, in turn, each of the four possible utilization rates classified in our conceptual framework in Section 3. First, the dependent variable Y_i is equal to one if woman i undergoes no test (P^n), and zero otherwise. Then, we re-estimate Equation (1) by using as the dependent variable Y_i , which is equal to one if woman i undergoes only the noninvasive subsidized screening test (ZP^s), and zero otherwise. To assess the presence of substitution effects, we also estimate Equation (1) after replacing Y_i with a variable equal to one if woman i undergoes only the invasive diagnostic test (P^d), and zero otherwise. Finally, Y_i equals one if woman i undergoes both types of test ($(1 - Z)P^s$), and zero otherwise.

C_i is the (presumed) conception date for woman i, while C^* is the cutoff date. The difference $(C_i - C^*)$ measures the number of weeks between the conception date and the cutoff date. A_i is the treatment assignment dummy variable, which is the primary variable of interest: it takes a value of one in the post-policy period ($A_i = 1$ if $C_i \geq C^*$), and zero otherwise. The coefficient δ is our key coefficient of interest and gives us the change in the likelihood of undergoing the prenatal test at $C_i = C^*$.

X_i is a vector of observable characteristics of woman i such as age, the highest level of education, parents' employment status, marital status, twin pregnancy, previous miscarriages and abortions, area of residence, and nationality. ε_i is the error term.

Throughout the study, we cluster the standard errors at the level of the week of conception, that is, the measurement unit of the conception date C_i, as suggested by Lee and Card (2008) for discrete running variables to account for any correlation within the clusters. We also show results with a more conservative clustering approach (Kolesàr and Rothe 2018), where standard errors are clustered on the mother's district of residence and quarter of conception. The Piedmont territory is divided into 14 districts or Local Health Authorities (LHA, Aziende Sanitarie Locali). Each authority is responsible for providing health services in its assigned geographical area, including implementing the policy under study. Additionally, we account for multiple hypotheses testing, adjusting p-values for the number of hypothesis tests performed, controlling the Type I error rate. We use the Bonferroni adjustment and the False Discovery Rate (FDR) adjustment (Benjamini and Hochberg 1995, and Anderson 2008).

Identification of the model requires no self-selection at the cutoff and no discontinuous differences in the characteristics of women at the cutoff date. We discuss both issues in Section 6.1 below.

Since we can only rely on data from Piedmont, another threat to identification is the presence of secular trends in the data. We adopt two strategies to rule out the possibility that our results are driven by trends in the take-up of screening tests over time. First, in our main analysis, we use a relatively narrow one-year interval around the cutoff point, conception dates 52 weeks before and 52 weeks after the policy implementation (Section 6.2). If it takes time for the effects of the reform to fully accrue, this approach is likely to provide a conservative effect of the policy. We also conduct robustness checks with narrower and broader time windows (from 12 to 104 weeks before and after the cutoff date) in Section 6.4, with narrow windows further reducing concerns about secular trends. Second, we run two sets of falsification tests, considering both the same cutoff date in different years and different cutoff dates in the same year 2009 of the policy change (Section 6.3).

Finally, to address heterogeneity issues, we also estimate Equation (1) for a set of subsamples of mothers according to their demographic and socioeconomic characteristics (age, level of education, nationality, and area of residence).

5.2 Women's and newborn health outcomes

Prenatal care, consisting of a clinical pathway of regular prenatal visits and testing, may affect the health of both the mother-to-be and the newborn in many different ways. Existing literature suggests that lack of prenatal care is associated with poor birth outcomes, such as low birth weight, preterm birth, and infant mortality (Woodhouse et al. 2014; Currie and Rossin-Slater 2015; Corman et al. 2018). Timely and sufficient prenatal care may also improve maternal health during pregnancy, at

and after childbirth, e.g., reducing weight gain and excessive hospitalizations, preventing smoking, and precipitous labor (Conway and Kutinova 2006, Yan 2017). We, therefore, analyze whether the take-up of prenatal genetic tests (as a proxy for prenatal care) has any consequences on health behaviors during pregnancy or newborn health.

Within prenatal care, prenatal genetic tests that detect congenital anomalies are different from other standard medical tests during pregnancy. Prenatal tests for chromosomal disorders are not routine tests, like, for example, blood pressure, Rh factor for blood type, or blood sugar. Women differ in the motivations for taking (or not) these tests and how they value the consequences of these tests (Markens et al. 1999; Creighton et al. 2003; Hall et al. 2006). Some women may perceive the information the test provides as stressful since it can lead to more invasive procedures with a risk of miscarriage. On the contrary, other women may value that information reassuring, even if they have no intention of terminating an affected pregnancy. These motivations are clearly unobserved; however, we exploit eligibility rules to account for potential endogeneity issues arising from self-selection into prenatal screening testing in our empirical analysis.

Our data set includes information related to lifestyle during pregnancy: weight gain, smoking, alcohol consumption, folic acid intake, and hospitalization.[19] We measure the general health of newborns at delivery by birth weight, length, head circumference, Apgar scores, whether the newborn was resuscitated, premature birth, stillbirth, and the newborn sex.

We estimate the following regression model on the sample of women and newborns:

$$HS_i = \beta_0 + \beta_1 Y_i + \beta_2 (C_i - C^*) + \beta_3 A_i \times (C_i - C^*) + X_i' \eta + u_i \,, \tag{2}$$

where the health status HS_i of the infant/mother *i* depends on Y_i, a dummy variable equal to one if the mother had prenatal screening tests and zero otherwise. In the specification, we also include the distance to the cutoff point $(C_i - C^*)$, the treatment assignment dummy A_i interacted with the distance to the cutoff point $(C_i - C^*)$, and the full set of mother characteristics X_i in Equation (1).

We analyze this relationship within a fuzzy RD design (Lee and Lemieux 2010). The utilization of prenatal tests is voluntary, and it is usually based on a woman's (unobserved) preferences and sociocultural background. Self-selection into prenatal care may bias our results upwards if healthier women value more health care and are more likely to use prenatal care, or downwards if women, expecting worse health, demand more prenatal care. We tackle this issue using a two-stage least squares (2SLS) strategy where the instrumental variable for prenatal test utilization is the

[19] Since we do not have information on pregnancies ending without a delivery, we are unable to provide an analysis of the impact of the policy on VTP. However, we discuss below aggregate regional data on VTP during the second trimester from which no effect can be found when the reform kicks in.

discontinuity in the eligibility of free prenatal screening tests triggered by the new policy. The policy that eliminated the co-payment affects the probability of undergoing a prenatal test, but it does not directly affect health outcomes and therefore offers an exogenous variation that we exploit in the 2SLS estimation. Equation (2) is thus estimated via 2SLS, where the first stage regression is:

$$Y_i = \alpha_0 + \delta A_i + \alpha_1(C_i - C^*) + \lambda A_i \times (C_i - C^*) + X_i'\theta + \varepsilon_i,$$

which is Equation (1) with $K = 1$ under the linear specification. The coefficient of interest is β_1 in Equation (2), which measures the causal effect of prenatal test utilization on women's health-related behaviors and newborn health. More specifically, we measure the average causal effect of screening tests on compliers, the subpopulation of women randomly assigned to treatment (the co-payment elimination group) who comply with the assignment by undergoing a screening test.[20] Finally, we investigate heterogeneous effects by estimating Equation (2) for subsamples of mothers based on age, education, nationality, and area of residence.

We adopt a conservative clustering approach throughout the study, and we cluster the standard errors at the level of the mother's district of residence and quarter of conception. We also account for multiple testing, and we adjust the p-values for the number of hypothesis tests performed using the Bonferroni and the False Discovery Rate (FDR) adjustments (Benjamini and Hochberg 1995, and Anderson 2008).

The potential mechanism at work is linked to the increased number of contacts with physicians, obstetricians, or midwives who may directly influence behaviors during pregnancy. For example, increasing the opportunity for clinicians to advise on lifestyle and potentially harmful effects of drinking or smoking that may affect women and newborns' health outcomes (Corman et al. 2018).[21] These additional meetings triggered by prenatal testing may be particularly influential because they occur early during pregnancy when they are potentially more effective in improving health outcomes (Raatikainen et al. 2007, Harris et al. 2012, Metcalfe et al. 2013). Moreover, more contacts can

[20] Our setting is characterized by partial compliance. The population can be divided into three subgroups: (i) compliers— women randomly assigned to treatment, who comply with it by undergoing a prenatal test; (ii) always-takers— women who are not assigned to treatment, who do undergo the prenatal test; or (iii) never-takers—women randomly assigned to treatment, who do not undergo a prenatal test. Our IV strategy allows us to identify the effect of prenatal tests on the group of compliers (see Angrist and Pischke 2009).

[21] For the region under study, the clinical pathway for prenatal screening tests has been defined in 2007 (before the co-payment removal) and it includes the following steps, all involving meetings with healthcare professionals: (1) information about the objectives, the reliability, and the risks of prenatal testing; (2) informed consent; (3) ultrasound scan for nuchal translucency; (4) collection of anamnestic and clinical data that may influence the risk; (5) two blood tests (one at the end of the first trimester, usually at the 11-13 gestational week, and the second one during the second trimester, usually at the 15-18 gestational week) for the concentration of some chemical markers; (6) computation of the personalised risk and delivery of the final test results; (7) depending on the test results, the diagnostic test; (8) depending on the results of the diagnostic tests, psychological support for VTP. Except for stage (1), provided to any pregnant woman, only women giving their consent experience the whole prenatal testing process. A woman can alternatively choose for a diagnostic test only, instead of a screening test in the first place. In this case the process encompasses a smaller number of steps: (1) information about the objectives, the reliability, and the risks of the diagnostic prenatal testing; (2) informed consent; (3) the diagnostic test, consisting of the removal of a sample of fluid from the amniotic sac; (4) depending on the results of the diagnostic tests, psychological support for VTP is available. Zhang et al. (2019) show the pathway structures for prenatal screening and diagnostic tests.

identify additional health needs triggering further contacts and health care that affect health outcomes during the pregnancy.

Findings from the existing literature are mixed. Most studies find that adequate quality and frequency of prenatal care effectively improves perinatal outcomes. Conversely, pregnancies with inadequate antenatal care present poor health outcomes (placental abruption, intrauterine infections, maternal intensive care units admission, preterm birth, low birth weight, low Apgar scores, and fetal or neonatal death), and high risk factors (smoke and alcohol consumption, high body mass index), even if delivery takes place in hospital (e.g., Raatikainen et al. 2007, Cox et al. 2011, Nabhan and Aflaifel 2015, Till et al. 2015). Another strand of literature considers the adverse effects of more frequent medical procedures as they may affect maternal anxiety, emotional distress, fetal attachment, and, eventually, mothers' and infants' health outcomes (Larsson et al. 2009, Biesecker 2019). Finally, some authors find limited or no behavioral impact from screening tests (Carroli et al. 2001, Georgsson Öhman et al. 2004, Harris et al. 2012 for a review). Unfortunately, our data does not contain any other measures of prenatal care, such as the number of prenatal care visits, the delay in initiating prenatal care, prenatal education, and the number and type of other prenatal tests and scans (Currie and Rossin-Slater 2015; Corman et al. 2018). Therefore, we are cautious that the precise underlying behavioral mechanisms between prenatal care and health outcomes remain to be identified.

6 Results

6.1 Statistical tests and descriptive evidence

We present two sets of statistical tests: sorting of women around the cutoff and balancing tests for the comparability of women around the cutoff. First, we test for strategic manipulation of the conception date: if women nonrandomly sort themselves around the cutoff date, for example, by delaying the conception date to obtain free access to the noninvasive screening test, the continuity assumption of average potential outcomes does not hold, and the causal effect is not identified. We perform several statistical tests for the presence of any discontinuities in the density around the cutoff date based on local polynomial density estimation techniques (McCrary 2008; Cattaneo et al. 2018). We do not find any statistically significant evidence of manipulation. We find no discontinuities in the density of women conceiving around the cutoff date, which supports the absence of self-selection or nonrandom sorting of women into control and treatment groups.[22]

[22] However, we are aware that classical statistical tests for the manipulation of the running variable may perform poorly when the running variable is discrete (Frandsen 2017). Figure A1 in the Appendix shows the density distribution (histogram) of women, and there is no evidence of discontinuities in the distribution of women at the cutoff.

Next, we test for changes in the observable characteristics of women around the cutoff date and therefore check the smoothness of the control variables around the policy change date. Table A3 in the Appendix reports the estimated coefficients for the treatment assignment dummy only. None of the characteristics of mothers show any statistically significant discrete jump at the cutoff date. Figure A2 in the Appendix plots the observed characteristics of pregnant women against the conception week relative to the eligibility cutoff date. Also, from visual inspection, we find that observable mothers' characteristics are smoothly distributed around the cutoff date.

Finally, we present graphical evidence for utilization rates. Figure 1 plots the four prenatal tests utilization rates against the assignment variable, that is, the number of weeks separating the conception date from the cutoff date. The y-axis measures the proportion of women undergoing no prenatal tests (top-left panel), a screening test only (top-right panel), a diagnostic test only (bottom-left panel), and both a screening and a diagnostic test (bottom-right panel). The x-axis reports $(C_i - C^*)$, the number of weeks between the conception date for woman i C_i , and the cutoff date C^*. The zero value represents the cutoff date of the policy change. All women whose conception date falls before the cutoff do not access free prenatal screening tests, while those who fall after the cutoff can access the free noninvasive screening test.

If the elimination of the co-payment had an effect, we expect a discontinuous jump in the utilization of prenatal tests at the cutoff date. Figure 1 documents a clear discontinuity in the probability of undergoing prenatal tests as a function of the conception date in two out of four groups. In particular, in the first two top panels of Figure 1, there is a discontinuous change in the utilization rates at the cutoff point: we observe a reduction in the probability of undergoing no prenatal tests after the program implementation and an increase in the take-up rate for noninvasive screening tests. On the other hand, we do not observe any significant change at the cutoff date for the proportion of diagnostic tests and both tests (two bottom panels of Figure 1).

6.2 Effect on prenatal tests take-up

Table 4 presents the RDD estimates for the effect of the elimination of the co-payment on the probability of undergoing prenatal tests. Results are based on the estimation of Equation (1) by OLS on the whole sample of nulliparous women who become pregnant within a time bandwidth of 52 weeks around the policy intervention. The specification is linear and standard errors are clustered at the week of the mother's conception level. We first estimate the effect of the policy change on the decision to undergo a prenatal test or not (outcome "No test," in column (1)). We then estimate the effect on the probability of undergoing a screening test (column (2)), a diagnostic test (column (3)), and both screening and diagnostic tests (column (4)) separately.

The results show that the policy had a positive statistically significant effect on the utilization rate, and this effect is mainly due to screening test take-up rates. The coefficient for the treatment assignment variable A_i is negative and statistically significant in column 1: the probability of undergoing no prenatal tests decreases by 6.6 percentage points after the elimination of the co-payment. We also control for observable characteristics in all specifications. Older women, more educated, employed, and natives show a higher probability of undergoing prenatal tests. Similarly, we find a significantly higher probability of undergoing tests for women whose partner is employed and those who experienced past abortions. Married women and women living in metropolitan areas have a higher likelihood of undergoing no prenatal tests.

In column (2) of Table 4, we find that after eliminating the co-payment, the probability of undergoing a noninvasive screening test increases by 5.5 percentage points. The probability of having an invasive diagnostic test only (column (3) of Table 4), and the probability of undergoing a screening test followed by a diagnostic test (column (4) of Table 4), are not affected by the policy.

We extend our baseline results on several dimensions as robustness checks. First, the main results are confirmed using alternative cluster robust standard errors (at the level of mother residence-quarter of conception, shown in Table 4), or multiple hypotheses testing. Second, we introduce higher-order polynomial parametric specifications. Table A4 in the Appendix shows OLS estimation results for Equation (1) with different nonlinear trends (quadratic and cubic, in columns (1) and (2), respectively). We find that the quadratic specification confirms the main results from the linear model, while there is a loss of precision with the cubic specification.

Third, we estimate a local linear nonparametric specification (column (3) of Table A4). We adopt the method proposed by Calonico et al. (2014). We allow for the selection of the optimal bandwidth by the MSE-optimal bandwidth selector, and the standard errors are cluster robust at the week of conception level. Coefficients are slightly smaller in size than the OLS results in Table 4. We find that after the policy change, the probability of undergoing no test significantly decreases by around 6.5 percentage points, while the take-up rate for screening tests significantly increases by 4.5 percentage points. There is no significant change for women who undergo only the diagnostic test or both tests.

Fourth, we alternatively define the observational unit for the forcing variable as the presumed conception day, replacing the presumed conception week. We have followed the literature in using the week of conception rather than the day of conception, given that the gestational age is estimated and measured in weeks by health practitioners, and even the individual may not know the exact day of conception. We proxy the day of conception based on the gestational age on the day of child birth,

and we replicate our analysis using days instead of weeks. Columns (4) and (5) in Table A4 show the estimation results for the linear parametric and the linear nonparametric specifications, respectively. Estimation results are very similar to those obtained when the forcing variable is defined in weeks.

6.3 Falsification tests

Since data on prenatal tests for other regions are not available, we do not have a good counterfactual for the change in the take-up of screening tests that would have been observed in the absence of the policy change. Therefore, to rule out the possibility that a secular trend in testing behavior drives our results, we run two sets of falsification tests (Carr and Packham 2021), considering the same cutoff date in different years and different cutoff dates in the same year 2009 of the policy change.

First, we replicate the analysis in equation (1) but rather than using October 1, 2009, as the policy cutoff date, we move backward and forward the cutoff date to previous years (between 2005 and 2012), namely October 1, 2008, October 1, 2007, and October 1, 2006, and to the following years, October 1, 2010, and October 1, 2011. We maintain the same length of the time window around the cutoff date. Figure 2 shows the results: we do not observe any significant effect on prenatal test take-up rates around the artificial policy cutoff dates. Similarly, Figure A3 in the Appendix shows the prenatal tests take-up rates around the policy change using a wider time window, spanning from the first week of October 2005 to the first week of May 2012. In line with our main findings, we find a discontinuous change in the utilization rates only at the actual policy cutoff point, which is significant for no prenatal and noninvasive screening tests.

Second, we move the cutoff date across different months in the year 2009. Figure A4 in the Appendix presents the results from these additional falsification tests. We consider a different cutoff date for the policy change in each replication, from January 1, 2009, to September 1, 2009. The plots show the point estimates and the 95 percent confidence intervals. We do not find any evidence of significant discontinuity before the new policy was implemented. Overall, these results are reassuring about the absence of a secular trend driving our main findings.

6.4 Additional robustness

First, we assess the robustness to different bandwidth choices around the cutoff date. Figure A5 in the Appendix shows the main results for fifteen alternative time windows around the policy change, ranging from 12 to 104 weeks around the cutoff date. The main results are confirmed: the probability of undergoing no test significantly decreases, while the probability of screening tests increases after the co-payment removal. In general, the effects on diagnostic tests do not generally change, or, for larger bandwidths, the effects increase though they are marginally statistically significant. Results are

rather imprecise only for smaller time windows (less than 20 weeks) around the cutoff date when the number of observations sharply decreases.

Second, we address the role of the actual date of the prenatal test. Our main analysis has used the full sample of women whose presumed conception date falls in the 52 weeks before and after the cutoff date. However, we do not observe the exact date when the women underwent the prenatal test. Medical guidelines establish that screening tests are delivered between the 11th and the 13th gestational week, and it may be that some of the women whose conception date falls in the pre-policy period actually paid the co-payment, while some others did not.[23] This may raise possible concerns with our baseline specification, as anticipation effects and measurement errors around the cutoff date may occur. As a robustness check, we then replicate the analysis by excluding women in the weeks before the cutoff date, and we safely exclude pregnancies whose dates were borderline for the co-payment elimination policy ("donut RDD," see Barreca et al. 2011).

Figure 3 shows some replications of Equation (1), where we alternatively exclude women whose conception date falls from 1 to 15 weeks before the policy cutoff date October 1, 2009. Estimation results are robust to different specifications, with marginally larger coefficients.

