Separated at Birth: The Fragmentation of Health Insurance and Underinvestment in Preventative Care*

Sarah Kotb†

November 16th, 2024

– Click Here for the Most Recent Draft –

Abstract

Health insurance in the United States is fragmented. Americans, and especially low-income groups, receive their insurance coverage through numerous plans throughout their lives and even at one point in time, their coverage benefits can be outsourced to different insurers. This study presents empirical, causal evidence showing that such fragmentation reduces the incentives for any one plan to invest in preventative health. Specifically, I study a policy in New York's Medicaid that carved out very low birthweight newborns from the responsibilities of private plans and placed them under the public state insurance program. Once the carve-out ended in 2012 and private plans became liable for the costs of very low birthweight newborns, pregnant enrollees covered by these plans experienced more preventative care that is specifically targeted towards monitoring and reducing the risk of preterm and low birthweight newborns. These increases were above and beyond secular changes in care experienced by pregnant enrollees in the public program. Moreover, the largest gains appear among African American enrollees who exhibit a disproportionately high risk of delivering a preterm and a low birthweight newborn, suggesting that the fragmented regime had not just under-allocated prevention but also misallocated it.

JEL Codes: I13, I18, H51, J13.
Keywords: Health Insurance, Fragmentation, Medicaid, Preventative Care

*I am thankful to my PhD advisors Tim Layton, Anna Sinaiko, and Marcella Alsan. This paper benefited from constructive feedback by Jacob Goldin, Maya Rossin-Slater, Maria Polyakova, Mark Shephard, Nathananiel Breg, Kacie Dragan, Rebecca Staiger, Corinna Haberland, Tom Rutter, Jessica Cohen, Jacob Wallace, and Melissa Franco. I am also grateful to Pati McGaughey, John Feister, and Ciaran Phibbs for clarifying the clinical aspects of this project. Tal Gross provided the overleaf template that facilitated producing this manuscript. For financial support, I am grateful to the Retirement and Disability Research Consortium at the US Social Security Administration. The Institute of Humane Studies also provided financial support to present an early draft in 2023.

†Harvard University. sarahkotb@g.harvard.edu

In many high-income countries, people receive their health insurance through the state: the government serves as their primary insurer and often their *only* insurer. This is not the case in the United States, where insurance coverage is highly fragmented. An American adult typically churns between employer-sponsored insurance when employed, state-run marketplace insurance when unemployed, and Medicare when she retires. If at any point her income dips below a certain threshold, she qualifies for Medicaid, the insurance system for low-income groups. Even within each of these large systems, an enrollee can churn between multiple plans and even within one plan, different health benefits can be outsourced to different insurers.

What does this fragmentation in coverage imply for the quality of healthcare in the United States? A reasonable hypothesis is that that it would result in too little preventative care (Cebul et al., 2008). Preventative care is often cheap and can prevent expensive spending down the line – consider for example an annual physical exam that detects hypertension. From a social planner's perspective, if forgoing this visit would result in a more expensive emergency department visit or an inpatient stay, the annual physical is socially efficient. From the insurer's perspective, however, the costs of preventative care are immediate, but its benefits in preventing costly bills can be delayed to when the patient has moved on to another plan or spillover to types of care that the insurer is not responsible for. As such, for insurers in a fragmented setting, offering preventative care too readily may not be profit-maximizing.

This study contributes to this discussion by providing rigorous quasi-experimental evidence on the causal effects of fragmentation. There are at least two reasons this evidence is important. First, it may explain why preventative care in the United States is stubbornly low. Despite almost universal coverage and a national task force dedicated to identifying and encouraging high-value preventative care, only 8% of American adults receive clinically recommended preventative services and the rate of preventable deaths remains high and increasing (Borsky et al., 2018; KFF-Peterson,

The research reported herein was performed pursuant to grant RDR18000003 from the US Social Security Administration (SSA) funded as part of the Retirement and Disability Research Consortium. The opinions and conclusions expressed are solely those of the author(s) and do not represent the opinions or policy of SSA, any agency of the Federal Government, or NBER. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the United States Government or any agency thereof.

2024). Most prior studies explaining this low rate have often pointed at the consumer: in that literature, the consumer's myopia, sensitivity to price, misperception of the value of care, and other demand-side factors are often the culprits behind low prevention, but structural, supply-side policies remain understudied. Second, the results from this project should inform policy debates on insurance design and, more specifically, on the costs of having a highly fragmented insurance system. Even if a complete upheaval of the American insurance system is not politically feasible or economically desirable, there are many smaller scale, less politically salient policies that intentionally fragment coverage and are often used without a complete understanding of their cost.

The empirical focus of this paper, the policy of carve-outs, is one such example. Coverage carve-outs are policies that exclude some benefits or populations from the primary insurer's responsibility and outsource them to another insurer, typically with the well-intended goal of providing better access to care under the latter. I study the effects of one such carve-out that existed in the New York State Medicaid program and that carved out infants born with a very low birthweight to mothers covered by private Medicaid plans. For the first six months of their lives, these newborns were moved to the public Medicaid program run by the state, which allowed them to use more of the expensive healthcare they need during this critical phase without being subjected to the rationing practices that private Medicaid plans engage in (NY DOH, 2012). After these initial six months, the newborns rejoin their mothers who had remained on their private Medicaid plans throughout this time. The hypothesis that this paper raises is that the fragmentation brought about by the newborn carve-out could have inefficiently impacted the incentives of private Medicaid plans when considering the care of pregnant enrollees. Specifically, I consider whether the carve-out has reduced their incentives to invest in the preventative care during pregnancy that would have reduced the likelihood of sicker infants in the first place.

To examine this hypothesis with a causal framework, I exploit the ending of the carve-out in early 2012, which made private Medicaid plans responsible for the care of all newborns born to enrolled mothers regardless of how sick–and how costly–these newborns are. I compare the behavior of the private plans before and after the carve-out ended as it pertained to preventative care during pregnancy. My primary outcomes of interest are the rates at which pregnant enrollees were monitored for the risk of having a preterm birth, which is often correlated with a costly low birthweight

newborn, as well as the take-up of clinical treatments that are aimed at preventing a preterm birth. To control for secular trends in these outcomes, I use a control group composed of pregnant enrollees covered by the public Medicaid program in New York, called Fee-for-Service, which unlike the private plan enrollees did not experience a change in incentives to reduce the rates of preterm births.