In Table A5, we also present three more parsimonious strategies: we exclude women who are between the 9th and the 13th gestational week, or between the 9th and the 14th gestational week, or between the 10th and the 14th gestational week, at the policy change date. These latter strategies should reflect the recommended timing of prenatal screening tests more accurately as we exclude observations that may be problematic for co-payment elimination. At the same time, we redefine the post-policy period as it now encompasses all women whose testing date (i.e., their 11th-13th gestational week) falls after October 1, 2009. Estimation results from Table A5 are smaller in magnitude than those from our baseline specification. We also find a marginally significant increase in the probability of a diagnostic test.[24]

We check our data for balance around the policy change, according to the new sample inclusion criteria. We find no self-selection at the cutoff and no discontinuous differences in the characteristics of women at the cutoff date when applying the different donut RDD strategies. We thus rule out any possible strategic behavior by women who may delay their testing date to avoid the co-payment.

[23] The medical guidelines refer to the integrated prenatal screening test, which represents the most common screening test in Piedmont. For illustrative purposes, women whose conception date falls in the last week of July should undergo the integrated prenatal test between the fourth week of September and the third week of October (which correspond to their 11-13th gestational week, see also section 4.1).

[24] Additionally, in Table A6, we replicate our analysis considering as treated all women who conceived during the 13, 12, 11, and 10 weeks before October 1, 2009. Coefficients for the assignment variable are very small in magnitude and imprecisely estimated. The presence of a mix of women, some who pay the co-payment and some who don't in the weeks before October 1, 2009, is the likely cause of these results.

One additional limitation of our database is that we only observe pregnancies conditional on births. Our results represent thus a lower bound of the policy impact because the policy could have generated pregnancy terminations that are not observed. We provide back-of-the-envelope calculations on the magnitude of this bias using numbers of first trimester pregnancy terminations from other sources (EUROCAT 2010, and Biggio et al. 2004). The surveillance program EUROCAT (2010) registers a prevalence rate of chromosomal disorders (encompassing trisomy 13, trisomy 18, and trisomy 21 that the prenatal tests can detect) in the whole population of around 2 cases per 1,000 births. We consider a hypothetical cohort of 10,000 pregnant women, and we assume that the take-up rate of prenatal screening tests is 70%, which coincides with the average take-up rate for screening tests and both tests in our sample. Sensitivity (the proportion of affected fetuses who are correctly identified as such) of prenatal screening tests is 85-90%, while specificity (the proportion of unaffected fetuses who are correctly identified as such) ranges between 95-99% (Biggio et al. 2004).[25] It follows that in the base case scenario, there are 20 expected cases in the whole population and 12-13 detected cases by screening tests. The proportions of women who consent to an invasive prenatal diagnostic procedure after a positive screening test and those terminating a pregnancy after a diagnosed disorder vary considerably. We use average rates suggested in Biggio et al. (2004): around 70% of women accept a diagnostic test after a positive screening test (and we assume that none would do after a negative screening test), while the elect termination rate for the fetal anomaly is around 90%. It follows that the number of terminations of pregnancy for fetal anomaly following prenatal diagnosis is about 8. These terminations of pregnancies, which represent less than 1% of the hypothetical population of 10,000 women, are unobserved in our sample that considers only pregnancies conditional on births. We expect the bias from these unobserved women to be minimal, as it would only marginally change our main estimation results on the take-up of screening tests, which still may be considered as a lower bound of the actual effects.

Overall, our evidence points to an increase in the take-up rate of screening tests following the elimination of the co-payment: the probability that a woman undergoes any prenatal tests increases by 6.6 percentage points. This result is driven by the increase in the take-up of screening tests (by 5.5 percentage points). Thus, the policy increased the take-up rate for screening tests by 8.7 percent, evaluated at the average pre-policy rate of women taking up only screening tests, 63.2 percent. However, as eliminating the co-payment did not affect the utilization rate for the riskier and costlier diagnostic tests, we do not find any substitution effects. Moreover, since it takes time for the effects of the policy to fully accrue, the RDD is likely an underestimate of the full effect.

[25] We should also take into account the spontaneous loss rate of affected pregnancies, which is around 25%, as well as the fetal loss rate after amniocentesis, around 0.9%. This would further reduce the number of expected cases, detected cases and terminations of pregnancy in the whole population.

6.5 Heterogeneity

This section explores whether the policy had more pronounced effects on prenatal test take-up among specific subgroups which differ by age, education, nationality, and residence.

We split the sample of mothers into four age groups, at conception: 18–24, 25–29, 30–34, or over 35 years. Women who are 35 years old or older at conception have access to invasive diagnostic prenatal tests free of charge because they are considered at high risk of congenital disorders. The elimination of the co-payment for this age group was introduced at the national level in 2001, which was well before the period we study (2008–2010). After eliminating the co-payment for screening tests, women older than 35 had free access to any prenatal screening or diagnostic tests. We are particularly interested in testing whether this group of older women changed their testing behavior after the policy change, given the higher probability of a substitution effect between the invasive and the noninvasive tests as they both became fully subsidized.

Table 5 shows the results for the OLS estimation of Equation (1) on different subsamples of women according to their age at conception. Standard errors are clustered at the level of the district of residence–quarter of conception. The odd-numbered columns show the linear specifications, and the even numbered columns the quadratic ones.

In columns (1) and (2) of Table 5, we find that the probability that a woman does not undergo any test decreases by 5 to 8 percentage points if she is younger than 35 at conception. Columns (3) and (4) show that this result is coupled with a higher probability of screening tests, which increases by 5 to 7 percentage points. No statistically significant change occurred to the probability of undergoing diagnostic tests or both tests. For the 18–24 age group, the sign and the magnitude of the coefficients are similar to the 25–29 and 30–34 age groups, but the standard errors are larger, possibly due to smaller sample sizes. The quadratic specifications (even columns of Table 5) lose precision for all age groups.

We find that for women who are 35 years old or older at conception, the probability of undergoing any prenatal test increases by 8 to 9 percentage points (columns (1) and (2) of Table 5). However, unlike other age groups, this effect is not completely determined by the increase in the noninvasive screening tests. Both the take-up rates of screening tests (columns (3) and (4) of Table 5) and diagnostic tests (columns (5) and (6) of Table 5) increase after the policy implementation, even if neither of the effects is statistically significant. A possible rationalization of this result is a behavioral response to the subsidized screening test by women who also have free access to subsidized diagnostic tests (Cohen et al. 2015). After the elimination of the co-payment, women may be more likely to collect additional information about prenatal tests in general. Thus, women with

higher risks, for whom all prenatal tests are free, choose the conclusive diagnostic invasive test more frequently.

To further explore the effect of the policy on the testing behavior of women over 35, we consider a smaller subsample defined by those who are 32–38 years old at conception. As a preliminary step, we also check whether there is any discontinuity in the age distribution around the age of 35 threshold. We perform some statistical manipulation tests (Cattaneo et al. 2018) and find no statistical evidence of systematic discontinuity in the density of mothers' age at conception.[26] At the bottom of Table 5, we present estimation results for Equation (1), where the treatment assignment dummy variable A_i enters both linearly and interacted with the binary indicator *Age Group 35+* (see Lalive 2008 for a similar exercise). The main finding is that prenatal test take-up increases after the policy; and this is mainly explained by a higher probability of undergoing a screening test (columns (1)–(4) of Table 5, coefficients for the *Treatment Assignment* A_i variable). We do not observe any statistically significant difference in the policy effects for women older than 35, as the interaction term is never statistically significant in all columns of Table 5. Finally, we observe that when a woman turns 35, the probability of a screening test decreases, while the probability of a diagnostic test increases (by 25–27 percentage points, columns (5)–(6) of Table 5, the coefficient for the binary variable *Age Group 35+*). We do not find evidence of substitution effects as eliminating the co-payment did not affect the utilization of the riskier diagnostic tests.

In Table 6, we further investigate whether women with different educational backgrounds respond differently to the elimination of the co-payment. We find that the probability of women with a low level of education undergoing prenatal tests increases by 5 to 6 percentage points (columns (1)–(2) of Table 6). However, the coefficients are not precisely estimated, likely due to the small sample. For women with a medium level of education, we find a significant increase of 7 to 9 percentage points. For women with high levels of education, we find an increase by 5 to 6 percentage points (although only the linear specification is statistically significant). We find a statistically significant increase in the take-up rate of noninvasive screening tests for both the medium- and high-education groups (about 6 percentage points). The increase in the probability of undergoing a screening test is about 5 percentage points, and imprecisely estimated, for the low-education group.

Table 6 also considers heterogeneity in women's responses according to nationality and area of residence. From columns 3-4 of Table 6, we find that non-native women have a significantly higher response rate for noninvasive screening tests (about 8 percentage points) than Italian natives (5 percentage points), even if quadratic specifications are less precisely estimated. Similarly, women in

[26] Figure A1, Panel B in the Appendix also shows the density distribution (histogram) of women's ages around the cutoff.

metropolitan areas have a significantly higher response rate for screening tests, by 12 to 14 percentage points, while there is no significant response by women in nonmetropolitan areas. These results are important, as the elimination of the co-payment was justified on *equity* grounds in addition to *efficiency* grounds. The larger response for non-native women and those living in a metropolitan area (characterized by lower utilization rates pre-policy, see Table 3) is consistent with the intentions of the policy.

6.6 Effects on women's and newborn health

As a final step, we estimate Equation (2) to analyze the effects of prenatal tests on women's and newborn health outcomes. Given that the co-payment reform significantly affected screening tests' take-up, we focus our analysis on the impact of this test on health outcomes.

Given the endogenous nature of the utilization decision, we adopt a 2SLS approach, where the "Screening Test" dummy is instrumented by the policy threshold considering a fuzzy RDD specification, as described in Section 5.2. We first report simple (but biased) OLS estimates of Equation (2) for mothers' and newborns' health outcomes at the bottom of Tables 7 and 8, respectively. Results suggest significant positive effects on folic acid supplements and significant negative effects on alcohol consumption during pregnancy when we do not consider self-selection into prenatal care. We also find significant negative effects of the screening test on almost all newborn health outcomes (except for newborn length, stillbirth, and sex). Reduced form estimates, reported in Table A7 in the Appendix, show that only weight gain, hospital admissions, and newborn length significantly reduce after the policy change. Figures A6 and A8 in the Appendix provide a visual interpretation of these results. Figures A7 and A9 show the same mothers' and newborns' health outcomes over a wider time interval.[27] In line with the evidence in Table A7, most of the outcomes appear to be pretty much stable over this longer period. On the contrary, we observe a much clearer drop for alcohol consumption around the threshold, and a clear upward trend during the period for folic acid consumption, that becomes more stable around the threshold. As for alcohol consumption, the wider perspective shed light on what seems to be a significant drop around the threshold which might be the result of additional counselling following the increase in the take up rate of the screening test.[28]

[27] We are forced to exclude hospital admissions from this exercise since data are not available.

[28] In a sample of 228 Italian women at their 38th gestational week in 2010, 35% declared that they had not received any information on the risks associated with alcohol consumption; only 27% declared to have received complete and consistent information (Battistella et al. 2010). According to regional official reports, the need for correct information is still high in the period under study (Regione Piemonte, 2018), and the additional meetings with the medical staff implied by the screening test have certainly helped in this direction.

Our key 2SLS results on women and newborn health outcomes are reported in Tables 7 and 8. As we already know from the estimation results of Equation (1) in previous sections, the policy threshold is a strong predictor of the take-up decision, and it is exogenous to health outcomes. The *F*-test on the excluded instrument in the first stage (reported at the bottom of Tables 7 and 8) is always larger than 30. As significant coefficients may emerge simply by chance when the number of outcomes increases (in our case, 6 outcomes for mothers and 9 for newborns), we compute both the Bonferroni corrected *p*-values and the sharpened False Discovery Rate *q*-values (Benjamini and Hochberg 1995, Anderson, 2008) for the 'Screening Test' coefficient to account for multiple inferences. Both sets of values are reported in Tables 7 and 8.

Our findings suggest only small effects on mothers' health outcomes and no effects on newborn outcomes. After accounting for the conservative multiple hypotheses testing adjustment, we find that screening tests reduce mothers' weight gain during pregnancy and the probability of hospitalization during pregnancy (columns (1) and (6) of Table 7) only at a 10 percent significance level. An increase in the screening test rate by one percentage point reduces the weight gain by 0.065 kilos, the probability of a weight gain above 15 kg by 0.445 percentage points, and the number of hospitalizations by 0.007. Table 8 does not show any causal effect of the prenatal screening test take-up decision on any newborn outcomes at any statistically significant level.

From these results, we draw two conclusions. First, screening tests are safe for the mother and the child, and, unlike invasive prenatal tests, they do not have any direct adverse impacts (e.g., Garrouste et al. 2011, Shurtz et al. 2016, Plachinski 2017). Second, we find that the impact of screening tests on mothers' weight gain and hospitalization during pregnancy are only statistically significant at the 10% level (and likely loose further significance looking over a longer time span). These two measures of mothers' health are usually not directly associated with newborn health outcomes. Excessive weight gain during pregnancy is mainly associated with poor maternal health outcomes, such as hypertension, gestational diabetes, miscarriage (Currie et al. 2010). Maternal prenatal hospitalization may be associated with complications during pregnancy and at delivery (Kim et al. 2021), which may differ depending on the leading causes of the hospitalization and on the pregnancy trimester the hospital admission occurs (that we do not observe).

We assess the robustness of the results on the subsamples of women based on education, nationality, and residence (Figures A10 and A11 in the Appendix). We do not find any significant effect of the prenatal screening tests on mothers' and newborn outcomes in any subsample once we account for multiple testing.

Overall, we do not find any evidence that prenatal screening tests affect newborn health outcomes. In contrast, we find a marginally statistically significant effect of prenatal screening tests on mothers' health outcomes: a reduction in weight gain during pregnancy and prenatal hospitalization. The possible behavioral mechanism through which prenatal screening tests may affect mothers' health status is the increased number of contacts between the mother-to-be and the healthcare providers. These additional encounters can promote healthy behaviors during pregnancy, especially because they occur early when prenatal screening tests are offered (Corman et al. 2018).

Finally, one important indirect outcome for women's and newborn health is whether the increase in the take-up of screening tests also affected the number of voluntary terminations of pregnancy (VTP). In Italy, voluntary termination of pregnancy is allowed at the woman's request within gestational limits: it is available in the first trimester free of charge; it is allowed in the second trimester conditional on saving a woman's life and protecting her physical and/or mental health; it is prohibited in the third trimester. As prenatal tests are taken in the first trimester, VTP occurring during the second trimester is more directly linked to test outcomes. The literature finds evidence of a decline in the likelihood of terminating a pregnancy after a genetic disorder diagnosis over the years. Natoli et al. (2012) notice that termination rates following a prenatal diagnosis of Down syndrome have decreased in recent years in the United States. Looking at data from Scotland, Jacobs et al. (2016) observe that advances in prenatal screening have improved detection rates for aneuploidy (including trisomy), and a reduction in termination rates has accompanied this. Unfortunately, our data do not allow a formal econometric test of this issue; but we can provide three arguments suggesting that VTP did not increase following the reform. First, the reform targeted noninvasive screening tests. These tests—when positive—require taking invasive diagnostic tests to identify chromosomal anomalies, eventually influencing the choice of a VTP. However, we do not find evidence of an impact of the co-payment reform on diagnostic tests.

Second, the share of all VTP occurring during the second trimester of pregnancy on the total number of births did not increase after the reform. Figure 4 shows the fit of a linear regression model, separately estimated on the two sides of the cutoff point (the policy change, in October 2009), on both the number of VTP and the abortion ratio (defined as the ratio between the number of VTP to the total number of births). The exercise is based on monthly data on the number of all VTP that occurred during the second trimester of pregnancy, within the administrative borders of Piedmont, between July 2008 and October 2010. The source of data is the Italian Statistical Office (Istat). We are unable to find an impact of the reform looking at both panels in Figure 4. Last, results in Table 8 (column (9)) rule out any significant change in the share of male births following the reform, which we interpret as indirect evidence of the absence of any increase in selective VTP following a prenatal

test. When looking at subsamples of women based on education, nationality, and residence, this finding also holds further reassuring about the absence of any effect on VTP.

7 Conclusions

Using an RDD framework, this study evaluates the effects of a 2009 policy that eliminated the co-payment for noninvasive prenatal screening tests in a large Italian region. It provides four key findings. First, eliminating the co-payment triggered an economically and statistically significant increase in the take-up rates of prenatal tests by 5.5 percentage points.

Second, we do not find any substitution effect with the more expensive and invasive diagnostic tests. The absence of substitution effects is mostly explained by the group of women older than 35 years at conception, who are at higher risk of congenital disorders, and who are thus more likely to undergo an invasive diagnostic test. At the time of the new policy, this group was already exempt from the co-payment for the invasive test and is therefore likely to be less sensitive to price changes.

Third, we find that the effect of the policy is larger for younger women, non-natives, those residing in metropolitan areas, and those with medium levels of education. Hence, the policy change produced the expected effects of increasing the take-up for more disadvantaged groups that had lower screening test rates before the co-payment was eliminated.

Finally, we do not find any important effect on health outcomes. For mothers' health behaviors, we find a reduction in weight gain and hospital admissions during pregnancy, which is only marginally statistically significant. None of the effects on the newborn health outcomes is statistically significant.

In terms of policy implications, our analysis suggests that eliminating co-payments can be an effective policy lever to encourage the take-up of noninvasive screening tests, which is a cost-effective method of identifying pregnancies at high risk of a genetic disorder.[29] This helps to address both *efficiency* concerns due to relevant market failures, such as the underestimation of benefits associated with prenatal care, and *equity* concerns, as the increase in prenatal care access is higher among more disadvantaged groups. However, since eliminating the co-payment for noninvasive screening tests does not lead to any substitution effect between noninvasive and invasive tests, the reduction in the inappropriate use of invasive tests should be targeted through other policy instruments. For instance, access to invasive diagnostic tests free of charge could be made conditional

[29] The resolution of the Regional Government (Delibera della Giunta Regionale) n. 38-11960, August 4, 2009 established that the annual additional costs for the public administration following the co-payment elimination was €500,000. The annual health expenditure of the Regional Government is around €8 billion.

on a positive test result of a noninvasive screening test rather than on the mother's age alone. New national guidelines, provided in the updated Essential Levels of Care legislation, are consistent with this recommendation.

8 Acknowledgments

The authors thank Douglas Almond, Sonia Bhalotra, Janet Currie, Osea Giuntella, Jo Thori Lind, Elena Lucchese, Gianmaria Martini, Catia Nicodemo, Anne-Fleur Roos, Philipp Steinbrunner, and Francesca Zantomio for their insightful comments on previous drafts. We are grateful to the participants at the 2019 Healthcare Management & Economics Workshop (HEMAW), the 6th Health Econometrics Workshop (HEW), the 2018 Annual Meeting of the European Public Choice Society (EPCS), and the XXX Annual Conference of the Italian Society of Public Economics (SIEP). We also thank Maria Maspoli, Paola Ghiotti, and Gloria Prina at the Assessorato Regionale alla Sanità, Regione Piemonte for giving us access to the CEDAP microdata. The comments of the editor-in-chief Christopher S. Carpenter and three anonymous referees are gratefully acknowledged. All remaining errors are our own.

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

Declarations of interest: none.

9 References

Aguilara, A., E. Gutierreza, and E. Seirab. 2021. The effectiveness of sin food taxes: Evidence from Mexico. *Journal of Health Economics*, Vol. 77, 102455.

Almond, D., K.Y. Chay, and D.S. Lee. 2005. The Costs of Low Birth Weight. *Quarterly Journal of Economics*, Vol. 120(3), 1031–1083.

Almond, D., and J. Currie. 2011. Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, Vol. 25(3), 153–172.

Almond, D., and L. Edlund. 2008. Son Biased Sex Ratios in the U.S. 2000 Census. *Proceedings of the National Academy of Sciences*, Vol. 105(15), 5681–5682.

Almond, D., L. Edlund, and K. Milligan. 2013. Son Preference and the Persistence of Culture: Evidence from South and East Asian Immigrants to Canada. *Population and Development Review*, Vol. 39(1), 75–95.