Using this straightforward difference-in-difference, I find that monitoring for preterm birth risk among pregnant private plan enrollees increased by 8 pp, a 32% increase off the pre-period mean of the same group. Consistent with higher monitoring, the results also show that the share of pregnant enrollees who were diagnosed by an immediate risk factor of preterm birth was 3 pp (or 88%) higher after the carve-out ended. Further down the pipeline of care, I find a 1.03 pp — a 78% increase off a low baseline of 1.32 pp — observed increase in the take-up of progesterone supplementation and another 0.19 pp — 50% off a baseline of 38% — increase in the take-up of cervical cerclage, two treatments that at the time were considered the standard care for women at risk of preterm birth.

Streamlining coverage also appeared to have non-uniform effects. Concerning is the finding that African American and Black women experienced the largest effects across all outcomes. Given that this group is known to have a higher risk of preterm and low birth weight births, this non-uniform effect suggests that streamlining coverage fueled the correlation between risk and prevention, disproportionately benefiting the most vulnerable group of enrollees. If one were to make the assumption that the effects of ending the carve-out in 2012 are equal in magnitude and opposite in direction to those of placing it, a possible conclusion is that the carve-out had led to not just an underinvestment in preventative care but a mis-allocation of it.

I consider a variety of competing hypotheses that point to other explanations for the observed increases in preventative care. Specifically, I consider whether pregnant enrollees among private Medicaid plans had a higher risk after the carve-out. I do not find evidence for this: even though public plan enrollees appeared to be slightly healthier, private plan enrollees did not experience a change in underlying risk. I then examine the institutional context surrounding preterm prevention technology at the time and ask whether the increased take-up of prevention was caused by the concurrent introduction of a new potentially safer version of progesterone, the hormone used to delay birth, in the year prior to the carve-out ending. However, I find that the

use of the hormone increased across all fronts: the use of the existing version tripled despite newly-raised safety concerns and the new version was quickly adopted and covered by private plans despite its substantially higher price tag. Finally, I consider whether the increased churn across private plans in the post-period, coinciding with the exit of a few plans, led enrollees to repeat preventative measures. However, the increased prevention was robust among enrollees unimpacted by the plan exits.

Concluding then that reducing fragmentation was the reason behind the increased prevention, I examine whether there are any unintended consequences of streamlined coverage. Cream-skimming is a prime candidate. Ostensibly, one of the reasons the carve-out was placed was to reduce plan incentives to avoid covering high-risk enrollees who could financially burden the plan with an expensive newborn. I find that following the end of the carve-out, pregnant enrollees were substantially more likely to change MC plans during the pregnancy. This change in churn in the post-period was higher than that experienced by men on MC plans and older MC women who are outside of the reproductive window. This increase in churn is consistent with a decrease in plan satisfaction among pregnant enrollees, which in turn could be driven by plan attempts to screen out enrollees who have become newly high-risk with the end of the carve out.

Finally, I ask whether the heightened monitoring and prevention efforts generated by streamlining coverage worked. I examine whether infants of private plan enrollees became healthier — e.g., had a higher birthweight, a longer gestation, etc — following the end of the carve-out. The results, however, are inconclusive, partly because they are not sufficiently powered to detect meaningful changes in these outcomes and partly because the prevention methods that are appropriate for this setting were later shown to be ineffective.Progesterone supplementation, in particular, which is one of the two clinical prevention methods I examine and one which was thought to be the golden standard at the time of the study, was recently proven ineffective in a confirmatory clinical trial (Ables, 2023). Although private insurers failed in reducing adverse health outcomes, the results suggest that it is not for lack of trying. Indeed, one can summarize the findings above by saying that once they became financially liable for covering adverse birth outcomes, insurers were highly intent on preventing them.

That insurance companies can influence care is not surprising. The increasingly

dominant system of public insurance programs contracting with private insurers was premised on the fact that private insurance firms can contain costs, not just by competing against each other, but also by "managing care" (and hence giving private Medicaid its name as Medicaid *Managed Care*). Instead, the more relevant question for regulation here is which levers insurers used to stimulate preventative care. I find no evidence that private plans expanded their networks or made them more suitable by, say, including more obstetricians or specialists in high risk pregnancies. In fact, the monitoring and prevention increases I estimate retain at least two thirds of their size when controlling for the primary physician during pregnancy. Two remaining channels that insurers could have used are increasing their reimbursement prices to physicians or relaxing prior authorization requirements - essentially making it more financially lucrative and/or administratively easier for physicians to pursue prevention. Because prices are often kept as trade secrets and because prior authorization requirements are only described behind locked portals for contracted physicians, I cannot substantiate the role of these two channels in increasing prevention.

This project contributes to two strands of literature: the first is a recent one on the role of fragmented insurance coverage. Without a single payer system for American healthcare, fragmentation of coverage has often been taken for granted as a necessary characteristic of financing the sector. Only very recently have researchers started scrutinizing it as a source of inefficiency and a potential barrier to quality healthcare. To my knowledge, existing studies on this topic have largely modelled the implications of fragmenting or streamlining coverage (Li, 2023; Fang and Gavazza, 2011; Starc and Town, 2020). This paper complements their work by using a clean empirical setting with an exogenous source of variation in coverage fragmentation, circumventing the need to rely on substantial modeling assumptions.

This paper also contributes to the related discussion on the specific policy of carve-outs, which is increasingly used in public insurance programs like Medicaid as well as in employer sponsored insurance (Layton et al., 2018). Earlier studies in the 90's and early 2000s have focused on the economic rationales for instituting carve-outs like containing moral hazard and adverse selection but have not examined their costs beyond a few mentions of coordination and hassle costs (Frank et al., 1996; Frank and McGuire, 1998; Frank and Garfield, 2007; Grazier and Eselius, 1999). Similarly, more recent studies have examined the potential for improving access through Medicaid carve-outs (Auty et al., 2021, 2022). This paper highlights that although

potentially useful in access, they may have unintended consequences in suppressing preventative care because, by definition, they fragment coverage. Although the topic of fragmentation often touches upon an important political debate (and an increasingly economics-based one like in Einav and Finkelstein (2023)) regarding the role of a single-payer for American healthcare, carve-outs like the one studied here represent a low-hanging fruit: a politics-free, bureaucratic avenue that can increase preventative care.