Anderson, M.L. 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, Vol. 103(484), 1481–1495.

Angrist, J., and J-S Pishke. 2009. Mostly Harmless Econometrics. Princeton University Press.

AReSS (Agenzia Regionale per i Servizi Sanitari). 2008. Tavolo A.Re.S.S. sullo Screening delle Anomalie Fetali Proposta di riorganizzazione del percorso per lo screening delle anomalie fetali. Technical Report.

Barreca, A., M. Guldi, J. Lindo, and G. Waddell. 2011. Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification. *Quarterly Journal of Economics*, Vol. 126(4), 2117–2123.

Battistella, G, S. Bazzo, M. Bottarel, L. Czerwinsky Domenis, G. Moino, P. Riscica. 2010. Il consumo di alcol in gravidanza, Azienda ULSS 9 di Treviso [Alcohol consumption during pregnancy, Local Health Authority of Treviso], Notiziario Istituto Superiore Sanità, vol. 23, ii-iv.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Royal Statistical Society*, Vol. 57(1), 289-300.

Bhalotra, S.R. and T. Cochrane. 2010. Where Have All the Young Girls Gone? Identification of Sex Selection in India. IZA Discussion Paper No. 5381.

Biesecker, B.B. 2019. The Psychological Well-being of Pregnant Women Undergoing Prenatal Testing and Screening: A Narrative Literature Review. *Hastings Center Report*, Vol. 49(3), S53– S60.

Biggio J.R., T.C. Morris, J. Owen, J.S. Stringer. 2004. An outcomes analysis of five prenatal screening strategies for trisomy 21 in women younger than 35 years. *American Journal of Obstetrics and Gynecology*, Vol. 190(3), 721-9.

Bitler, M. P., and C. S. Carpenter. 2016. Health Insurance Mandates, Mammography, and Breast Cancer Diagnoses. *American Economic Journal: Economic Policy,* Vol. 8(3), 39–68.

Bitler, M. P., and C. S. Carpenter. 2017. Effects of State Cervical Cancer Insurance Mandates on Pap Test Rates. *Health Services Research*, Vol. 52(1), 156–175.

Borra, C., L. González, and A. Sevilla. 2016. Birth Timing and Neonatal Health. *American Economic Review*, Vol. 106 (5), 329–332.

Boyd, P., C. DeVigan, B. Khoshnood, M. Loane, E. Garne, and H. Dolk. 2008. Survey of Prenatal Screening Policies in Europe for Structural Malformations and Chromosome Anomalies, and their Impact on Detection and Termination Rates for Neural Tube Defects and Downs Syndrome. *BJOG: An International Journal of Obstetrics & Gynaecology*, Vol. 115(6), 689–696.

Calonico, S., M.D. Cattaneo, and R. Titiunik. 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs, *Econometrica*, Vol. 82(6), 2295–2326.

Calonico, S., M.D. Cattaneo, and R. Titiunik. 2015. Optimal Data-Driven Regression Discontinuity Plots, *Journal of the American Statistical Association*, Vol. 110(512), 1753–1769.

Calonico, S., M.D. Cattaneo, and M.H. Farrell. 2018. On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*, Vol. 113(522), 767–779.

Carr J.B., and A. Packham. 2021. SNAP Schedules and Domestic Violence. *Journal of Policy Analysis and Management*, Vol. 40(2), 412–452.

Carroli G., J. Villar, G. Piaggio, D. Khan-Neelofur, M. Gülmezoglu, M. Mugford, P. Lumbiganon, U. Farnot, and P. Bersgjø P. 2001. WHO Antenatal Care Trial Research Group. WHO systematic review of randomised controlled trials of routine antenatal care. *Lancet*. Vol. 357(9268), 1565-70.

Cattaneo, M.D., M. Jansson, and X. Ma. 2018. Manipulation Testing Based On Density Discontinuity. *Stata Journal*, Vol. 18(1), 234–261.

Cohen, J., P. Dupas, and S. Schaner. 2015. Price Subsidies, Diagnostic Tests, and Targeting of Malaria Treatment: Evidence from a Randomized Controlled Trial. *American Economic Review*, Vol. 105(2), 609–645.

Conway, K.S., and A. Kutinova. 2006. Maternal Health: Does Prenatal Care Make a Difference? *Health Economics*, Vol. 15(5), 461–488.

Corman, H., D.M. Dave, and N. Reichman. 2018. Effects of Prenatal Care on Birth Outcomes: Reconciling a Messy Literature. NBER Working Paper No. 24885.

Cox RG, L. Zhang, M.E. Zotti, J. Graham. 2011. Prenatal care utilization in Mississippi: racial disparities and implications for unfavorable birth outcomes. *Maternal and Child Health Journal*, Vol. 15, 931-942.

Creighton, S., E.W. Almqvist, D. MacGregor, B. Fernandez, H. Hogg, J. Beis, J.P. Welch, C. Riddell, R. Lokkesmoe, M. Khalifa, J. MacKenzie, A. Sajoo, S. Farrell, F. Robert, A. Shugar, A. Summers, W. Meschino, D. Allingham-Hawkins, T. Chiu, A. Hunter, J. Allanson, H. Hare, J. Schween, L. Collins, S. Sanders, C. Greenberg, S. Cardwell, E. Lemire, P. MacLeod, M.R. Hayden. 2003. Predictive, pre-natal and diagnostic genetic testing for Huntington's disease: the experience in Canada from 1987 to 2000. *Clinical genetics*, Vol. 63(6), 462-475.

Crombag N., Y. Vellinga, S. Kluijfhout, L. Bryant, P. Ward, R. Iedema-Kuiper, P. Schielen, J. Bensing, G. Visser, A. Tabor, and J. Hirst. 2014. Explaining Variation in Down's Syndrome Screening Uptake: Comparing the Netherlands with England and Denmark Using Documentary Analysis and Expert Stakeholder Interviews, *BMC Health Services Research*, Vol. 14, 437.

Crombag, N. 2016. Explaining Low Uptake for Down Syndrome Screening in the Netherlands (and Predicting Utilisation of Other Programmes). Mimeo, Utrecht University, the Netherlands.

Currie, J. 2006. The Take-up of Social Benefits. In A. Auerbach, D.Card, and J. Quigley (eds), "*Poverty, the Distribution of Income, and Public Policy*," New York: Russell Sage, 80–148.

Currie, J., S. DellaVigna, E. Moretti, and V. Pathania. 2010. The Effect of Fast Food Restaurants on Obesity and Weight Gain. *American Economic Journal: Economic Policy*, Vol. 2(3), 32–63.

Currie, J., and M. Rossin-Slater. 2015. Early-Life Origins of Life-Cycle Well-Being: Research and Policy Implications. *Journal of Policy Analysis and Management*, Vol. 36(4), 974–974.

Dormandy, E., S. Michie, R. Hooper, and T. Marteau. 2005. Low Uptake of Prenatal Screening for Down Syndrome in Minority Ethnic Groups and Socially Deprived Groups: A Reflection of Women's Attitudes or a Failure to Facilitate Informed Choices? *International Journal of Epidemiology*, Vol. 34, 346–352.

EUROCAT. 2010. Special Report: Prenatal Screening Policies in Europe 2010. EUROCAT (European Surveillance of Congenital Anomalies) Central Registry, University of Ulster.

Fajnzylber, E., V.J. Hotz, and S.G. Sanders. 2010. An Economic Model of Amniocentesis Choice, NBER Working Paper 16306.

Fetter, D. 2013. How Do Mortgage Subsidies Affect Home Ownership? Evidence from the Mid-Century GI Bills. *American Economic Journal: Economic policy*, Vol. 5(2), 111-47.

Figlio, D., J. Guryan, K. Karbownik, and J. Roth. 2014. The Effects of Poor Neonatal Health on Children's Cognitive Development. *American Economic Review*, Vol. 104(12), 3921–3955.

Frandsen, B.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, Advances in Econometrics*," Vol. 38, Emerald Publishing Limited, Bingley, 281-315.

Fransen, M.P., M.H. Schoonen, J.P Mackenbach, E.A. Steegers, H.J. de Koning, J.A. Laudy, R.J. Galjaard, C.W. Looman, M.L. Essink-Bot, and H.I. Wildschut. 2010. Ethnic Differences in Participation in Prenatal Screening for Down syndrome: A Register-Based Study. *Prenatal Diagnosis*, Vol. 30, 988–994.

Gajdos, T., C. Garrouste, and P. Geoffard. 2016. The Subjective Value of a Life with Down Syndrome: Evidence from Amniocentesis Decision. *Journal of Economic Behavior & Organization*, Vol. 127, 59–69.

Garrouste, C., J. Le, and E. Maurin. 2011. The Choice of Detecting Down Syndrome: Does Money Matter? *Health Economics*, Vol. 20, 1073–1089.

Georgsson Öhman, S., S. Saltvedt, C. Grunewald, and U. Waldenström. 2004. Does fetal screening affect women's worries about the health of their baby? *Acta Obstetricia et Gynecologica Scandinavica*, Vol. 83, 634-640

Grimes, D., and K. Schulz. 2002. Uses and Abuses of Screening Tests. *Lancet*, Vol. 359(9309), 881–884.

Hall, J., D.G. Fiebig, M.T. King, I. Hossain, and J.J. Louviere. 2006. What influences participation in genetic carrier testing? Results from a discrete choice experiment. *Journal of Health Economics*, Vol. 25(3), 520-537.

Halla M., G.J. Pruckner, and T. Schober. 2016. Cost Savings of Developmental Screenings: Evidence from a Nationwide Program. *Journal of Health Economics*, Vol. 49, 120–135.

Hansen B., K. Miller, and C. Weber. 2020. Federalism, partial prohibition, and cross-border sales: Evidence from recreational marijuana. *Journal of Public Economics*, Vol. 187, 104-159.

Harris J.M., L. Franck, and S. Michie. 2012. Assessing the psychological effects of prenatal screening tests for maternal and foetal conditions: a systematic review. *Journal of Reproductive and Infant Psychology*, Vol. 30(3), 222-246

Hausman, C. and D. Rapson. 2018. Regression Discontinuity in Time: Considerations for Empirical Applications. *Annual Review of Resource Economics*, Vol. 10, 533–552.

Imbens, G., and T. Lemieux. 2008. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, Vol. 142, 615–635.

Institute of Medicine. 2009. Weight Gain During Pregnancy: Re-examining the Guidelines. Report Brief.

Ito, K. 2015. Asymmetric Incentives in Subsidies: Evidence from a Large-Scale Electricity Rebate Program. *American Economic Journal: Economic Policy*, Vol. 7(3), 209–237.

Jacobs M., S.A. Cooper, R. McGowan, S.M. Nelson, and J.P. Pell JP. 2016. Pregnancy Outcome Following Prenatal Diagnosis of Chromosomal Anomaly: A Record Linkage Study of 26,261 Pregnancies. *PLoS One*, Vol. 11(12), e0166909.

Javaher, P., E. Nyoungui, H. Kääriäinen, U. Kristoffersson, I. Nippert, J. Sequeiros, and J. Schmidtke. 2010. Genetic Screening in Europe. *Public Health Genomics*, Vol. 13, 524–537.

Kim, J., A. Lee, and M. Rossin-Slater. 2021. What to Expect When It Gets Hotter. The Impacts of Prenatal Exposure to Extreme Temperature on Maternal Health. *American Journal of Health Economics*, forthcoming.

Kolesár, M., and C. Rothe. 2018. Inference in Regression Discontinuity Designs with a Discrete Running Variable. *American Economic Review*, Vol. 108(8), 2277-2304.

Lalive R. 2008. How Do Extended Benefits Affect Unemployment Duration? A Regression Discontinuity Approach. *Journal of Econometrics*, Vol. 142, 785–806.

Larsson, A., E.C. Svalenius, K. Maršál, M. Ekelin, P. Nyberg, and A. Dykes. 2009. Parents' Worried State of Mind When Fetal Ultrasound Shows an Unexpected Finding. *Journal of Ultrasound in Medicine*, Vol. 28, 1663-1670.

Lautharte, I. 2021. Babies and Bandidos: Birth outcomes in pacified favelas of Rio de Janeiro. *Journal of Health Economics*, Vol. 77.

Lee D., and D. Card. 2008. Regression Discontinuity Inference with Specification Error. *Journal of Econometrics*, Vol. 142, 655–674.

Lee D., and T. Lemieux. 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2), 281–355.

Li P., Y. Lu, and J. Wang. 2020. The effects of fuel standards on air pollution: Evidence from China. *Journal of Development Economics*. Vol. 146, 102488.

Loane M., J. Morris, M. Addor, L. Arriola, J. Budd, B. Doray, E. Garne, M. Gatt, M. Haeusler, and B. Khoshnood. 2013. Twenty-Year Trends in the Prevalence of Down Syndrome and Other Trisomies in Europe: Impact of Maternal Age and Prenatal Screening. *European Journal of Human Genetics*, Vol. 21(1), 27–33.

Malone, F.D., J. Canick, R.H. Ball, D.A. Nyberg, C.H. Comstock, R. Bukowski, R.L. Berkowitz, S.J. Gross, L. Dugoff, S.D. Craigo, I.E. Timor-Tritsch, and S.R. Carr. 2005. First-Trimester or Second-Trimester Screening, or Both, for Down's Syndrome. *New England Journal of Medicine*, Vol. 353(19), 2001–2011.

Markens, S., C.H Browner, and N. Press. 1999. 'Because of the risks': how US pregnant women account for refusing prenatal screening. *Social Science & Medicine*, Vol. 49(3), 359-369.

McCrary, J. 2008. Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, Vol. 142(2), 698–714.

Metcalfe, A., L.M. Lix, J. Johnson, F. Bernier, G. Currie, A.W. Lyon, and S.C. Tough. 2013. Assessing the Impact of the SOGC Recommendations to Increase Access to Prenatal Screening on Overall Use of Health Resources in Pregnancy. *Journal of Obstetrics and Gynaecology Canada*, Vol. 35 (5), 444–453.

Moulton, J.G., B.D. Waller, and S.A. Wentland. 2018. Who Benefits from Targeted Property Tax Relief? Evidence from Virginia Elections. *Journal of Policy Analysis and Management*, Vol. 37(2), 240–264.

Nabhan A.F., and N. Aflaifel. 2015. High feedback versus low feedback of prenatal ultrasound for reducing maternal anxiety and improving maternal health behaviour in pregnancy. *Cochrane Database of Systematic Reviews*, Issue 8. Art. No.: CD007208.

Natoli, J. L., D.L. Ackerman, S. Mcdermott, S., and J.G. Edwards. 2012. Prenatal Diagnosis of Down Syndrome: A Systematic Review of Termination Rates (1995–2011). *Prenatal Diagnosis*, Vol. 32(2), 142–153.

Ohno M, A. Caughey. 2013. The Role of Noninvasive Prenatal Testing as a Diagnostic Versus a Screening Tool – A Cost-Effectiveness Analysis. *Prenatal Diagnosis*, Vol. 33, 630–635.

Palomaki, G.E., K. Steinort, G.J. Knight, and J.E. Haddow. 2006. Comparing Three Screening Strategies for Combining First- and Second-Trimester Down Syndrome Markers. *Obstetrics & Gynecology*, Vol. 107, 367–375.

Palomaki, G.E., G.J. Knight, E.R. Ashwood, R.G. Best, J.E. Haddow. 2013. Screening for Down Syndrome in the United States: Results of Surveys in 2011 and 2012. *Archives of Pathology & Laboratory Medicine.* Vol. 137, 921–926.

Plachinski, L. 2017. The Effect of Access to Prenatal Genetic Testing on Test Utilization and Birth Outcomes: Evidence from Down Syndrome. Wellesley College, Honors Thesis Collection n. 475.

Quintana-Domeque, C., and P. Ródenas-Serrano. 2017. The hidden costs of terrorism: The effects on health at birth. *Journal of Health Economics*, Vol. 56, 47-60.

Raatikainen, K., N. Heiskanen, and S. Heinonen. 2007. Under-attending free antenatal care is associated with adverse pregnancy outcomes. *BMC Public Health*, Vol. 7, 268.

Regione Piemonte. 2018. Nascere in Piemonte: pecorso nascita regionale. 2° Rapporto sui dati dei Certificati di Assistenza al Parto: Anni 2006-2016. Assessorato della sanità, Direzione Sanità.

Santalahti, P., E. Hemminki, A.-M. Latikka, and M. Ryynänen. 1998. Women's Decision-Making in Prenatal Screening. *Social Science and Medicine*, Vol. 46(8), 1067–1076.

Seror, V. 2008. Fitting Observed and Theoretical Choices--Women's Choices about Prenatal Diagnosis of Down Syndrome. *Health Economics*, Vol. 17(5), 557–577.

Shurtz, I., A. Brzezinski, and A. Frumkinc. 2016. The Impact of Financing of Screening Tests on Utilization and Outcomes: The Case of Amniocentesis. *Journal of Health Economics*, Vol. 48, 61–73.

Till S.R., D. Everetts, and D.M. Haas. 2015. Incentives for increasing prenatal care use by women in order to improve maternal and neonatal outcomes. *Cochrane Database of Systematic Reviews*, Issue 12.

Turati, G. 2013. The Italian Servizio Sanitario Nazionale: A Renewing Tale of Lost Promises. In J. Costa-Font, S.C. Greer (Eds.), *Federalism and Decentralization in European Health and Social Care: Competition, Innovation, and Cohesion*, Palgrave MacMillan, London, 47–66.

Vassy, C., S. Rosman, and B. Rousseau. 2014. From Policy Making to Service Use. Down's Syndrome Antenatal Screening in England, France and the Netherlands. *Social Science & Medicine*, Vol. 106, 67–74.

Verloove, P., R.A. Verwey, R. Brand, and M.J. Keirse. 1986. Importance of Gestational Age. *Lancet,* Vol. 1(8496), 1494.

Wald, N. J., C. Rodeck, A. K. Hackshaw, J. Walters, L. Chitty, and A.M. Mackinson - SURUSS Research Group. 2003. First and Second Trimester Antenatal Screening for Down's Syndrome: The Results of the Serum, Urine and Ultrasound Screening Study (SURUSS). *Health Technology Assessment*, Vol. 7(11), 1–77.

Woodhouse, C., J. Lopez Camelo, and G.L. Wehby. 2014. A Comparative Analysis of Prenatal Care and Fetal Growth in Eight South American Countries. *PLoS One*, Vol. 9(3), e91292.

World Health Organization. 2016. *WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience*. World Health Organization Report.

Yan, J. 2017. The Effects of Prenatal Care Utilization on Maternal Health and Health Behaviors. *Health Economics*, Vol. 26(8), 1001–1018.

Zhang, W., T. Mohammadi, J. Sou, and A.H. Anis. 2019. Cost-effectiveness of prenatal screening and diagnostic strategies for Down syndrome: A microsimulation modeling analysis. *PLoS ONE* Vol. 14(12), e0225281.

Table 1. Number and Proportion of Women Undergoing Different Prenatal Tests

	Whole Sample	Pre-Policy	Post-Policy
No Test	7,410	4,534	2,876
	22.71	27.00	18.16
Screening Test Only	21,893	10,606	11,287
	67.11	63.17	71.28
Diagnostic Test Only	2,309	1,143	1,166
	7.08	6.81	7.36
Both Tests	1,011	505	506
	3.10	3.00	3.20
Total	32,623	16,788	15,835
	100	100	100

Notes: The first row shows *frequencies,* and the second row shows *column percentages*.