A second question that this paper contributes to is why preventative care is underutilized. Existing studies have focused on demand-side factors ranging from discounting (e.g., Fang and Wang, 2015) and information gaps (e.g., Parente et al., 2005) to an overarching high price elasticity for preventative healthcare regardless of whether one assumes a neoclassical or behavioral perspective (e.g., Ringel et al., 2002; Brot-Goldberg et al., 2017). Some studies discussed important implications of this high elasticity when it comes to insurance design. Ellis and Manning (2007), for example, show that even if there is no financial risk inherent in preventative care, it is still efficient to cover it because it reduces future premiums. Others have focused on proposing value-based insurance design policies where preventative care receives no cost-sharing in order to reap its health and cost saving benefits (Chernew et al., 2007). This paper points to the fact that even when conditioning on insurers already providing coverage for preventative care *and* when cost-sharing does not play a role like in the Medicaid population, the larger structure of insurance still matters.

1 Background

1.1 Fragmentation of Health Insurance in the United States

In the United States, where one receives health insurance depends on one's long list of demographics: age, employment status, retirement status, income and assets, whether one has any disabilities, and whether these disabilities prohibit formal employment. Because these demographics change over a person's lifetime, people will often churn between their parents' coverage when they are young, employer-sponsored insurance when they are employed, exchange plans when they are unemployed or self-employed, Medicaid if they are low-income, and Medicare if they are above retirement age or

disabled. The majority of these systems are now delivered in a managed competition framework, meaning that within one system at least a few plans compete to attract enrollees. Although this set-up provides the economic benefits of private markets and choice, it creates room for coverage fragmentation as people move from one plan to another over time.

Even at one point in time, one may receive their insurance benefits from different insurers. Medicare enrollees, for example, may be receiving coverage from two different plans: one for hospital care and office visits and one for prescription benefits — the famous parts A, B, and D of Medicare. In other cases, a person will be enrolled in only one plan but this plan will outsource the coverage of some benefits to a third party. Prescription drugs, in most employer-sponsored insurance plans, are often outsourced to pharmacy benefit managers. Many state Medicaid programs will often carve out specific benefits, like dental care, behavioral health, and substance use benefits from the coverage responsibilities of the private Medicaid plan that is considered the primary insurer and outsource them to either specialty insurers or to the public version of the Medicaid program (Layton et al., 2018).

Despite the prevalence of insurance fragmentation, we know little about its effects. As a first-order approximation, some evidence points to the fact that co-existing insurers often impose externalities on each other. For example, the presence of supplemental insurance for the elderly, Medigap, which covers cost-sharing on Medicare leads to additional healthcare use that would not have otherwise existed. Cabral and Mahoney (2019) estimate that this offset effect cost Medicare a substantial 22% increase in spending. Even intensive margin choices with the design of this supplemental insurer can increase spending on part of the primary insurer. For example, when these supplemental Medigap plans increased their cost-sharing for office visits and prescription drugs for a group of retired enrollees in California, patients experienced an increase in inpatient hospital visits – a benefit that is paid for by Medicare (Chandra et al., 2010). Although these externalities speak to the effects of fragmentation, they are largely generated by moral hazard: consumers respond to cheaper (or more expensive) prices posed by one insurer and the second insurer internalizes the effects of this response.

Abstracting away from the demand side, there is some evidence that insurers have a lower incentive to provide preventative care in a more fragmented environ-

ment. The most relevant evidence comes from Starc and Town (2020), who compare private Medicare plans that only cover prescription drugs and Medicare plans that cover hospitalizations along with prescriptions. They show that the former type of plans, covering prescriptions only, spend less on drugs that reduce the likelihood of hospitalization, allowing the hospitalization externalities to fall on the latter type of plans. Similarly, Li (2023) shows that industry sectors with a high turnover rate invest less in preventative care, knowing that the costs of this under-provision are more likely to fall on another insurer. Taking a more life cycle perspective on insurance fragmentation, Fang and Gavazza (2011) show that individuals who have had a longer average job tenure in their working years used much less care when they retired. Coupled with the statistic that short-tenure jobs often provide less insurance benefits and worse preventative care coverage, the authors conclude that short tenures at jobs and thus short-term health insurance contracts suppress preventative care. Of course, what what makes one industry a high turn-over one and what makes a Medicare plan cover hospitalizations is likely correlated with preventative care coverage and, as such, these three studies largely relied on structurally modeling insurer incentives to isolate the effects of fragmentation. This study supplements their work by using a clean policy-induced exogenous variation to overcome the need for structural assumptions. The policy inducing this variation in question is described in more detail over the next two sections.

1.2 Insurance Carve-outs

Carve-outs refer to the exclusion of some benefits from the coverage responsibilities of the primary health insurer and outsourcing them to another insurer. In the 1980s and 90s, carve-outs were becoming increasingly common in employer-sponsored insurance with the goal of reducing moral hazard in mental healthcare (Frank et al., 1996; Frank and McGuire, 1998). The small private plans that employers often contracted with could not reign in the growing costs and employers resorted to outsourcing mental health benefits to private insurers that were specialized in mental health. The selling point for these specialized insurers is that they could identify and contract with efficient, or at least lower-cost, mental health providers *and* they had higher bargaining power when they negotiated rates with these providers. A review by Frank and Garfield (2007) suggests that these specialty insurers were successful in reducing costs without compromising access.

More relevant to the setting in hand, economists studying carve-outs also identified adverse selection and access to care as a potential reason why carve-outs would emerge (Frank et al., 1996). If there are services that are particularly expensive or would attract costly enrollees, private plans may choose to avoid covering them or ration their use with utilization management techniques or other implicit ways. The resulting concern about access is more pronounced among public insurance programs, like Medicaid, than among employers – which may explain why with these type of carve-outs, public insurers rather than private insurers often covered the carved-out services (e.g., the state Medicaid program covers these services while private Medicaid plans take over the rest of the benefits).

Some examples of these services in question include substance use treatment, coverage of expensive drugs, and the low birthweight newborn carve-out that is the empirical setting of this project. Worth noting here, is that public insurers have no particular comparative advantage in covering these services. Instead, the motivation for these carve-outs is to ensure unfettered access to these services. A few recent studies validated this motivation: Instituting a carve-out for Hepatitis C drugs, which often cost upwards of $25,000 for a single course of treatment, led to almost a twofold increase in their fills among Medicaid enrollees in four states (Auty et al., 2021). Instituting a carve-out for substance use treatment in Maryland's Medicaid has doubled outpatient admissions while removing a similar carve-out in Nebraska's Medicaid led to a 90% decline in admissions (Auty et al., 2022).

This access motivation is important. If the goal of carve-outs is to support a high-risk group (e.g., people at risk of substance use disorder or people at risk of contracting sexually transmitted diseases), then providing them with access to expensive services without exposing them to cream-skimming is indeed consistent with this goal. But, as discussed so far, the likelihood of them being "high-risk" and needing this access is not always exogenous. While Hepatitis C and substance use disorders are arguably more preventable by behavior rather than by preventative healthcare, other sources of vulnerability, the treatments for which have been carved-out, are amenable to care. HIV is one example.[1] The VLBW carve-out, which is the empirical setting of this study, is another. The inherent fragmentation of carving-out treatments in these settings may then hurt the same populations they were intended to help.

[1]Contracting HIV has been preventable with Pre-Exposure Prophylaxis since 2012. But in 2022, HIV treatments were carved out in at least two state Medicaid programs (Dawson et al., 2023).

This cost of carve-outs have been largely understudied. Most studies on carve-outs point to worse coordination as their primary costs and do so theoretically (Frank and McGuire, 1998; Grazier and Eselius, 1999). A simple example that is often discussed is when an enrollee receives care for most services through an insurer with a specific network and care for the carved-out service from another insurer with a non-overlapping network. This specific cost of fragmentation is obvious at a first glance, and perhaps remediable at a second one with coordination efforts. However, no where in this carve-out literature has appeared the link between the carve-out endogenously creating to the medical need in the first place. [2]

1.3 The NY Medicaid Setting and the VLBW Carve-out

The setting for this project is New York Medicaid in the early 2010s. Like most states, NY Medicaid operates through contracting with private plans, called Managed Care (MC) plans, to offer insurance coverage. By 2010, the state had completed its staggered transition to MC (Rockefeller Institute, 2015). Because of the late transitions in upstate NY that overlap with the first years of this study, my analysis will focus largely on enrollees in New York City as well as enrollees in upstate counties that transitioned by 2004.

Pregnant women, as most adults, were a part of the population mandated to be under MC, although some groups either had to or chose to stay at the public Medicaid program – called Fee-for-Service (FFS). Their newborns automatically qualified for Medicaid (NY DOH, 2013). More importantly, the default option was to enroll the newborn in the same type of Medicaid and the same plan as the mother: If the mother was on FFS, so was the newborn and if the mother was on MC, so was the newborn. Even though mothers on MC plans had the freedom to choose any plan available for their newborn, between 2009 and early 2012, 95% of them went with the default option of enrolling their newborns in the same plan (NY DOH, 2011b, 2012).[3]

Newborns under 1,200 grams – considered by the NY Medicaid program as very

[2] Grazier and Eselius (1999) do briefly discuss the potential for discouraging preventative care but they do so in the context of mental health carve-outs where the specialty mental healthcare insurer provides less preventative care because of high turnover and absence of lock-in policies.

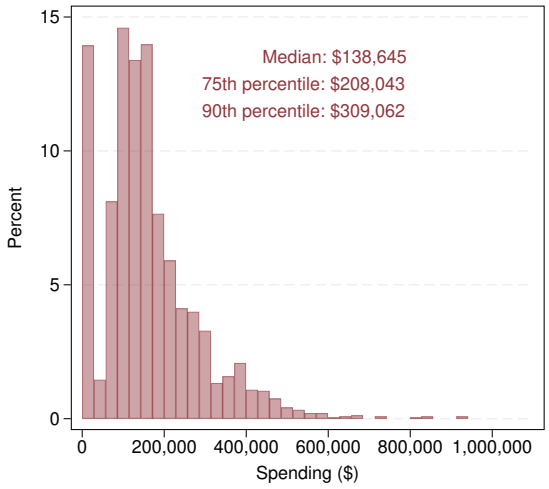
[3] Estimate comes from my primary dataset which includes enrollment information and is discussed in more detail in section 3.

low birthweight (or VLBW for short) – formed the exception to this rule.[4] Since 1999, a part of the contract between the state and MC plans stipulated that should a VLBW infant be born under a plan's roster, this newborn would be carved-out of the plan's responsibility and insured directly by the state (NY DOH, 2012; Stankaitis et al., 2005). That is, they would be covered by the state's FFS program for the first six months of their lives.

Figure 1
Spending on FFS Newborns In Their First Six Months

(a) By the Birthweight of the Newborn

(b) VLBW Newborns



Notes: Both graphs reflect FFS spending in the pre-period of this study from 2008 through March 2011 in NYC. The graph excludes newborns who had both MC and FFS during their first six months, newborns who were enrolled for fewer than six months, and newborns who died within their first six months of life.

To my knowledge, there is no publicly available official documentation regarding why this carve-out has come to exist. However, it is highly likely that the high spending of these newborns is responsible. Panel (A) in Figure 1 shows the distribution of health spending by the FFS program, showing that the average newborn under 1,250 grams cost the state about $160,000 in their first six months of life. This is almost a staggering 12 times the spending on a newborn with a healthy birthweight, reflecting the expensive, long stays in neonatal intensive care units (NICUs) that VLBW newborns require.[5] Even when comparing it to low (but not very low) birthweight

[4]This cutoff differs from the clinical standards which define VLBW infants as those with birthweight 1,500 grams or lower (Cutland et al., 2017).

[5]The majority of the spending is driven by inpatient care. Appendix figure A1 shows the breakdown by inpatient versus outpatient care.

newborns who also use NICUs, VLBW newborns are almost three times as expensive. Panel (b) shows that the distribution of spending in dollars rather than logs is highly skewed such that spending at the left-tail, reaching upwards of $300,000, still occurred among 10% of the newborns.

If these levels of spending on VLBW newborns carry from the FFS program to MC plans, then once they are forced to cover them, private plans may resort to rationing the care that these newborns receive. This stands in contrast to the FFS program, which is fully run and funded by the state, meaning that the program has no profit-maximization incentives and would consequently cover these newborns with no rationing. Indeed, two studies that compared infants born in a small window around the two sides of the birthweight cutoff found that the newborns carved-out to FFS benefited from higher hospital inpatient spending and experienced fewer hospital readmissions (Lee, 2020; Liu and Lim, 2021). Alternatively, MC plans may choose to avoid covering pregnant enrollees who exhibit a high likelihood of having a preterm newborn. In other words, they would engage in cream-skimming their enrollees. Resorting to either of these strategies would be profit-maximizing but obviously socially undesirable. The socially desirable *and efficient* strategy would be that plans hold their pregnant enrollee pool constant and ensure that they have health pregnancies such that they do not lead to a VLBW delivery.