Table 2. Summary Statistics of Health Outcomes

Panel A. Mothers' Health Outcomes

	Whole Sample			Pre-Policy		Post-Policy	
	Mean	***SD***	***N***	***Mean***	***SD***	***Mean***	***SD***
Weight Gain in Pregnancy	13.047	4.002	27,594	13.087	4.014	13.006	3.989
Weight Gain in Preg. >15 kg	0.236	0.425	27,594	0.240	0.427	0.233	0.422
Smoking in Pregnancy	0.074	0.262	32,623	0.073	0.260	0.076	0.265
Alcohol consumption in Preg.	0.041	0.199	32,623	0.054	0.226	0.028	0.164
Folic Acid in Pregnancy	0.836	0.370	32,623	0.814	0.389	0.859	0.348
Hospital Admissions in Preg.	0.038	0.191	31,004	0.039	0.195	0.037	0.188

Panel B. Newborn Health Outcomes

	Whole Sample			Pre-Policy		Post-Policy	
	Mean	***SD***	***N***	***Mean***	***SD***	***Mean***	***SD***
Low Weight	0.056	0.230	32,829	0.055	0.228	0.056	0.230
Newborn Length	49.505	2.013	32,094	49.545	1.996	49.462	2.029
Head Circumference	33.983	1.319	32,180	33.996	1.314	33.970	1.323
Low Apgar score –1 min	0.200	0.400	31,937	0.202	0.402	0.199	0.399
Low Apgar score –5 min	0.055	0.228	31,942	0.057	0.231	0.053	0.224
Resuscitation	0.035	0.184	32,830	0.036	0.186	0.034	0.182
Preterm	0.054	0.225	32,830	0.054	0.225	0.054	0.225
Stillbirth	0.003	0.053	32,830	0.003	0.055	0.002	0.050
Male	0.515	0.500	32,811	0.513	0.500	0.517	0.500

Notes: All variables are binary variables, except for "Newborn Length" and "Head Circumference," which are both expressed in centimeters (cm), and "Weight Gain in Pregnancy," which is expressed in kilograms (kg).

Table 3. Percentage of Women Who *Do Not Undergo Any* Prenatal Tests, Within Different Subsamples

	Whole Sample	Pre-Policy	Post-Policy
Age groups:			
Age group 18–24	0.378	0.419	0.335
Age group 25–29	0.233	0.274	0.191
Age group 30–34	0.184	0.224	0.140
Age group 35+	0.170	0.212	0.118
Education levels:			
Low Education	0.306	0.349	0.258
Medium Education	0.205	0.247	0.160
High education	0.176	0.216	0.138
Nationality:			
Non-native	0.379	0.420	0.338
Native	0.180	0.225	0.132
Area of residence:			
Nonmetropolitan area	0.208	0.240	0.172
Metropolitan area	0.257	0.319	0.195

Table 4. Effect of Policy Change on the Prenatal Tests' Take-Up Rates: Baseline Results

Dep. Var.:	*No Tests*	*Screening Test*	*Diagnostic Test*	*Both Tests*
	(1)	(2)	(3)	(4)
Assignment A_i	-0.066***	0.055***	0.008	0.003
(s.e. clustered at week level)	(0.01)	(0.01)	(0.01)	(0.00)
{s.e. clustered at residence-quarter level}	{0.03}	{0.03}	{0.01}	{0.00}
Linear Trend, (C_i - C^*)	-0.001***	0.001**	0.000	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)
Assignment A_i x Linear Trend, (C_i - C^*)	0.001***	-0.001*	-0.001***	0.000
	(0.00)	(0.00)	(0.00)	(0.00)
Age group 25–29	-0.071***	0.076***	-0.005***	-0.000
	(0.01)	(0.01)	(0.00)	(0.00)
Age group 30–34	-0.094***	0.082***	0.003	0.009***
	(0.01)	(0.01)	(0.00)	(0.00)
Age group 35+	-0.099***	-0.238***	0.271***	0.065***
	(0.01)	(0.01)	(0.01)	(0.00)
Medium Education	-0.022***	0.015**	0.003	0.004**
	(0.01)	(0.01)	(0.00)	(0.00)
High education	-0.024***	0.014*	0.004	0.006*
	(0.01)	(0.01)	(0.00)	(0.00)
Employed	-0.073***	0.062***	0.007**	0.004*
	(0.01)	(0.01)	(0.00)	(0.00)
Married	0.034***	-0.012**	-0.018***	-0.004*
	(0.00)	(0.00)	(0.00)	(0.00)
Native	-0.112***	0.092***	0.015***	0.005**
	(0.01)	(0.01)	(0.00)	(0.00)
Father Employed	-0.155***	0.129***	0.020***	0.006*
	(0.01)	(0.01)	(0.00)	(0.00)
Twin	0.017	0.006	-0.010	-0.013
	(0.02)	(0.03)	(0.02)	(0.01)
Past Miscarriage	-0.010	0.001	0.005	0.004
	(0.01)	(0.01)	(0.00)	(0.00)
Past Abortion	-0.027***	0.018*	0.007	0.002
	(0.01)	(0.01)	(0.00)	(0.00)
Metropolitan Area	0.037***	-0.031***	-0.006**	0.000
	(0.01)	(0.01)	(0.00)	(0.00)
Constant	0.580***	0.430***	-0.008	-0.002
	(0.02)	(0.02)	(0.01)	(0.00)
Adj. R^2	0.08	0.10	0.20	0.02
Pre-policy Mean Dep. Var.	0.27	0.63	0.07	0.03
N. Obs.	32,623	32,623	32,623	32,623

Notes: Each column presents the estimation of Equation (1) by OLS, on the sample of women becoming pregnant during the 52 weeks before and the 52 weeks after the policy change. The dependent variables are: No Test (equal to one if the woman had no prenatal tests, and zero otherwise) in column (1); Screening Test (equal to one if a screening test is undertaken, and zero otherwise) in column (2); Diagnostic Test (equal to one if an invasive diagnostic test is undertaken, and zero otherwise) in column (3); and Both Tests (equal to one if the woman had a screening test followed by an invasive diagnostic test, and zero otherwise) in column (4). Assignment A_i is equal to one in the post-policy period. All equations include the term $(C_i - C^*)$, that is, the number of weeks between the conception date C_i, and the cutoff date C^*, and its interaction with the treatment assignment variable. Standard errors (in parentheses) are clustered at the level of the week of conception (104 clusters), while standard errors in brace are clustered at the level of mother's district of residence–quarter of conception. * p<0.1; ** p<0.05; *** p<0.01.

Table 5. Effect of Policy Change on the Prenatal Tests' Take-Up Rates: Age Groups

Dep. Var.:	***No Test***		***Screening Test***		***Diagnostic Test***		***Both Tests***	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification:	*Linear*	*Quadratic*	*Linear*	*Quadratic*	*Linear*	*Quadratic*	*Linear*	*Quadratic*
1. Subsample: Age 18–24 *(N. Obs. 5,393)*								
Assignment A_i	-0.045	-0.074	0.048	0.076	0.003	-0.000	-0.006	-0.001
	(0.05)	(0.06)	(0.05)	(0.06)	(0.00)	(0.01)	(0.00)	(0.01)
Adj. R^2	0.07	0.07	0.06	0.06	0.007	0.007	0.003	0.003
Mean Dep. Var.	0.378		0.607		0.006		0.009	
2. Subsample: Age 25–29 *(N. Obs. 9,221)*								
Assignment A_i	-0.078**	-0.077	0.074**	0.068	0.000	-0.000	0.004	0.009
	(0.04)	(0.05)	(0.04)	(0.05)	(0.00)	(0.00)	(0.00)	(0.01)
Adj. R^2	0.05	0.05	0.04	0.04	0.003	0.003	0.001	0.001
Mean Dep. Var.	0.233		0.746		0.008		0.014	
3. Subsample: Age 30–34 *(N. Obs. 11,115)*								
Assignment A_i	-0.060*	-0.056	0.055*	0.043	0.003	0.012	0.003	0.000
	(0.03)	(0.04)	(0.03)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.06	0.06	0.04	0.04	0.003	0.003	0.001	0.001
Mean Dep. Var.	0.184		0.772		0.019		0.025	
4. Subsample: Age 35+ *(N. Obs. 6,894)*								
Assignment A_i	-0.080**	-0.086*	0.041	0.045	0.031	0.043	0.007	-0.001
	(0.04)	(0.05)	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)
Adj. R^2	0.06	0.06	0.01	0.01	0.02	0.02	0.004	0.004
Mean Dep. Var.	0.171		0.459		0.289		0.082	
5. Subsample: Age 32–38 *(N. Obs. 10,048)*								
Assignment A_i	-0.069**	-0.080	0.067*	0.066	0.006	0.019	-0.004	-0.006
	(0.03)	(0.05)	(0.03)	(0.05)	(0.01)	(0.02)	(0.01)	(0.01)
Age Group 35+	0.005	-0.024	-0.298***	-0.292***	0.249***	0.268***	0.044**	0.048*
	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.03)
Assignment A_i x Age Group 35+	-0.008	0.018	-0.016	0.012	0.009	-0.017	0.014	-0.013
	(0.03)	(0.04)	(0.04)	(0.06)	(0.04)	(0.05)	(0.02)	(0.04)
Adj. R^2	0.07	0.07	0.10	0.10	0.13	0.13	0.01	0.01
Mean Dep. Var.	0.175		0.646		0.130		0.049	

Notes: Each row presents the estimation of Equation (1) by OLS, for different subsamples of women. The table reports only the post-policy dummy variable Assignment A_i. All equations include the term $(C_i - C^*)$, its square (only in even columns (2), (4), (6), and (8)), and the corresponding interactions with the treatment assignment variable. All equations also include the full set of observable characteristics for women. Standard errors (in parentheses) are clustered at the level of the district of residence–quarter of conception. * p<0.1; ** p<0.05; *** p<0.01.

Table 6. Effect of Policy Change on the Prenatal Tests' Take-Up Rates: Education Level, Nationality, and Residence

Dep. Var.:	*No Test*		*Screening Test*		*Diagnostic Test*		*Both Tests*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Specification:	*Linear*	*Quadratic*	*Linear*	*Quadratic*	*Linear*	*Quadratic*	*Linear*	*Quadratic*
1. Subsample: Low Education *(N. Obs. 9,113)*								
Assignment A_i	-0.056	-0.049	0.054	0.047	0.003	-0.004	-0.001	0.006
	(0.04)	(0.05)	(0.04)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.12	0.12	0.10	0.10	0.20	0.20	0.03	0.03
Mean Dep. Var.	0.306		0.624		0.049		0.021	
2. Subsample: Medium Education *(N. Obs. 16,379)*								
Assignment A_i	-0.072**	-0.092**	0.055*	0.062*	0.011	0.026*	0.006	0.005
	(0.03)	(0.04)	(0.03)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.06	0.06	0.11	0.11	0.21	0.21	0.02	0.02
Mean Dep. Var.	0.205		0.689		0.073		0.033	
3. Subsample: High Education *(N. Obs. 7,131)*								
Assignment A_i	-0.064**	-0.051	0.058**	0.053	0.008	0.006	-0.002	-0.008
	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)
Adj. R^2	0.03	0.03	0.09	0.09	0.18	0.17	0.02	0.02
Mean Dep. Var.	0.176		0.691		0.094		0.040	
4. Subsample: Native *(N. Obs. 24,968)*								
Assignment A_i	-0.065**	-0.070*	0.049*	0.051	0.011	0.015	0.005	0.004
	(0.03)	(0.04)	(0.03)	(0.04)	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.04	0.04	0.11	0.11	0.20	0.20	0.02	0.02
Mean Dep. Var.	0.180		0.697		0.086		0.036	
5. Subsample: Non-native *(N. Obs.7,655)*								
Assignment A_i	-0.071*	-0.075	0.079*	0.073	-0.001	0.005	-0.007	-0.004
	(0.04)	(0.05)	(0.04)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.07	0.07	0.06	0.06	0.11	0.11	0.02	0.02
Mean Dep. Var.	0.379		0.585		0.020		0.015	
6. Subsample: Metropolitan *(N. Obs. 12,788)*								
Assignment A_i	-0.145***	-0.156**	0.124**	0.138**	0.020*	0.019	0.001	-0.002
	(0.06)	(0.07)	(0.05)	(0.06)	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.08	0.08	0.09	0.09	0.19	0.19	0.02	0.02
Mean Dep. Var.	0.257		0.640		0.071		0.032	
7. Subsample: Nonmetropolitan *(N. Obs. 19,835)*								
Assignment A_i	-0.014	-0.014	0.010	0.001	0.000	0.009	0.004	0.004
	(0.03)	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.09	0.09	0.10	0.10	0.21	0.21	0.02	0.02
Mean Dep. Var.	0.208		0.691		0.071		0.030	

Notes: Each row presents the estimation of Equation (1) by OLS, for different subsamples of women. The table reports only the post-policy dummy variable Assignment A_i. All equations include the trend $(C_i - C^*)$, its squared term (only in even columns (2), (4), (6), and (8)), and the corresponding interactions with the treatment assignment variable. All equations also include the full set of observable characteristics for women. Standard errors (in parentheses) are clustered at the level of the district of residence–quarter of conception. * p<0.1; ** p<0.05; *** p<0.01.

Table 7. Effect of the Prenatal Screening Tests on Health Outcomes of the Mother: 2SLS Results

Dep. Var.:	*Weight Gain in Pregnancy*	*Weight Gain in Pregnancy >15 kg*	*Smoking in Pregnancy*	*Alcohol consumption in Pregnancy*	*Folic Acid in Pregnancy*	*Hospital Admissions in Pregnancy*
	(1)	(2)	(3)	(4)	(5)	(6)
Screening Test	-6.506	-0.445	-0.083	-0.373	-0.480	-0.680
	(2.57)	(0.20)	(0.12)	(0.20)	(0.53)	(0.26)
[p-value]	*[0.011]*	*[0.026]*	*[0.478]*	*[0.056]*	*[0.369]*	*[0.008]*
{Bonferroni corr. p-value}	*{0.169}*	*{0.392}*	*{1.000}*	*{0.837}*	*{1.000}*	*{0.116}*
{Sharpened FDR q-value}	*{0.093}*	*{0.128}*	*{0.681}*	*{0.201}*	*{0.560}*	*{0.093}*
Linear Trend, (C_i - C*)	0.009**	0.001*	0.000	0.000	0.003*	0.002***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Assignment A_i x Linear Trend ((C_i - C*)	-0.003	-0.000	-0.000	-0.000	-0.002	-0.002***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age group 25–29	0.553**	0.032	-0.007	0.034**	0.062	0.046**
	(0.24)	(0.02)	(0.01)	(0.02)	(0.04)	(0.02)
Age group 30–34	0.166	-0.003	-0.022*	0.050***	0.068	0.053**
	(0.31)	(0.02)	(0.01)	(0.02)	(0.05)	(0.02)
Age group 35+	-2.070***	-0.161***	-0.049**	-0.042	-0.068	-0.127***
	(0.44)	(0.03)	(0.02)	(0.03)	(0.10)	(0.05)
Medium Education	0.057	-0.009	-0.030***	0.017***	0.048***	0.015**
	(0.09)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
High education	-0.441***	-0.062***	-0.060***	0.013*	0.060***	0.011
	(0.10)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Employed	0.644***	0.036**	-0.001	0.027*	0.073**	0.035*
	(0.19)	(0.02)	(0.01)	(0.01)	(0.04)	(0.02)
Married	-0.640***	-0.057***	-0.058***	-0.008*	0.005	-0.007
	(0.08)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Native	0.649**	0.031	0.063***	0.037*	0.142***	0.070***
	(0.28)	(0.02)	(0.01)	(0.02)	(0.05)	(0.03)
Father Employed	0.785***	0.038*	0.006	0.056*	0.215**	0.085**
	(0.24)	(0.02)	(0.02)	(0.03)	(0.09)	(0.04)
Twin	2.485***	0.251***	-0.009	-0.001	0.013	0.056***
	(0.31)	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)
Past Miscarriage	0.153*	0.013	0.019***	0.012**	0.063***	0.023***
	(0.09)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Past Abortion	0.676***	0.066***	0.067***	0.046***	0.044**	0.029**
	(0.11)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Metropolitan Area	-0.084	0.002	0.006	0.044***	0.020	-0.010
	(0.11)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Constant	16.666***	0.543***	0.156***	0.165*	0.763***	0.378***
	(1.29)	(0.10)	(0.05)	(0.09)	(0.25)	(0.12)
F-Test excluded instrument [*p*-value]	62.38 [0.00]	62.38 [0.00]	35.44 [0.00]	35.44 [0.00]	35.44 [0.00]	37.39 [0.00]
OLS Results:						
Screening Test	0.095	0.003	0.002	-0.030***	0.132***	-0.003
	(0.07)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
Mean Dep. Var.	13.05	0.24	0.07	0.04	0.84	0.04
N. Obs.	27,594	27,594	32,623	32,623	32,623	31,004

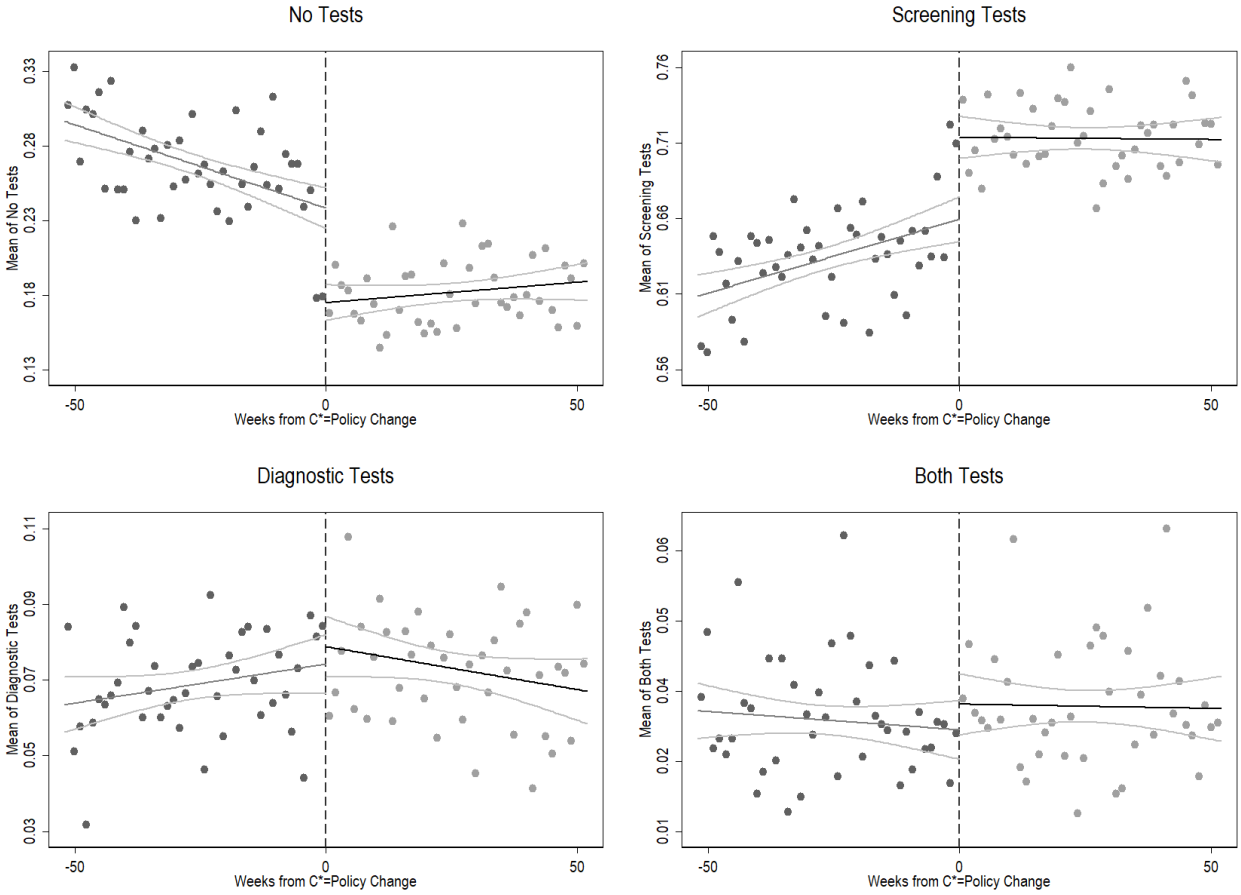
Notes: Each column presents the estimation of Equation (2) by 2SLS. The Screening Test variable is a binary variable equal to one if a woman had a noninvasive screening test during pregnancy, and zero otherwise. Assignment A_i, which is equal to one in the post-policy period, is the instrumental variable for Screening Test. Standard errors (in parentheses) are clustered at the level of the district of residence–quarter of conception. OLS results for Equation (2) are reported for the Screening Test coefficient only. Below the 'Screening test' coefficient, the table reports the original *p*-values (square brackets) and the "adjusted" *p*-values, i.e. the Bonferroni corrected *p*-values and the sharpened False Discovery Rate, FDR, *q*-values by Anderson, 2008 (in brackets), to account for multiple hypotheses testing. * p<0.1; ** p<0.05; *** p<0.01.