1.3.1 Can Healthcare "Prevent" VLBW Deliveries?

In order to avoid premature newborns with a low birthweight, health practitioners and guidelines have pointed to a pipeline of care (Patel and Rumore, 2012). Although these recommendations have changed over time, this section describes clinical recommendations prior to the end of the carve-out in 2012. The pipeline starts with physicians identifying high-risk pregnancies. This is typically done by inquiring about their past pregnancies because prior preterm birth is often the strongest predictor of a recurrent preterm birth (Stewart and Graham, 2010). Alternatively (and more testable using claims data), physicians use a transvaginal ultrasound, an imaging procedure that measures the length of the cervix and checks whether it is too short – also another risk factor for a preterm birth (Wolfberg, 2012). Although it is not typically indicated by guidelines for all pregnancies, there was an ongoing conversation around whether universal screening of cervical length should be followed and high-

risk pregnant women often undergo this procedure in the first trimester to accurately check whether the cervix length is too short (Ressel, 2002; Berghella, 2012; Society for Maternal-Fetal Medicine and Berghella, 2012). Based on prior history and this first measurement, a pregnant woman may undergo this procedure more than once in order to continuously measure the cervix length and detect any changing patterns.[6]

Once a prior preterm birth or a short cervix has been diagnosed, guidelines point to two clinical solutions. The first is hormonal supplementation with progesterone. Although progesterone has been in use since the 70s (Johnson et al., 1975; Goldstein et al., 1989), the first clinical trial was conducted in the early 2000s and showed that weekly progesterone shots reduced the rate of preterm births by one third and the likelihood of a low birthweight also by one third (Meis et al., 2003). In addition to the shots, progesterone administered vaginally was also found to reduce the risk of preterm births by about 50% and sometimes increase birthweight among women with a short cervix (Hassan et al., 2011; Fonseca et al., 2007). Second, pregnant women at risk of preterm birth, especially those diagnosed with a short/shortening cervix, can undergo a cervical cerclage procedure. Metanalyses have shown that the one-day, outpatient procedure can prolong pregnancy by an average of a month and increase infant survival by 30% (Berghella et al., 2011). Although cervical cerclages were not endorsed by the American College of Obstetricians and Gynecologists (ACOG), about 71% of surveyed specialists believed they were likely to help prevent a premature birth (Ressel, 2004; Fox et al., 2008). Appendix tables A1 through A3 summarize these studies.

Ending the Carve-out and Plan Incentives

In April 2012, the VLBW carve-out ended, making private Medicaid plans responsible for covering all newborns born to enrolled mothers regardless of how sick or costly these newborns were (NY DOH, 2012).[7] This policy change presents an opportunity

[6]Moreover, a transvaginal ultrasound is also used to confirm, above and beyond results from a regular ultrasound, whether the placenta is abnormally placed in the uterus - a diagnosis called placenta previa, also considered a risk factor for preterm birth - although not as clinically alarming as a short cervix or a prior preterm birth (Rao et al., 2012).

[7]The carve-out ended as part of a redesign effort in NY Medicaid meant to reduce spending, partly by ensuring every Medicaid enrollee is "enrolled in care management". This included infants with disabilities, people living with HIV, people with nursing home needs, foster care children, etc. For more details, see a description of proposal #1458 in NYSDOH (2011b).

to understand the effects of fragmentation on the provision of preventative care during pregnancy. In order, however, for this change to increase preventative care and reduce the likelihood of VLBW births, there must be a risk transfer that is large enough to influence the incentives for the insurer.

As discussed above, VLBW newborns require expensive care and the price tag facing FFS for these newborns was high. One can argue, however, that MC plans would not face the same price tag as the FFS program *even for the same newborns*. For one, MC plans are premised on eliminating wasteful care and having more efficient, cheaper networks (i.e., hospitals that negotiate lower prices with insurers). Moreover, there was a generous risk-adjustment system that compensated plans for enrolling expensive enrollees, including sick infants. This means that the FFS spending would be an upper bound for what MC plans would actually pay.

The details of this risk adjustment system are important because they determine how the plan incentives changed with the carve-out ending. The risk adjustment system had three main tenets. First, MC plans received capitated monthly payments that are risk adjusted to the health of the enrollee, including the prematurity of a newborn (Winkelman and Mehmud, 2012).[8] Second, NY Medicaid had a generous stop-loss policy for MC plans (NY Medicaid, 2022). The MC plan was responsible for paying up to $100,000 for inpatient charges per enrollee per calendar year. If the plan was liable for additional inpatient charges, NY Medicaid reimbursed them 80 cents on each additional dollar, until the plan reached a total of $250,000. This leaves the plan with a maximum inpatient ceiling of $130,000 per newborn.[9] The third tenet was a one-type kick-payment that NY Medicaid plans also received at each delivery: one for the mother giving birth and one for the newborn. The newborn one typically ranged between $2,200 to $6,700 depending on the enrollee's county (DiNapoli, 2014).

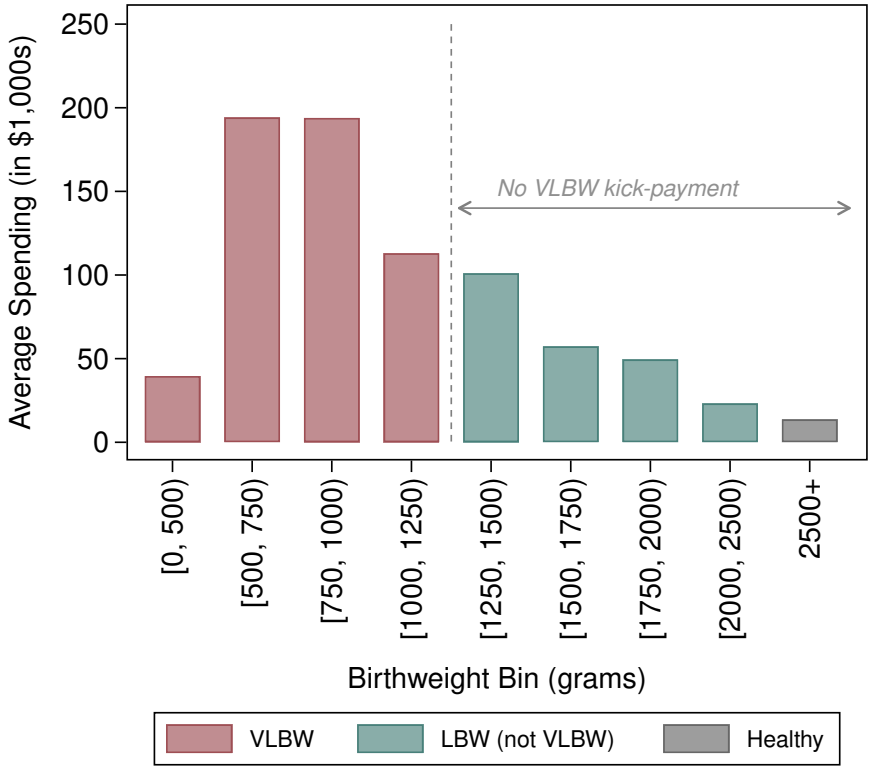
Finally, and most importantly, in 2012 with the end of the carve-out, NY Medicaid added an additional one-time kick-payment called the "Supplemental Low birthweight Newborn Capitation Payment" (which I will refer to as the VLBW kick-payment after) (DiNapoli, 2014). This kick-payment was large in absolute and relative terms - ranging from $68,000 to $105,000 per newborn.