Table 8. Effect of the Prenatal Screening Tests on Health Outcomes of the Newborn: 2SLS Results

Dep. Var.:	***Low Weight***	***Newborn Length***	***Head Circumfer.***	***Low Apgar score - 1 min***	***Low Apgar score - 5 min***	***Resuscitat.***	***Preterm***	***Stillbirth***	***Male***
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Screening Test	0.079	-3.021	-1.710	0.008	-0.031	-0.141	0.097	-0.025	-0.200
	(0.09)	(2.17)	(1.18)	(0.34)	(0.13)	(0.11)	(0.11)	(0.02)	(0.21)
[p-value]	*[0.391]*	*[0.163]*	*[0.149]*	*[0.980]*	*[0.809]*	*[0.220]*	*[0.359]*	*[0.268]*	*[0.339]*
{Bonferroni corr. p-value}	*{1.000}*	*{1.000}*	*{1.000}*	*{1.000}*	*{1.000}*	*{1.000}*	*{1.000}*	*{1.000}*	*{1.000}*
{Sharpened FDR q-value}	*{0.560}*	*{0.428}*	*{0.428}*	*{0.957}*	*{0.896}*	*{0.531}*	*{0.560}*	*{0.560}*	*{0.560}*
Linear Trend, (C_i - C*)	0.000	0.000	0.000	-0.000	0.000	0.000*	-0.000	0.000	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Assignment A_i x Lin. Trend (C_i - C*)	-0.000*	0.006	0.003	0.000	-0.000	-0.000*	-0.000	-0.000	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age group 25–29	-0.004	0.215	0.170*	0.007	0.005	0.012	-0.007	0.003	0.018
	(0.01)	(0.18)	(0.10)	(0.03)	(0.01)	(0.01)	(0.01)	(0.00)	(0.02)
Age group 30–34	0.001	0.203	0.174	0.016	0.010	0.018*	-0.003	0.004**	0.011
	(0.01)	(0.21)	(0.12)	(0.03)	(0.01)	(0.01)	(0.01)	(0.00)	(0.02)
Age group 35+	0.038**	-0.665*	-0.273	0.035	0.012	-0.010	0.035*	-0.003	-0.044
	(0.02)	(0.38)	(0.21)	(0.06)	(0.02)	(0.02)	(0.02)	(0.00)	(0.04)
Medium Education	-0.012***	0.165***	0.068**	-0.026***	-0.011***	-0.002	-0.009**	-0.000	0.006
	(0.00)	(0.05)	(0.03)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
High education	-0.016***	0.264***	0.181***	-0.027***	-0.006	-0.004	-0.013***	-0.001	0.010
	(0.00)	(0.06)	(0.04)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Employed	-0.010	0.245*	0.135*	-0.003	0.000	0.004	-0.009	0.001	0.012
	(0.01)	(0.15)	(0.08)	(0.02)	(0.01)	(0.01)	(0.01)	(0.00)	(0.02)
Married	-0.003	0.007	0.046**	-0.001	0.001	0.001	-0.001	0.000	-0.009
	(0.00)	(0.04)	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Native	0.005	-0.018	-0.045	-0.020	-0.006	0.007	-0.015	0.002	0.012
	(0.01)	(0.22)	(0.12)	(0.04)	(0.01)	(0.01)	(0.01)	(0.00)	(0.02)
Father Employed	-0.022	0.551*	0.202	-0.021	-0.019	0.016	-0.029*	-0.001	0.032
	(0.01)	(0.33)	(0.18)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)
Twin	0.562***	-3.626***	-1.504***	0.223***	0.130***	0.060***	0.539***	0.001	-0.047**
	(0.02)	(0.17)	(0.09)	(0.02)	(0.02)	(0.01)	(0.03)	(0.00)	(0.02)
Past Miscarriage	0.003	0.063	0.030	-0.004	0.001	0.007**	0.007	-0.001	0.005
	(0.00)	(0.05)	(0.03)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Past Abortion	-0.002	0.116*	-0.044	-0.008	-0.009*	0.005	-0.001	0.000	-0.011
	(0.00)	(0.07)	(0.04)	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
Metropolitan Area	0.006	-0.050	-0.106**	-0.020	0.003	-0.001	0.003	-0.001	-0.004
	(0.00)	(0.08)	(0.04)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
Constant	0.019	50.826***	34.814***	0.236	0.098	0.108**	0.025	0.018*	0.612***
	(0.04)	(0.99)	(0.54)	(0.17)	(0.06)	(0.05)	(0.05)	(0.01)	(0.10)
F-Test excluded instrument [*p*-value]	33.72	30.77	29.46	33.29	33.52	33.71	33.71	33.71	33.95
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
OLS results:									
Screening Test	-0.009***	0.024	-0.045**	-0.020***	-0.013***	-0.007***	-0.006**	-0.001	0.003
	(0.00)	(0.03)	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Mean Dep. Var.	0.06	49.50	33.98	0.20	0.06	0.04	0.05	0.003	0.51
N. Obs.	32,829	32,094	32,180	31,937	31,942	32,830	32,830	32,830	32,811

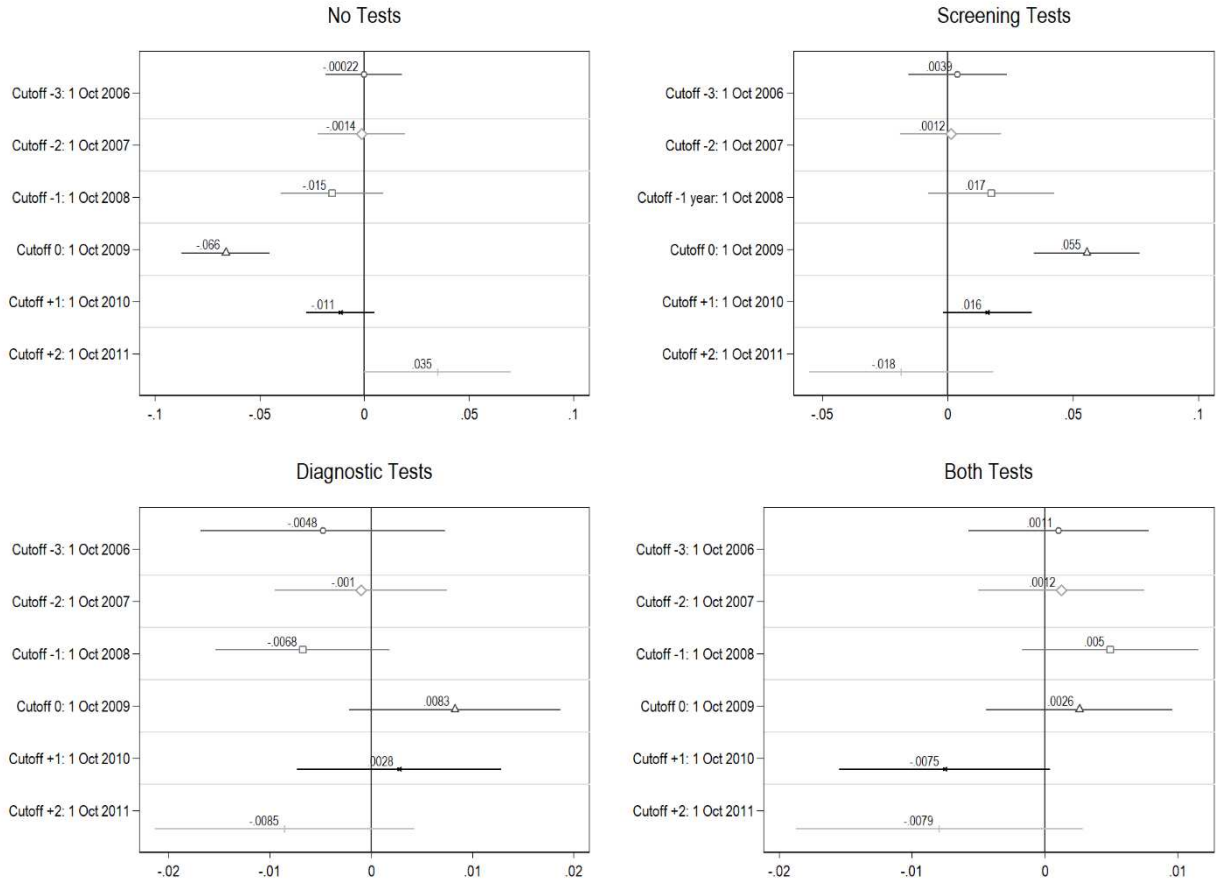
Notes: Each column presents the estimation of Equation (2) by 2SLS. The *Screening Test* variable is a binary variable equal to one if a woman had a noninvasive screening test during pregnancy, and zero otherwise. *Assignment* A_i, equal to one in the post-policy period, is the instrumental variable for *Screening Test*. Standard errors (in round parentheses) are clustered at the level of the district of residence–quarter of conception. OLS results for Equation (2) are reported for the Screening Test coefficient only. Below the 'Screening test' coefficient, the table reports the original p-values (square brackets) and the "adjusted" *p*-values, i.e. the Bonferroni corrected p-values and the sharpened False Discovery Rate, FDR, q-values by Anderson, 2008 (in brackets), to account for multiple hypothesis testing. * p<0.1; ** p<0.05; *** p<0.01.

Figure 1. Prenatal Tests Take-Up Rates Around the Policy Change



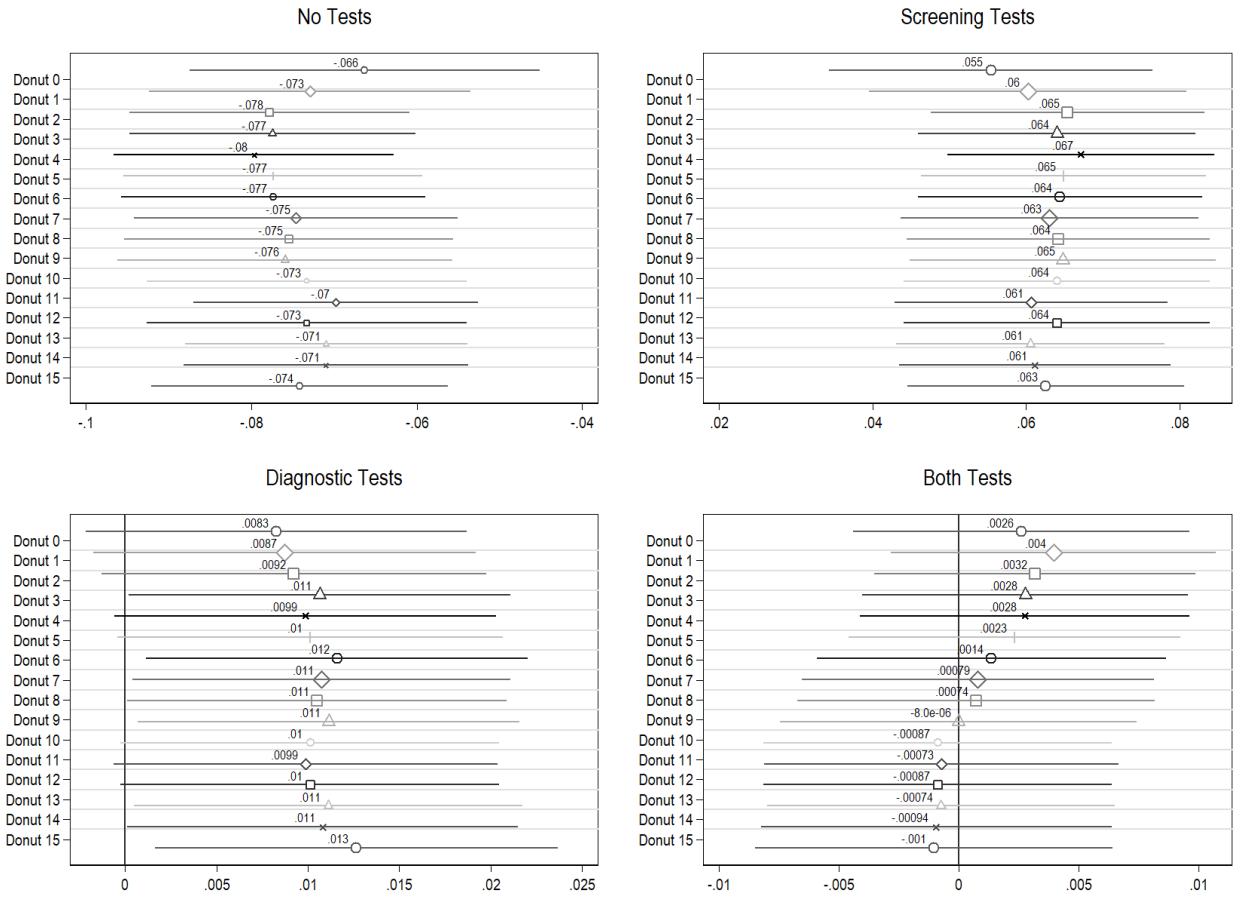
Notes: The y-axis measures the proportion of women undergoing no prenatal tests (top-left panel), a screening test (top-right panel), a diagnostic test (bottom-left panel), and both tests (bottom-right panel). The x-axis measures the number of weeks to (from) the policy change date (the zero value). Each dot represents the proportion for that week. The solid lines are the fit of a linear regression model, separately estimated on both sides of the cutoff point. The light grey lines represent the 95 percent confidence interval.

Figure 2. Falsification Tests on the Years Preceding and Following the Policy Change



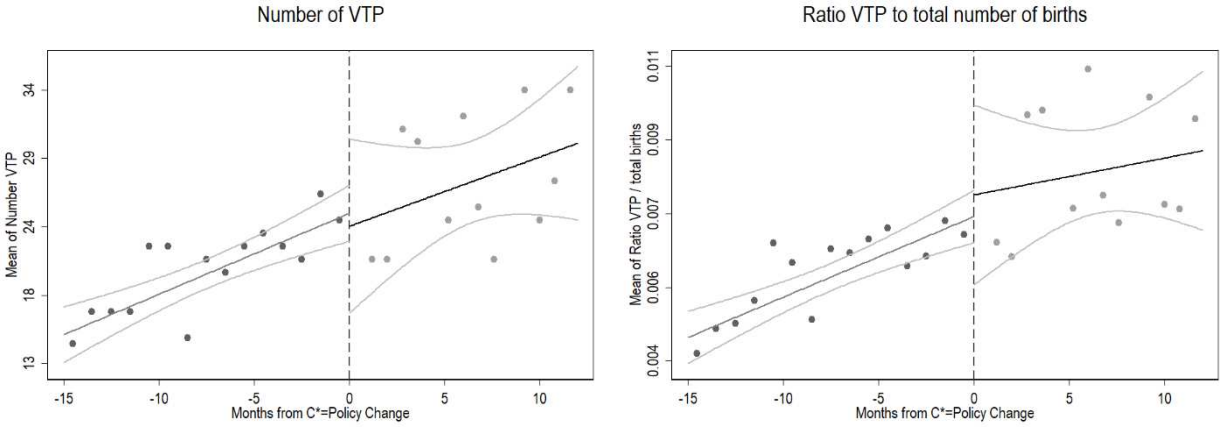
Notes: Each plot shows six replications of Equation (1). The dependent variables are no prenatal tests (top-left panel), screening test (top-right panel), diagnostic test (bottom-left panel), and both tests (bottom-right panel). In each replication, we consider a different cutoff date. The plots show the point estimates and the 95 percent confidence intervals (standard errors clustered at the level of the week of conception). The x-axis measures the estimated coefficients for the treatment assignment A_i, a dummy variable equal to one in the post-policy period, in Equation (1), for the linear specification only. The y-axis reports the considered cutoff dates. The cutoff 0 corresponds to the actual policy change date 1 October 2009. In all specifications, the data considered cover the 52 weeks before and 52 weeks after the cutoff date, except for Cutoff +2, 1 October 2011, where only 35 weeks before and 25 weeks after the cutoff date are considered because of data unavailability.

Figure 3. Alternative Sample Inclusion Criteria ('Donut Hole' Strategies)



Notes: Each plot shows several replications of Equation (1). The dependent variables are no prenatal tests (top-left panel), screening tests (top-right panel), diagnostic tests (bottom-left panel), and both tests (bottom-right panel). In each replication, we consider different inclusion criteria in the sample. The 'Donut 0' corresponds to the baseline sample. In the 'Donut 1' sample, we exclude women whose conception date falls in the week preceding the policy change, in the 'Donut 2' we exclude all women conceiving during the two weeks that precede the policy change, etc. The plots show the point estimates and the 95 percent confidence intervals (standard errors clustered at the level of the week of conception). The x-axis measures the estimated coefficients for the treatment assignment A_i, a dummy variable equal to one in the post-policy period, in Equation (1), for the linear specification only. The y-axis reports the considered sample inclusion criteria ("donut hole" strategy).

Figure 4. Voluntary Termination of Pregnancy Around the Policy Change



Notes: The y-axis measures the number of voluntary terminations of pregnancy (VTP, left panel), and the abortion ratio equal to the number of VTP to the total number of births (right panel). The x-axis measures the number of months to (from) the policy change date (the zero value, October 2009). Each dot represents the number of VTP or the VTP ratio for that month. The solid lines are the fit of a linear regression model, separately estimated on both sides of the cutoff point. The light grey lines represent the 95 percent confidence interval. Data include all VTPs that occurred during the second trimester of pregnancy, within the administrative borders of Piedmont, between July 2008 and October 2010. The source of data is the Italian Statistical Office (Istat).

Appendix

Table A1. Definition of Mothers' Characteristics

Variable	*Definition*
Age at Conception	Age at conception, expressed in years
Age group 18–24	Binary variable equal to one if age at conception is in the 18–24 range
Age group 25–29	Binary variable equal to one if the age at conception is in the 25–29 range
Age group 30–34	Binary variable equal to one if the age at conception is in the 30–34 range
Age group 35+	Binary variable equal to one if the age at conception is equal to or above 35
Low Education	Binary variable equal to one if the woman completed compulsory school or has no education
Medium Education	Binary variable equal to one if the woman completed high school
High education	Binary variable equal to one if the woman has a university or higher degree
Employed	Binary variable equal to one if the woman is employed
Married	Binary variable equal to one if the woman is married
Native	Binary variable equal to one if the woman was born in Italy
Father Employed	Binary variable equal to one if the father is employed
Twin	Binary variable equal to one for a twin pregnancy
Past Miscarriage	Binary variable equal to one if the woman had at least one miscarriage
Past Abortion	Binary variable equal to one if the woman had at least one abortion
Metropolitan Area	Binary variable equal to one if the woman lives in a metropolitan area; Metropolitan areas are defined by the metropolitan area of Torino—the regional capital—and of the other seven largest towns of the region (provincial capitals).

Table A2. Summary Statistics of Mothers' Characteristics

	Whole Sample		**Pre-Policy**		**Post-Policy**	
	Mean	***SD***	***Mean***	***SD***	***Mean***	***SD***
Age at Conception	30.015	5.373	30.010	5.351	30.021	5.396
Age group 18–24	0.165	0.371	0.167	0.373	0.164	0.370
Age group 25–29	0.283	0.450	0.277	0.448	0.288	0.453
Age group 30–34	0.341	0.474	0.348	0.477	0.333	0.471
Age group 35+	0.211	0.408	0.208	0.406	0.215	0.411
Low Education	0.279	0.449	0.288	0.453	0.271	0.444
Medium Education	0.502	0.500	0.507	0.500	0.496	0.500
High education	0.219	0.413	0.205	0.404	0.233	0.423
Employed	0.713	0.452	0.710	0.454	0.717	0.451
Married	0.596	0.491	0.593	0.491	0.599	0.490
Native	0.765	0.424	0.770	0.421	0.760	0.427
Father Employed	0.912	0.283	0.901	0.298	0.923	0.267
Twin	0.010	0.099	0.009	0.094	0.011	0.106
Past Miscarriage	0.123	0.328	0.120	0.325	0.125	0.330
Past Abortion	0.076	0.265	0.071	0.257	0.081	0.272
Metropolitan Area	0.392	0.488	0.377	0.485	0.408	0.491

Notes: All variables are binary variables, except for Age at Conception, which is expressed in years.

Table A3. Test of the Smoothness of Mothers' Characteristics Around the Cutoff Date of the Policy

	Assignment A_i	*Std. Err.*
Dep. Var.		
Age at Conception	0.060	(0.29)
Age group 18–24	-0.017	(0.02)
Age group 25–29	0.021	(0.02)
Age group 30–34	-0.006	(0.02)
Age group 35+	0.002	(0.02)
Low Education	0.016	(0.02)
Medium Education	-0.028	(0.02)
High education	0.012	(0.02)
Employed	-0.015	(0.02)
Married	-0.010	(0.03)
Native	-0.0002	(0.04)
Father Employed	-0.0005	(0.02)
Twin	-0.003	(0.00)
Past Miscarriage	-0.012	(0.01)
Past Abortion	0.003	(0.01)
Metropolitan Area	0.029	(0.16)

Notes: In each row, we estimate the following equation by OLS:

$$X_i = \tau_0 + \gamma A_i + \sum_{k=1}^{2} \tau_k (C_i - C^*)^k + \sum_{k=1}^{2} \pi_k A_i \times (C_i - C^*)^k + \mu_i,$$

where each single observed mother characteristic X_i is considered a placebo outcome and is regressed on: 1) the treatment assignment dummy variable A_i, which is equal to one in the post-policy period; 2) the term (C_i – C*), that is, the number of weeks between the conception date C_i and the cutoff date C*; 3) the term (C_i – C*) squared; and 4) the interaction of the trend (C_i – C*) with the treatment assignment variable. The table reports only the coefficient γ, for the post-policy dummy variable. Standard errors (in parentheses) are clustered at the level of the month of conception. The number of observations is 32,623 in all equations. * p<0.1; ** p<0.05; *** p<0.01.

Table A4. Effect of Policy Change on Prenatal Tests' Take-Up Rates: Alternative Specifications

	(1)	(2)	(3)	(4)	(5)
	Parametric Quadratic Weekly Trend	*Parametric Cubic Weekly Trend*	*Nonparametric Linear Weekly Trend*	*Parametric Linear Daily trend*	*Nonparametric Linear Daily Trend*
Dep. Var.: No Test					
Assignment A_i	-0.071***	-0.025	-0.065***	-0.062***	-0.061***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Adj. R^2	0.08	0.08		0.08	
Bandwidth			42		322
Dep. Var.: Screening Test					
Assignment A_i	0.055***	0.002	0.045***	0.053***	0.044***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Adj. R^2	0.10	0.10		0.10	
Bandwidth			37		282
Dep. Var.: Diagnostic Test					
Assignment A_i	0.013*	0.011	0.009	0.006	0.006
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.20	0.20		0.20	
Bandwidth			59		364
Dep. Var.: Both Tests					
Assignment A_i	0.002	0.012	0.004	0.004	0.005
	(0.01)	(0.01)	(0.003)	(0.002)	(0.003)
Adj. R^2	0.02	0.02		0.03	
Bandwidth			41		306

Notes: Each row presents the estimation results for Equation (1). The table reports only the post-policy dummy variable Assignment A_i. Columns (1)-(2)-(4) report OLS results that include the trend ($C_i - C^*$), its squared term (columns (1) and (2)), its cubic term (column (2)), and all the corresponding interactions with the treatment assignment variable. Columns (3) and (5) report results for local linear nonparametric estimation methods. The optimal bandwidth is chosen by the MSE-optimal bandwidth selector, and the observations are weighted by a triangular kernel (Calonico et al., 2014, 2015). All equations also include the full set of observable characteristics for women (age, highest education level, mother's employment status, father's employment status, marital status, twin pregnancy, previous miscarriages and abortions, area of residence, and nationality). In columns (1)-(2)-(3) the running variable term ($C_i - C^*$) is the number of weeks between the conception date C_i and the cutoff date C^*. In columns (4) and (5) the running variable term ($C_i - C^*$) is the number of days between the conception date C_i and the cutoff date C^*. Standard errors (in parentheses) are clustered at the level of the conception day/week. * p<0.1; ** p<0.05; *** p<0.01.

Table A5 Effect of Policy Change on Prenatal Tests' Take-Up Rates: Alternative Specifications

	(1)	(2)	(3)
	Donut 9-13	*Donut 9-14*	*Donut 10-14*
Dep. Var.: No Test			
Assignment A_i	-0.049***	-0.051***	-0.046***
	(0.01)	(0.01)	(0.01)
Adj. R^2	0.09	0.09	0.09
Dep. Var.: Screening Test			
Assignment A_i	0.041***	0.042***	0.039***
	(0.01)	(0.01)	(0.01)
Adj. R^2	0.10	0.10	0.10
Dep. Var.: Diagnostic Test			
Assignment A_i	0.011**	0.013**	0.012**
	(0.01)	(0.01)	(0.01)
Adj. R^2	0.20	0.20	0.20
Dep. Var.: Both Tests			
Assignment A_i	-0.003	-0.004	-0.004
	(0.00)	(0.00)	(0.00)
Adj. R^2	0.03	0.03	0.03

Notes: Each row presents the estimation results for Equation (1). The table reports only the post-policy dummy variable Assignment A_i. All equations also include the trend ($C_i - C^*$), and the full set of observable characteristics for women (age, highest education level, mother's employment status, father's employment status, marital status, twin pregnancy, previous miscarriages and abortions, area of residence, and nationality). In the 'Donut 9-13' (column 1), 'Donut 9-14' (column 2), and 'Donut 10-14' (column 3) samples, we exclude all women whose conception date falls in the intervals 9-13, 9-14 and 10-14 weeks at the policy change, respectively. In all specifications, the data cover the 52 weeks before and 52 weeks after the cutoff date. Standard errors (in parentheses) are clustered at the level of conception week. * p<0.1; ** p<0.05; *** p<0.01.

Table A6 Effect of Policy Change on Prenatal Tests' Take-Up Rates: Alternative Specifications

	(1)	(2)	(3)	(4)
	Cutoff at -13	*Cutoff at -12*	*Cutoff at -11*	*Cutoff at -10*
Dep. Var.: No Test				
Assignment A_i	0.0006	-0.003	-0.016	-0.023*
	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.08	0.08	0.08	0.08
N. Obs.	31,841	31,907	31,974	32,037
Dep. Var.: Screening Test				
Assignment A_i	0.005	0.012	0.021*	0.026**
	(0.01)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.10	0.10	0.10	0.10
N. Obs.	31,841	31,907	31,974	32,037
Dep. Var.: Diagnostic Test				
Assignment A_i	-0.0008*	-0.005	-0.006	-0.005
	(0.00)	(0.01)	(0.01)	(0.01)
Adj. R^2	0.20	0.20	0.20	0.20
N. Obs.	31,841	31,907	31,974	32,037
Dep. Var.: Both Tests				
Assignment A_i	-0.003	-0.004	-0.0002	0.002
	(0.00)	(0.00)	(0.00)	(0.00)
Adj. R^2	0.02	0.03	0.02	0.02
N. Obs.	31,841	31,907	31,974	32,037

Notes: Each row presents the estimation results for Equation (1). The table reports only the post-policy dummy variable Assignment A_i. All equations also include the trend ($C_i - C^*$), and the full set of observable characteristics for women (age, highest education level, mother's employment status, father's employment status, marital status, twin pregnancy, previous miscarriages and abortions, area of residence, and nationality). In the 'Cutoff -13' (column 1), 'Cutoff -12' (column 2), 'Cutoff -11' (column 3), and 'Cutoff -10' (column 4) samples, we consider all women who conceived during the 13, 12, 11, and 10 weeks before October 1, 2009, as treated. In all specifications, the data cover the 52 weeks before and 52 weeks after the cutoff date. Standard errors (in parentheses) are clustered at the level of conception week. * p<0.1; ** p<0.05; *** p<0.01.

Table A7. Reduced Form Estimation Results for Health Outcomes of the Mother and the Newborn

Dep. Var.	*Assignment A_i*	*Std. Err.*	*Adj. R2*	*N. Obs.*
Mother health outcomes				
Weight Gain	-0.537***	(0.12)	0.03	27,594
Weight Gain >15 Kg.	-0.037***	(0.01)	0.02	27,594
Smoking	-0.005	(0.01)	0.03	32,623
Alcohol consumption	-0.022	(0.01)	0.03	32,623
Folic Acid	-0.028	(0.02)	0.06	32,623
Hospital Admissions	-0.042***	(0.01)	0.01	31,004
Newborn health outcomes				
Low Weight	0.004	(0.00)	0.10	32,829
Newborn Length	-0.165**	(0.08)	0.06	32,094
Head Circumference	-0.091	(0.06)	0.03	32,180
Low Apgar score - 1 minute	0.0002	(0.02)	0.01	31,937
Low Apgar score - 5 minutes	-0.002	(0.01)	0.01	31,942
Resuscitation	-0.008	(0.01)	0.004	32,830
Preterm	0.005	(0.01)	0.09	32,830
Stillbirth	-0.001	(0.00)	0.001	32,830
Male	-0.011	(0.01)	0.00	32,811

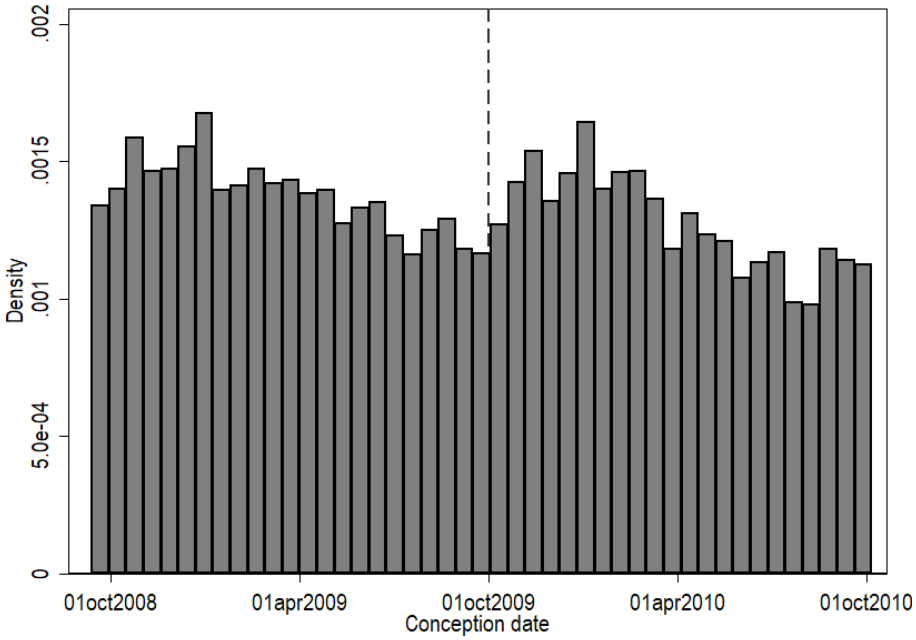
Notes: In each row, we estimate the following reduced form equation by OLS:

$$HS_i = \tau_0 + \gamma A_i + \tau_1(C_i - C^*) + \tau_2 A_i(C_i - C^*) + X_i'\eta + \mu_i,$$

where the health status HS_i of the infant/mother *i* is regressed on: 1) the treatment assignment dummy variable A_i, which is equal to one in the post-policy period; 2) the term ($C_i - C^*$), that is, the number of weeks between the conception date C_i and the cutoff date C^*; 3) the interaction of the trend ($C_i - C^*$) with the treatment assignment variable; 4) The full set of mothers' characteristics X_i. The table reports only the coefficient γ, for the post-policy dummy variable. Standard errors (in parentheses) are clustered at the level of the district of residence–quarter of conception. * p<0.1; ** p<0.05; *** p<0.01.

Figure A1. Distribution of Pregnant Women

Panel A. Histogram of Pregnant Women Density according to Conception Week, around the Policy Change



Panel B. Histogram of Pregnant Women Density according to Age at Conception (the Vertical Line is Age 35)

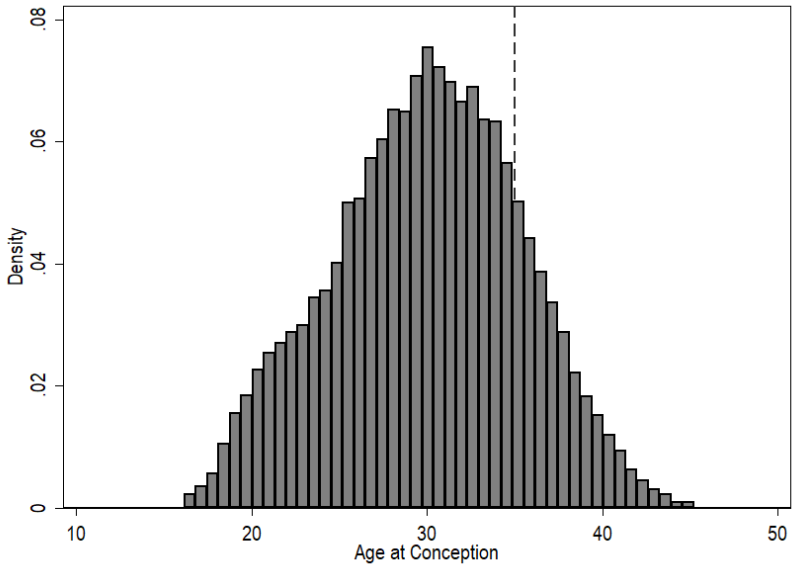


Figure A2. Mothers' Characteristics Around the Policy Change

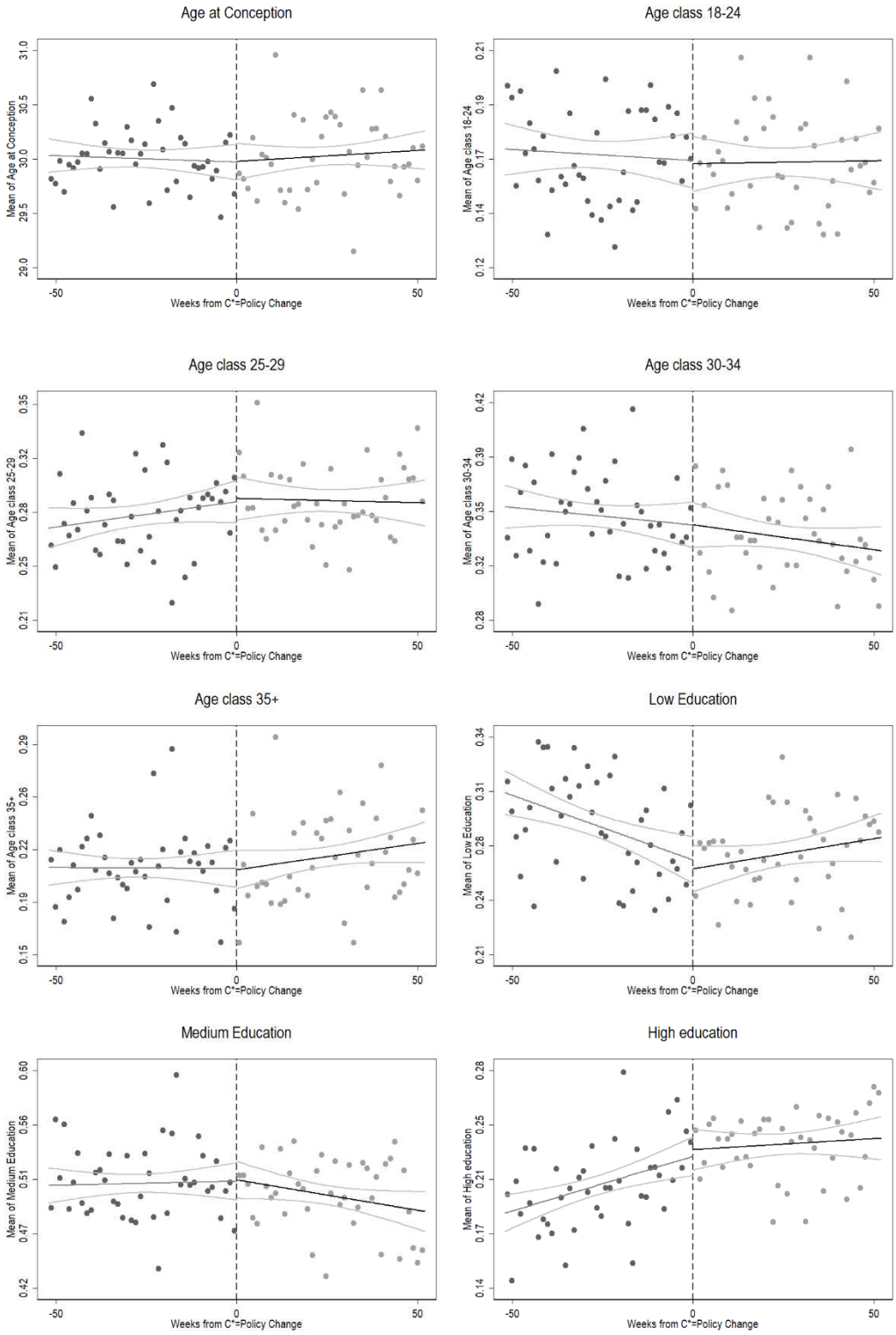
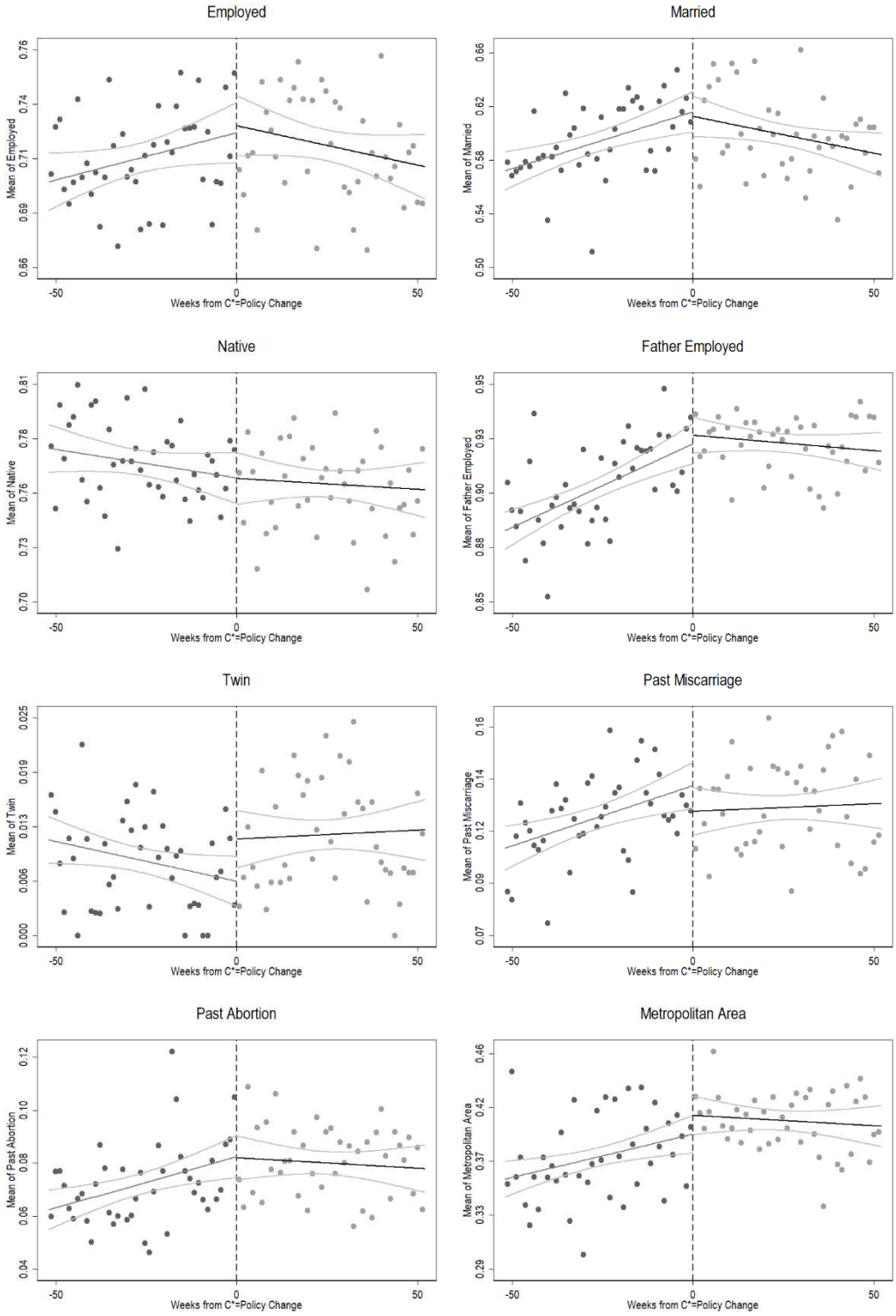
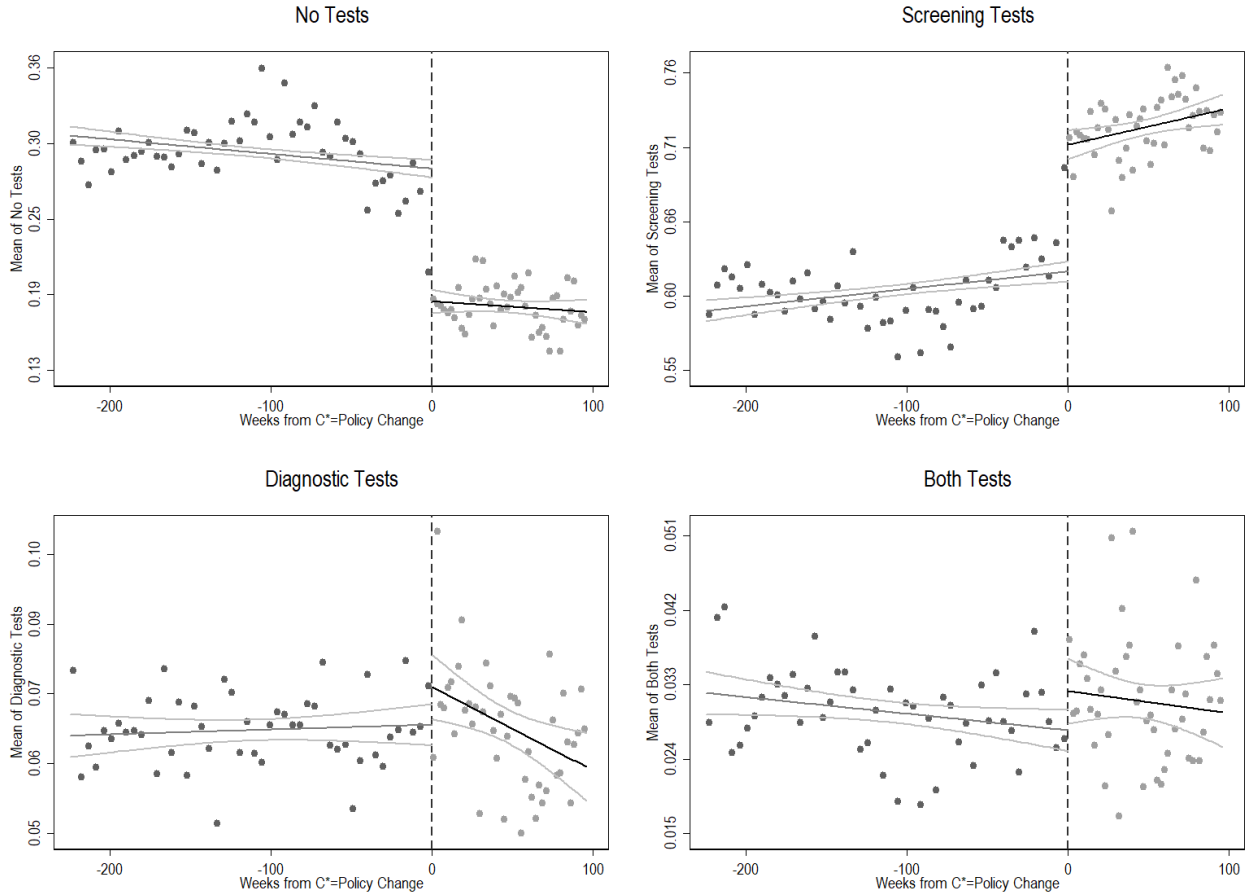