[8]However, risk adjusting these premiums did not make a substantial difference. According to the primary dataset introduced in section 3, the average monthly capitation after the carve-out ended was similar for VLBWs, LBWs, and newborns with a healthy birthweight.

[9]Before 2010, the first ceiling was $50,000 instead of $100,000.

While the new risk-adjustment policy was meant to help MC plans cover the high costs of care for these VLBW infants, one concern is that the size of the kick-payment may have reduced or eliminated financial risk for the plans. Figure 2 uses FFS spending on newborns by birthweight to demonstrate the trade-off that the kick-payment now created for plans wanting to use preventative care. Specifically, medical spending declines with birthweight, making preventative care strategic. However, reducing birthweight too much may result in foregoing the kick-payment. The question thus is which incentive dominated the calculus for MC plans. In the next section, I outline a simple conceptual framework and a simulation exercise to answer this question.

Figure 2
FFS Spending by Birthweight in the Pre-Period



2 Conceptual Framework and Simulation of Plan Incentives

The goal of this conceptual framework is to examine the incentives of MC plans in providing preventative care before and after the carve-out ended and to bring to the forefront any assumptions I make in the process.

In their profit-maximization process, assume that MC plans consider the spending and revenue from each pregnancy and the resulting newborn combined – following the insurer's perspective, this section will use the terms pregnancy and newborn interchangeably. Further, assume that there is a distribution of birthweights from mothers on MC plans and that within these distributions, there are two groups of interest: The first is inframarginal newborns, indexed with I. They are inframarginal because preventative treatment can improve their birthweight (and/or reduce their healthcare cost) but it cannot push them over the carve-out cutoff of 1200 grams. The second is marginal newborns, indexed with M, for whom preventative treatment during pregnancy would push them over the cutoff.[10]

There is, of course, a third prominent group of newborns that are to be born with healthier birthweights but given that the carve-out does not impact them, they do not feature here. Moreover, assume that the MC plan cannot perfectly identify whether a pregnancy would lead to an inframarginal or marginal birthweight, however, they do know their fractions γ_I and γ_M based on births in prior years.

Now, consider the plan's profits when the carve-out is in place. Assume that the net profit from each group j depends on the revenues R_j they bring in from monthly capitation payments and the newborn kick-payments as well as the spending S_j they expend on the pregnancy and the first six months of the newborn's life. Let $S_j \equiv S_j(p)$ where p stands for the preventative care given during pregnancy.

Before the carve-out ended, we have ${}^{\delta S_I}/_{\delta p} > 0$ because preventative care imposes an immediate cost and thus increases the spending on pregnancy. After delivery, the

[10]The extent to which this categorization represents the true effectiveness of preventative treatment across the birthweight distribution is unclear. Existing clinical studies published prior or during the pre-period point to an *average* decline in the number of newborns under 1500 grams but they do not qualify whether this decline is uniform across all newborns under 1500 grams or if it is concentrated at specific parts of the distribution. See Appendix tables A1 through A3 for a summary of these studies.

inframarginal newborns are carved-out. For marginal newborns, preventative care also increases plan spending (*i.e.*, ${}^{\delta S_M}/_{\delta p} > 0$) although it does so through both the immediate cost of the preventative care and the resulting cost of the newborn that would *not* be carved-out and would stay as part of the plan's responsibilities. I can now describe the profit function for the average MC plan with the carve-out as:

$$\pi_{before} = \gamma_I \left(R_I - S_I(p)\right) + \gamma_M \left(R_M - S_M(p)\right) \tag{1}$$

It is then obvious from equation (1) that, with the carve-out in place, increasing preventative care for either group is not profit-maximizing.

Ending the carve-out introduces a few changes. First, the new VLBW kick-payment, $kick_j$, enters into the plan's calculus. For the inframarginal newborns, this kick-payment, $kick_I$, is guaranteed because they would always fall under the VLBW cutoff. For the marginal newborns, however, the kick-payment depends on the level of preventative care $kick_M \equiv kick_M(p)$ because preventative care increases the likelihood of moving across the cutoff and hence reduces the likelihood of collecting the kick-payment. Second, the relation between spending and preventative care also changes. In this world, without the carve-out, preventative care has a cost-saving effect: it reduces the cost of the newborn because they are more likely to be born at term or heavier. Assuming that this cost-saving effect is larger than the immediate cost of the treatment, we then have ${}^{\delta S_j}/_{\delta p} < 0$ for $j = I, M$.

I can then summarize the profit function after the carve-out ended as:

$$\pi_{after} = \gamma_I \left(R_I + kick_I - S_I(p)\right) + \gamma_M \left(R_M + kick_M(p) - S_M(p)\right)$$

with ${}^{\delta S_M}/_{\delta p} < 0$, ${}^{\delta S_I}/_{\delta p} < 0$, and ${}^{\delta kick_M}/_{\delta p} < 0$.

I now proceed to examine the effect of increasing preventative care in the post-period after the carve-out ends. Taking the first derivative with respect to p, I find the equation:

$$\frac{\delta \pi_{after}}{\delta p} = \gamma_I \left(-\frac{\delta S_I(p)}{\delta p}\right) + \gamma_M \left(\frac{\delta kick_M(p)}{\delta p} - \frac{\delta S_M(p)}{\delta p}\right) \tag{2}$$

In other words, encouraging preventative care now increases plan profit but only to the extent that inframarginal cases dominate the enrollee pool. The marginals can make it such that preventative care costs more if the kick-payment forgone from preventative care is larger than the cost-saving effect (i.e. when $\frac{\delta kick_M(p)}{\delta p} - \frac{\delta S_M(p)}{\delta p}$ is negative). We can then re-arrange the terms from equation (2) to find the condition that is necessary for the first-derivative to be positive (i.e., when preventative care is profit-increasing, if not maximizing):

$$\underbrace{\gamma_I \left(-\frac{\delta S_I(p)}{\delta p}\right) + \gamma_M \left(-\frac{\delta S_M(p)}{\delta p}\right)}_{\text{cost-saving effect}} \geq \underbrace{-\gamma_M \left(\frac{\delta kick_M(p)}{\delta p}\right)}_{\text{forgone kick-payment for marginals}} \tag{3}$$

The inequality in (3) now reduces to the two competing effects of preventative care: the cost-saving effect driven from both inframarginals and marginals and the foregone kick-payment effect from the marginals.