Figure A2. Mothers' Characteristics Around the Policy Change (continued)



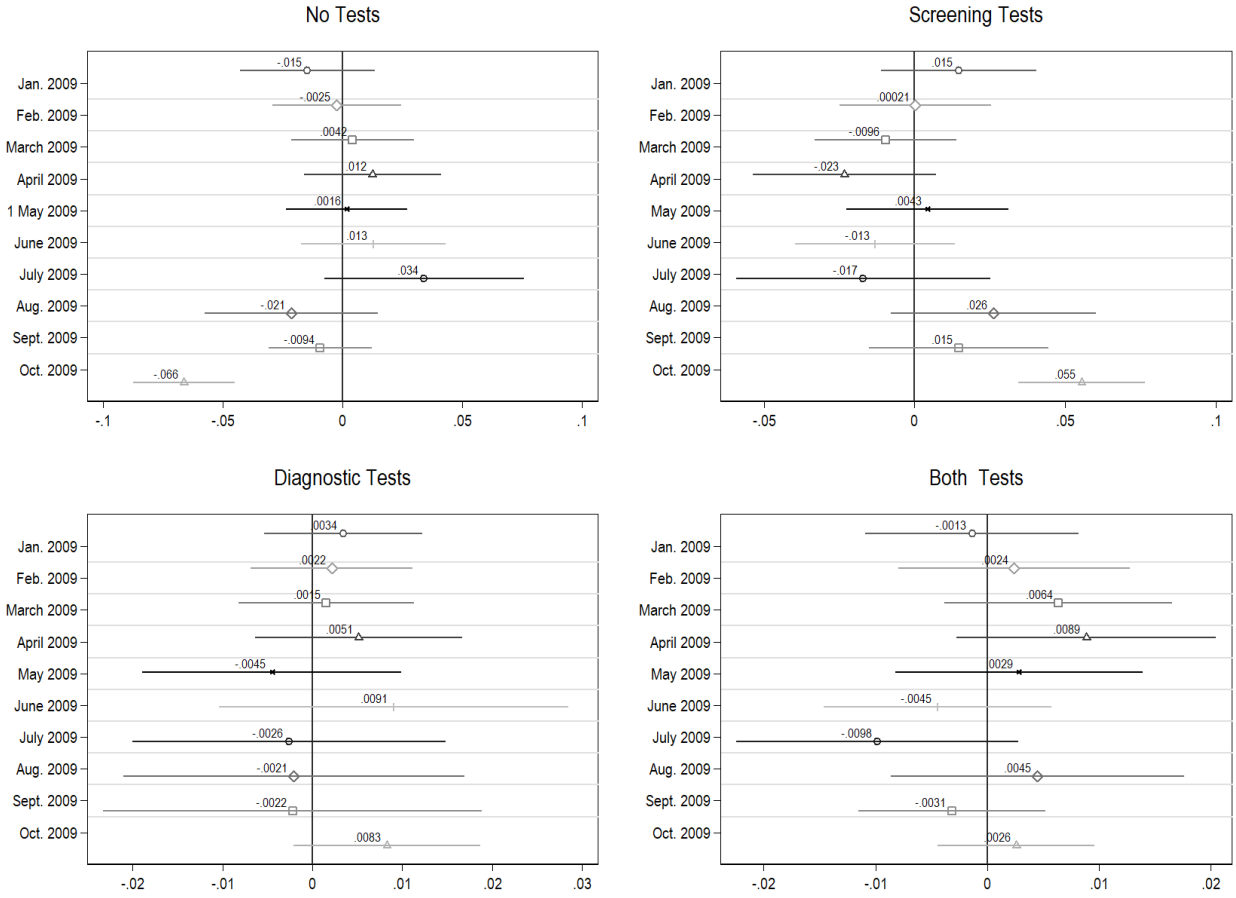
Notes: The y-axis measures the mothers' characteristics. The x-axis measures the number of weeks to (from) the policy change date (the zero value). Each dot represents the average mother's characteristics for that week. The solid lines are the fit of a linear regression model, separately estimated on both sides of the cutoff point. The light grey lines represent the 95 percent confidence interval.

Figure A3. Prenatal Tests Take-Up Rates Around the Policy Change using a Wider Time Window



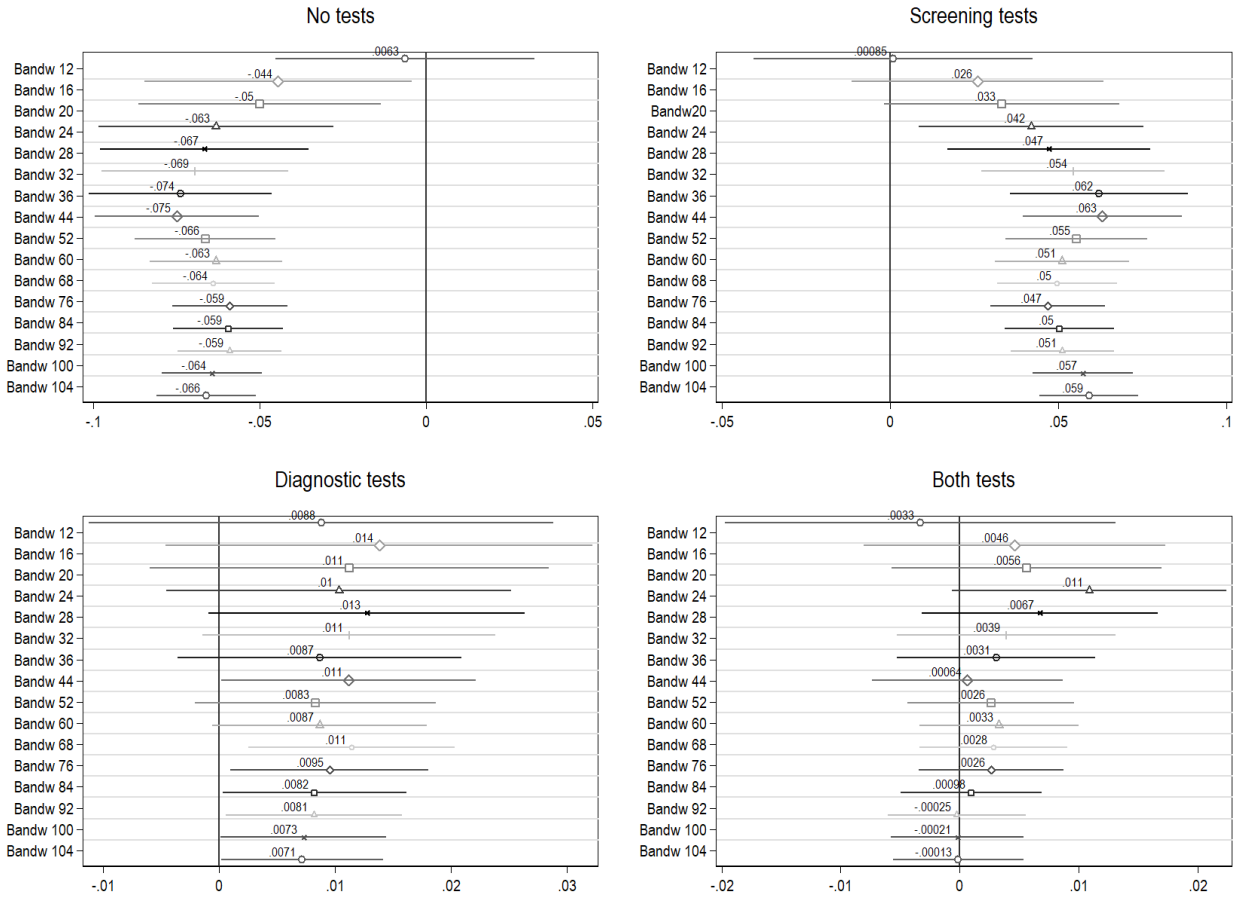
Notes: The y-axis measures the proportion of women undergoing no prenatal tests (top-left panel), a screening test (top-right panel), a diagnostic test (bottom-left panel), and both tests (bottom-right panel). The x-axis measures the number of weeks to (from) the policy change date (the zero value). The considered time interval goes from the first week of October 2005 (week -247) to the first week of May 2012 (week +96). Each dot represents the proportion for that week. The solid lines are the fit of a linear regression model, separately estimated on both sides of the cutoff point. The light grey lines represent the 95 percent confidence interval.

Figure A4. Falsification Test: Estimated Treatment Effects for Cutoff Dates Artificially Anticipated



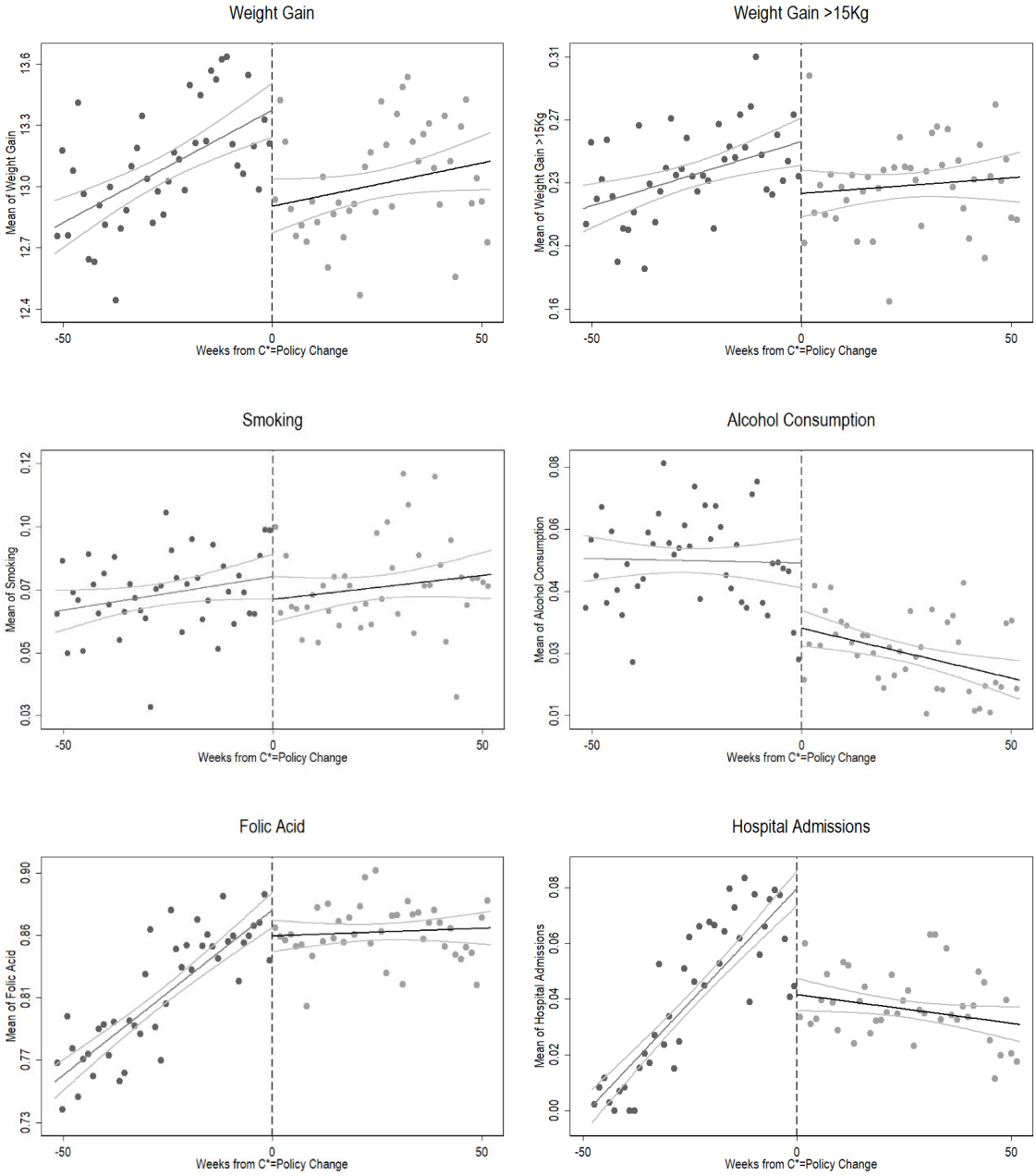
Notes: Each plot shows ten replications of Equation (1). The dependent variables are no prenatal tests (top-left panel), screening test (top-right panel), diagnostic test (bottom-left panel), and both tests (bottom-right panel). In each replication, we consider a different cutoff date for the policy change, from January 1, 2009, to October 1, 2009. The plots show the point estimates and the 95 percent confidence intervals (standard errors clustered at the level of conception week). The x-axis measures the estimated coefficients for the treatment assignment A_i, a dummy variable equal to one in the post-policy period, in Equation (1), for the linear specification only. The y-axis reports the considered cutoff date. The estimated coefficient labeled Oct. 2009 is the true cutoff date.

Figure A5. Effect of Policy Change on Prenatal Tests' Take-Up Rates: Alternative Bandwidths around the Policy Change



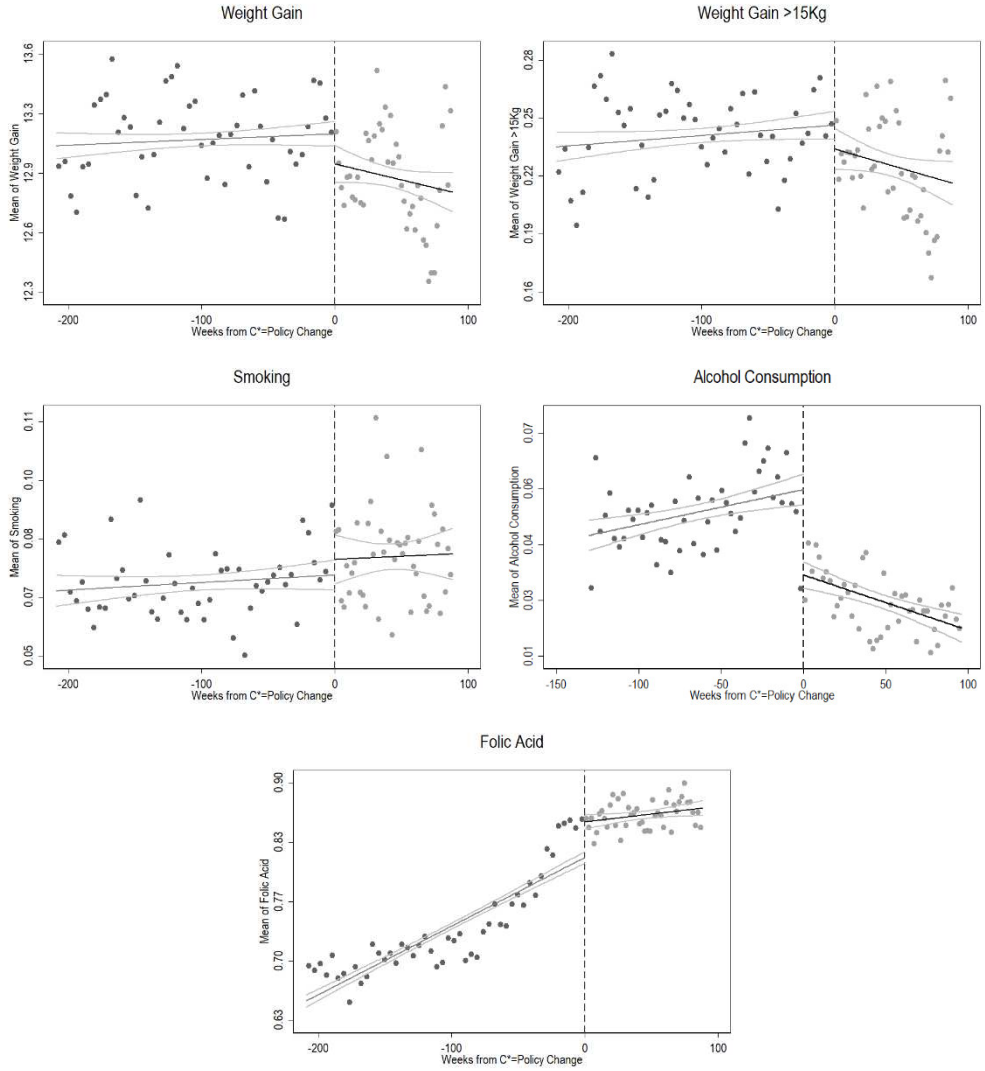
Notes: Each plot shows some replications of Equation (1). The dependent variables are no prenatal tests (top-left panel), screening test (top-right panel), diagnostic test (bottom-left panel), and both tests (bottom-right panel). In each replication, we consider a different bandwidth, i.e. a different number of weeks around the policy change. The plots show the point estimates and the 95 percent confidence intervals (standard errors clustered at the level of conception week). The x-axis measures the estimated coefficients for the treatment assignment A_i, a dummy variable equal to one in the post-policy period, in Equation (1), for the linear specification only. The y-axis reports the considered bandwidth. The bandwidth 52 represents the bandwidth for the baseline estimation results (52 weeks before and 52 weeks after the policy change).