For further simplification, assume that preventative care confers to inframarginals cost savings that are a constant $k \in (0, +\infty)$ the benefit it confers to marginals:

$$\frac{\delta S_I(p)}{\delta p} = k \frac{\delta S_M(p)}{\delta p} \tag{4}$$

Given this cost-savings relation and given the two fractions γ_I and γ_M make up the entire population in question, this simplifies the last inequality to:

$$(1 - \gamma_M) \left(-k \frac{\delta S_M(p)}{\delta p}\right) + \gamma_M \left(-\frac{\delta S_M(p)}{\delta p}\right) \geq -\gamma_M \left(\frac{\delta kick_M(p)}{\delta p}\right) \tag{5}$$

I consider the inequality above at two definitions of k that generate intuitive and testable conditions:

- $k = 0$, i.e., when prevenative care only impacts marginals, the inequality reduces to:

$$-\gamma_M \left(\frac{\delta S_M(p)}{\delta p}\right) \geq -\gamma_M \left(\frac{\delta kick_M(p)}{\delta p}\right) \tag{6}$$

which naturally tells us that only marginals enter the calculus and in order for preventative care to be profitable, the cost-savings from this group specifically have to exceed the forgone kick-payments.

- $k = 1$, i.e., when preventative care impacts inframarginals and marginals equally, the inequality reduces to:

$$\gamma_M \leq \frac{-\delta S(p)/\delta p}{-\delta kick_M(p)/\delta p} \tag{7}$$

Where equation 7 puts an upper bound on the fraction of marginals in order for preventative care to be profit-maximizing in the post-period. This upper bound the condition specifies is the ratio between the cost-saving effect of preventative care and the foregone kick-payment when preventative care is used. If the cost-saving effect is larger than the foregone kick-payment then the ratio is larger than one, the condition always holds, and preventative care will always be profit-maximizing. However, the magnitude of cost-savings can be lower than that of forgone kick-payments and preventative care would still be profit-maximizing. Mathematically, this is because the size of cost-savings relative to the forgone kick-payments only has to be as large as that of the share of marginals. Intuitively, it is because cost-savings are generated from marginals and inframarginals while the forgone kick-payments are generated from marginals only.

Now, I consider whether these conditions are likely to hold empirically by simulating the average cost savings holding forgone kick-payments constant and the average forgone kick-payment holding cost savings constant for an MC plan providing preventative care to inframarginal and marginal pregnancies. I focus on these two definitions of k partly because they are easily testable and partly because, to the best of my knowledge, the actual k is not clinically obvious from the medical literature – in other words, we do not know which part of the birthweight distribution is most amenable to preventative care.[11]

Of course, there are other cases besides $k = 0$ and $k = 1$. A value of k that is larger than 1 means that preventative care generates more cost-savings from inframarginal newborns than marginal ones. It is theoretically plausible possibly because

[11]The clinical trial for progesterone showed that the share of newborns who are under 1500 gms decreases by 38% on average. However, it does not show which part of the distribution under 1500 grams that this decline comes from. See appendix section A for more details.

inframarginal newborns are costlier, or because there is more scope for them to gain birthweight, or both. However, as k becomes larger, the left-hand side in equation 3 becomes larger, and the easier it is for preventative care to become profitable. As such, I focus on the more conservative cases where inframarginals and marginals are impacted equally (i.e. $k = 1$) and where only marginals can generate cost-savings (i.e. $k = 0$).

In order to test the conditions in 6 and 7, I first simulate the distribution of MC plan profits in the pre-period given the fraction of marginal pregnancies that plans cover based on their pre-period rates.[12] I vary two parameters, the first is k as just described, and the second is s, a measure of perceived effectiveness, normalized to the effect reported in the progesterone clinical trial. For purposes of brevity, I describe the exact details of how I operationalize this simulation in the appendix and summarize the results here.

Figure 3
Expected Change in Profit

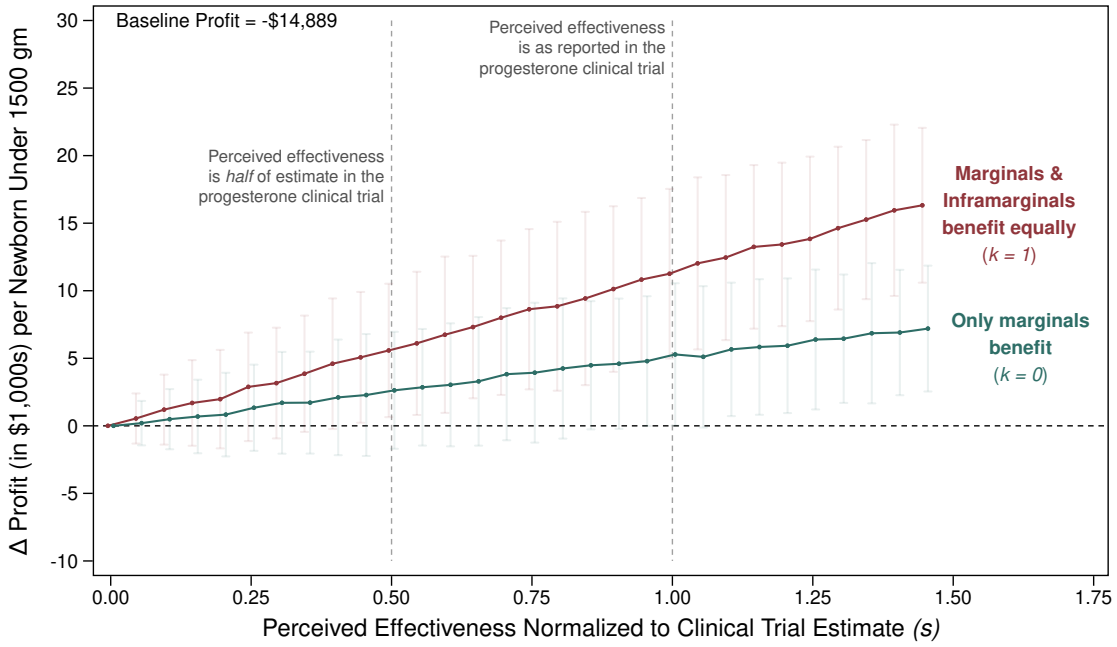


Figure 3 shows average changes in plan profits given potential levels of perceived