Figure A6. Mother's Health Outcomes around the Policy Change



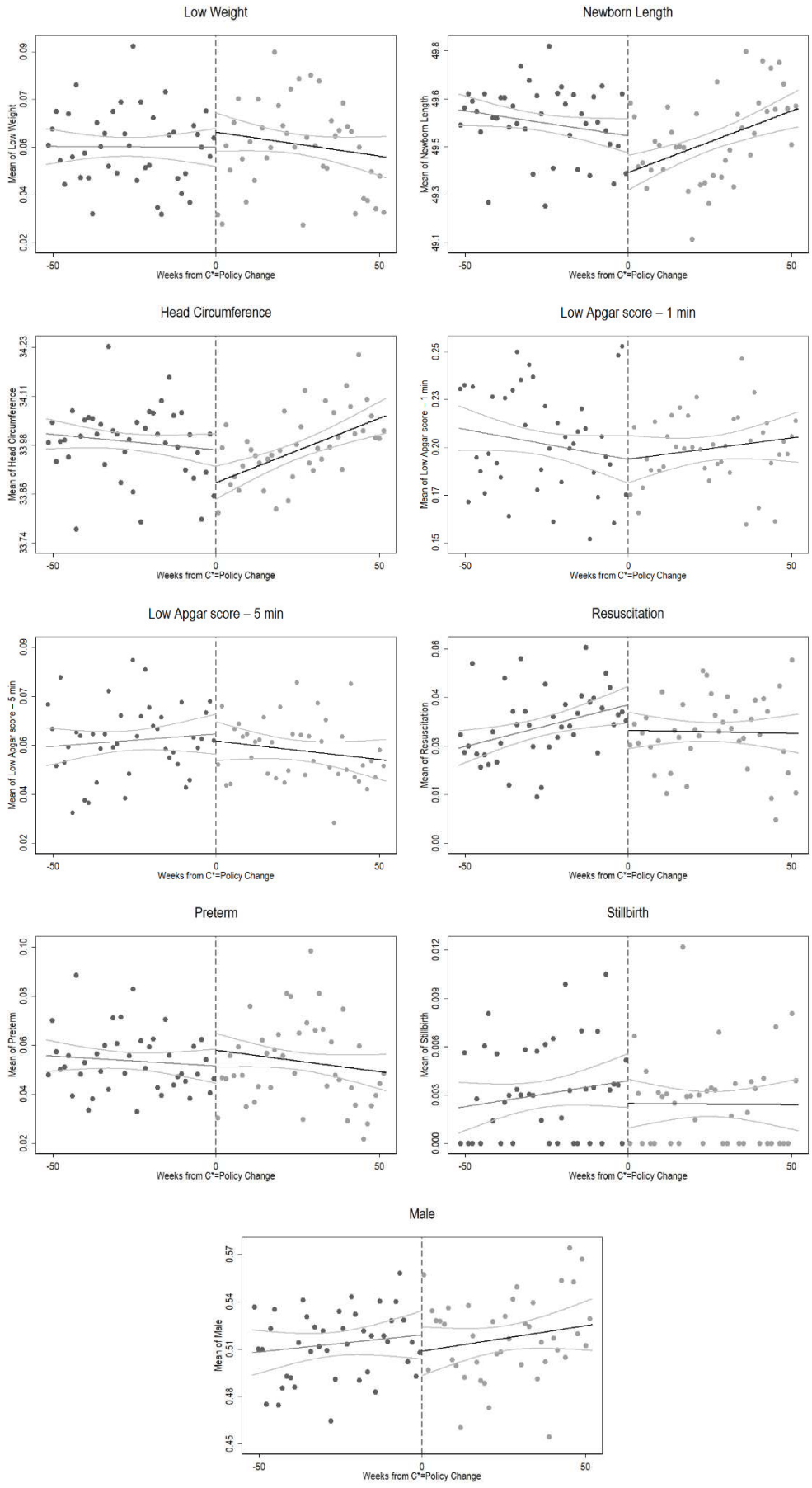
Notes: The y-axis measures the mother's health outcomes. The x-axis measures the number of weeks to (from) the policy change date (the zero value). Each dot represents the average mother's outcome for that week (Bin Average). The solid lines are the fit of a linear regression model, separately estimated on both sides of the cutoff point. The light grey solid lines represent the 95 percent confidence interval.

Figure A7. Mother's Health Outcomes around the Policy Change using a Wider Time Window



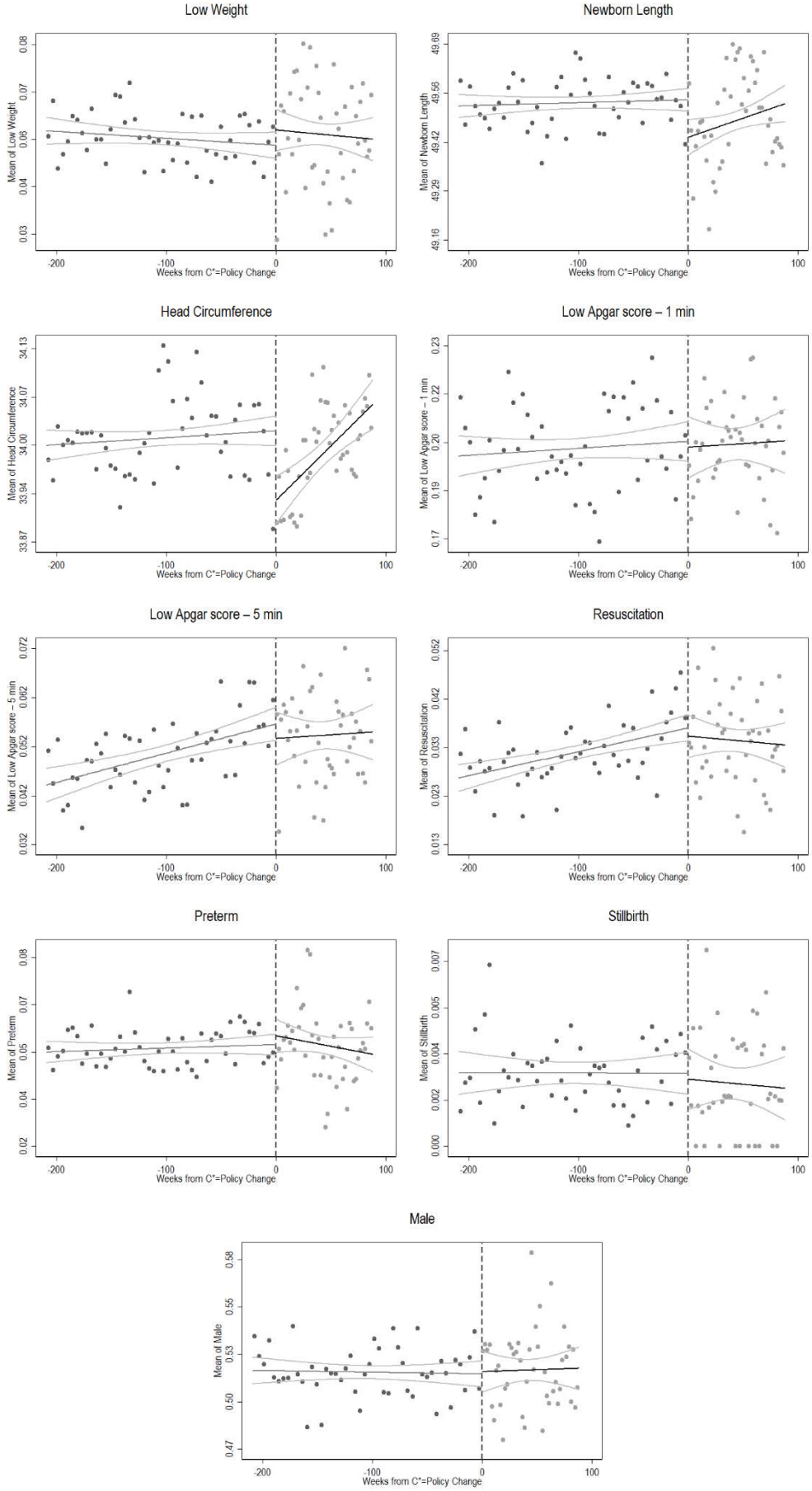
Notes: The y-axis measures the mother's health outcomes. The x-axis measures the number of weeks to (from) the policy change date (the zero value). Each dot represents the average mother's outcome for that week (Bin Average). The solid lines are the fit of a linear regression model, separately estimated on both sides of the cutoff point. The light grey solid lines represent the 95 percent confidence interval. The considered time interval goes from the first week of October 2005 (week -247) to the first week of May 2012 (week +96), except for Alcohol consumption (data unavailable before January 2007), and Hospital Admissions, (data unavailable before October 2008).

Figure A8. Newborn's Health Outcomes around the Policy Change



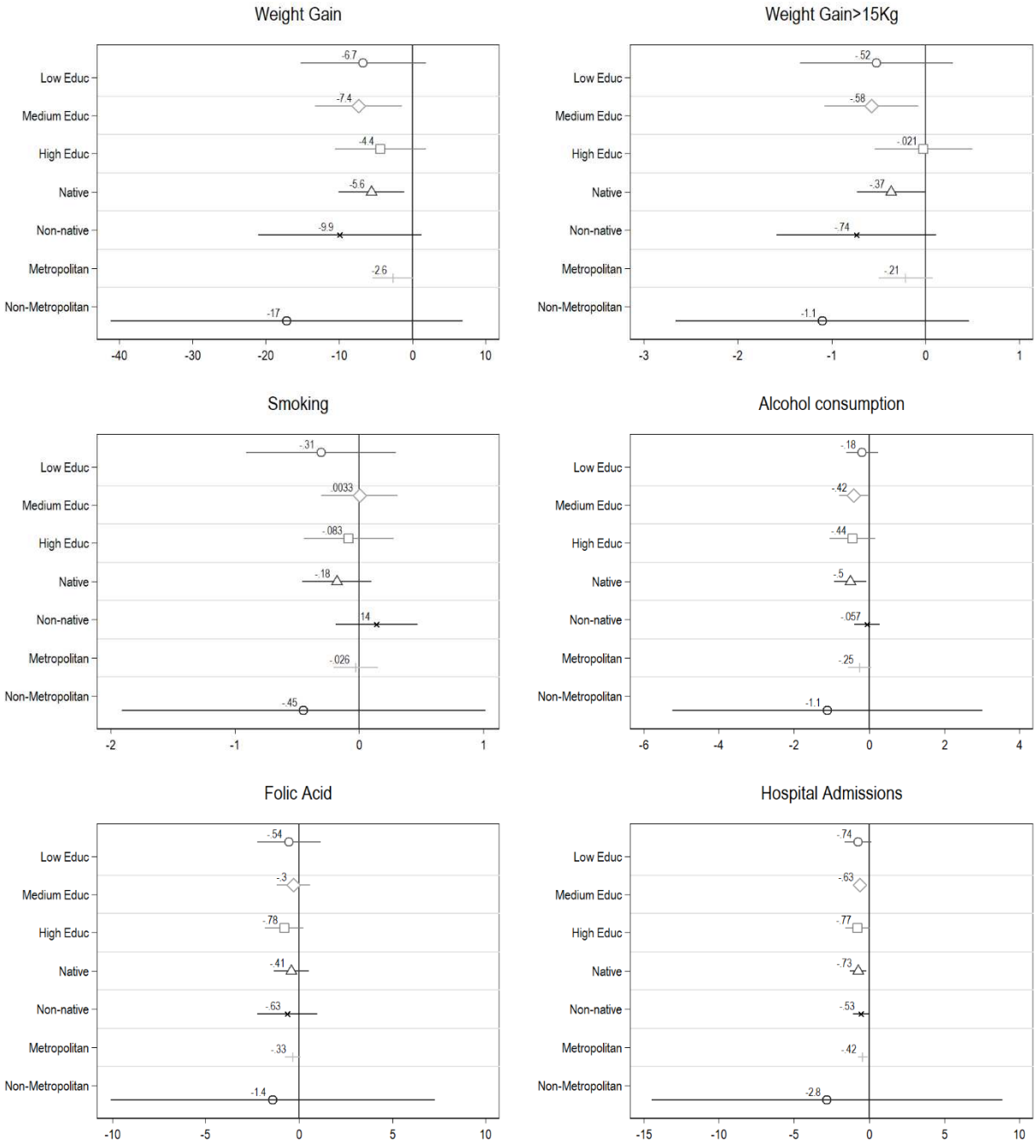
Notes: The y-axis measures the newborn's health outcomes. The x-axis measures the number of weeks to (from) the policy change date (the zero value). Each dot represents the average newborn's outcome for that week (Bin Average). The solid lines are the fit of a linear regression model, separately estimated on both sides of the cutoff point. The light grey solid lines represent the 95 percent confidence interval.

Figure A9. Newborn's Health Outcomes around the Policy Change using a Wider Time Window



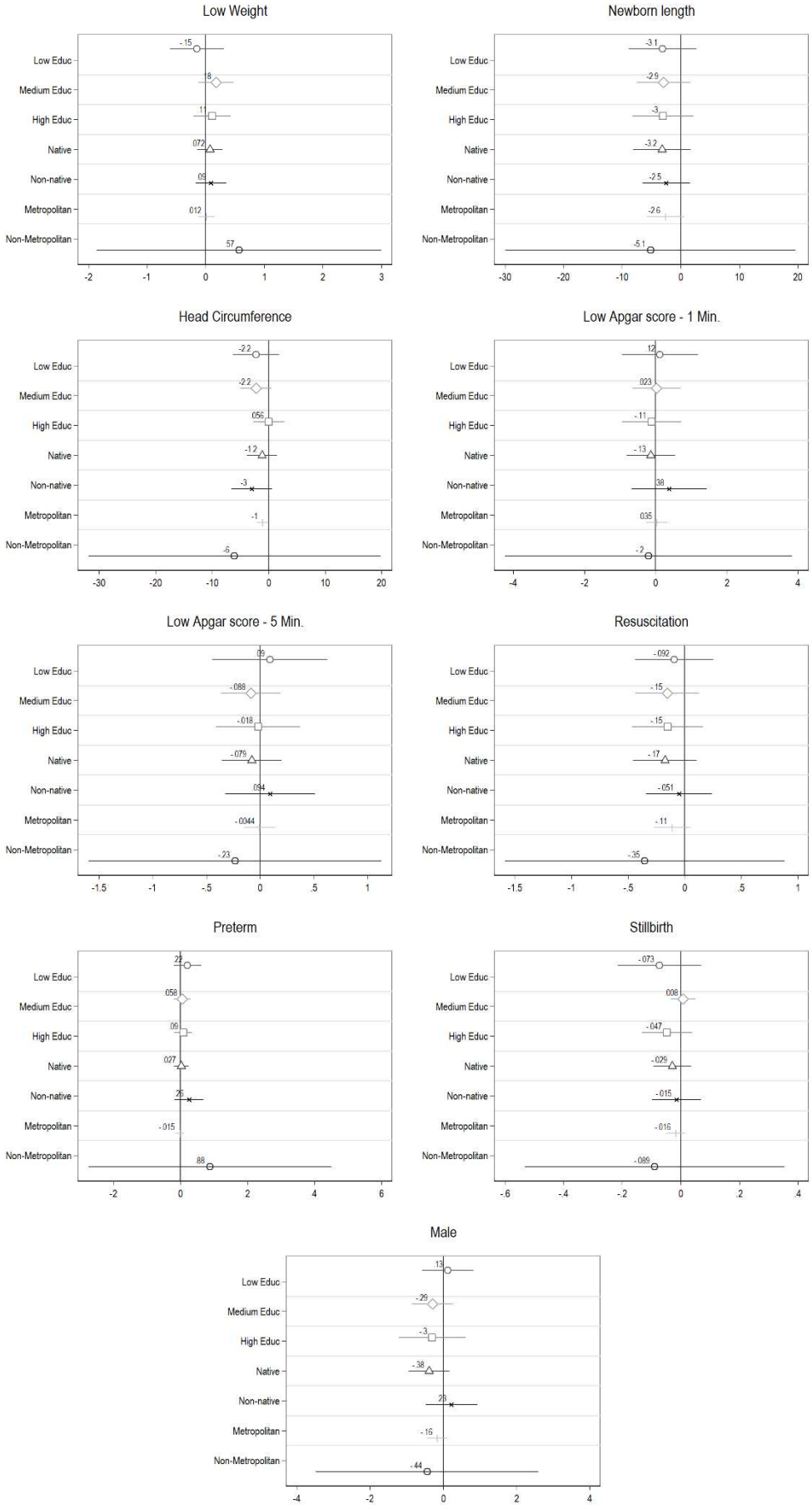
Notes: The y-axis measures the newborn's health outcomes. The x-axis measures the number of weeks to (from) the policy change date (the zero value). Each dot represents the average newborn's outcome for that week (Bin Average). The solid lines are the fit of a linear regression model, separately estimated on both sides of the cutoff point. The light grey solid lines represent the 95 percent confidence interval. The considered time interval goes from the first week of October 2005 (week -247) to the first week of May 2012 (week +96).

Figure A10. The Effect of Screening Tests on Mother's Health Outcomes by Subsamples



Notes: Each plot shows 2SLS results of Equation (2) for seven subsamples of women (low educated, medium educated, high educated, native, non-native, living in metropolitan areas, and living in nonmetropolitan areas). The dependent variables are weight gain during pregnancy and weight gain during pregnancy larger than 15 kg (top panels), smoking in pregnancy and alcohol consumption during pregnancy (middle panels), and folic acid supplements during pregnancy and hospital admission in pregnancy (bottom panels). The plots show the point estimates and the 95 percent confidence intervals (standard errors clustered at the level of the district of residence–quarter of conception). The x-axis measures the estimated coefficients for the screening test, a dummy variable equal to one if the woman had a screening test, and zero otherwise. The instrumental variable for the screening test variable is the post-policy dummy variable Assignment A_i. The y-axis reports the considered subsamples.

Figure A11. The Effect of Screening Tests on Newborn's Health Outcomes by Subsample



Notes: Each plot shows 2SLS results of Equation (2) for seven subsamples of women (low educated, medium educated, high educated, native, non-native, living in metropolitan areas, and living in nonmetropolitan areas). The dependent variables are health outcomes for newborns. The plots show the point estimates and the 95 percent confidence intervals (standard errors clustered at the level of the district of residence–quarter of conception). The x-axis measures the estimated coefficients for the screening test, a dummy variable equal to one if the woman had a screening test, and zero otherwise. The instrumental variable for the screening test variable is the post-policy dummy variable Assignment A_i. The y-axis reports the considered subsamples.