[12]Profits are unobserved because spending is unobserved. However, MC spending can be inferred using an adjusted FFS spending distribution, adding capitation, and applying risk adjustment rules.

effectiveness of progesterone, one of the two primary methods for reducing the risk of preterm (and consequently low birthweight) newborns. The primary insight from this simulation is that the kick-payment did not de-incentivize private Medicaid plans from using preventative care. This stands regardless of how effective (or ineffective) insurers perceived preventative care was. It also stands regardless of which part of the birthweight distribution they believed would be impacted. Appendix Table A5 shows the reason. Even with the set of most conservative assumptions – when inframarginals do not contribute to cost-savings ($k = 0$) and when perceived effectiveness is half of that reported in the clinical trial ($s = 0.5$) – the cost-savings are always larger than the foregone kick-payment.

Other reasons – besides this simulated set of results – suggest that ending the carve-out propelled insurers towards prevention rather than collecting the kick-payment. First, over time MC plans stalled in covering VLBWs newborns. Second, although the carve-out was supposed to last for the first six months of the newborn's life, it often shielded insurers from nine to twelve months of the newborn's coverage. I return to these points when I discuss the first stage in the 5 section. But before then, I describe the data and the general methods.

3 Data and Sample

The primary data for this project come from Medicaid Analytic eXtract (MAX), collected and published by the Centers for Medicaid and Medicare. The data contain information on Medicaid enrollees across the United States, their demographic characteristics, how long they were enrolled in Medicaid, and under which plans. The data also include information on the care they receive while enrolled, as recorded by their medical claims. Each claim records diagnoses and procedures if any, identifies who the provider was, and whether the provider billed FFS or an MC plan. These claims are available for both the private MC plans and the public FFS program.[13]

[13]Two missing pieces of information in this dataset are reimbursements from MC plans to providers as well as rebates on any prescription drugs for both MC and FFS.

3.1 Sample Construction

I use this dataset to construct my primary sample - a sample of NYC Medicaid enrollees who had a pregnancy episode while enrolled and who did so between 2008 and 2015. The chosen study period allows me to examine preventative care at least three years before the announcement of the carve-out ending, four years before it actually ends, and four years after.[14]

The sample was also subject to some exclusion criteria. Specifically, I exclude pregnant enrollees outside a typical reproductive window – under 14 or over 45. Second, I exclude dual eligibles who are insured by both Medicare and Medicaid because Medicare serves as their primary payer. Third, I exclude enrollees who churn between Medicaid FFS and Medicaid MC during their pregnancies because any MC enrollment entails exposure to the "treatment": a plan that experiences a change in fragmentation and an ensuing change in incentives to provide preventative care. Finally, I exclude abortions. I do this last step for two reasons: first, because unwanted pregnancies are unlikely to receive preventative treatment especially given that at the time period abortion was legal and covered. In New York City, where the majority of the pregnancies were taking place, abortion was covered by Medicaid regardless of medical necessity. Second, as appendix section B describes in detail, the timing of abortions appears to be incorrectly coded, especially when compared relative to public data from the Department of Health information.

I also create a secondary sample from the data, where I matched each mother with a recorded live birth delivery episode to her newborn(s). In order to do that, I matched mothers and newborns based on a family case number assigned by Medicaid and the date of birth. On a given study year, the percentage of pregnancy episodes ending in deliveries that were successfully linked to the newborn(s) ranged between 82 and 87%.[15] Among those that were successfully linked, I could extract the outcomes of the newborns during birth.

[14]The study period ends in September 2015 to avoid coding discrepancies resulting from the switch from ICD-9 to ICD-10 for claims.

[15]These rates are less than 100% because the family case number number changes over time and does not change consistently between mothers and newborns. Newborns linked to more than one mother were removed.

3.2 Sample Description

The resulting primary sample had 503,734 pregnancy episodes belonging to 384,113 enrollees. As shown in Table 1 below, the majority of mothers in the pre-period episodes identified as either Hispanic (32%) or African American (35%) and lived in NYC.

I also construct a measure of risk of an adverse pregnancy outcome, which can be a preterm birth, a low birth weight newborn, or a miscarriage/still/ectopic pregnancy. To do this, I combine these race and ethnicity data with other demographic information like age, three-digit zip code, and eligibility category, along with clinical risk factors, like diabetes and weight issues, to predict a risk score. Crucially, I identify these risk factors from the year prior to the pregnancy. This timing is important because any diagnoses during the pregnancy are potentially a function of monitoring, which is in itself a function of insurer incentives and the end of the carve-out. All these predictors were then input into a Boosting algorithm which in turn produces a predicted risk.[16] The average pregnancy episode in the baseline period had a 28% chance of a preterm birth, a low-birth weight newborn, or a non-live birth end of the pregnancy.

Table 1 shows some differences in the demographic and health characteristics of these episodes across MC and FFS enrollees: episodes covered by MC tended to have fewer African American women. As far as pregnancy risk goes, however, MC plans appeared to be responsible for episodes with a higher chance of having any risk factor (31% instead of 19% in FFS), although predicted risk scores appear to be roughly similar across the two groups.[17]

As mentioned earlier, NY had passed mandates that had transitioned most of the NYC Medicaid enrollees into the MC System by the time of the study. The remaining two groups were beneficiaries who had to stay on FFS and beneficiaries who could choose to stay on FFS. The existence of the latter group, at least in theory, is akin to selecting into treatment, making it worth examining the extent to which they feature

[16]The training data was made up of 70% all NYC pregnancies in the pre-period. The exact algorithm and why it was chosen is described more in appendix section C.

[17]It is worth remembering here that MC plans have an incentive for upcoding diagnoses to generate larger reimbursements. As such, some of the difference in levels in risk factors may be a product of that.

in the sample.

Table 1
Primary Sample Characteristics in the Pre-Period

	(1)	(2)	(3)	(4)
	All	FFS	MC	Difference
Demographics				
Age	27.4	26.8	27.7	0.838
NYC	0.696	0.625	0.727	0.102
White	0.485	0.406	0.520	0.114
Black/African American	0.349	0.386	0.332	-0.054
Asian	0.078	0.053	0.089	0.036
Other Race	0.057	0.080	0.046	-0.033
Hispanic or Latino	0.318	0.368	0.296	-0.072
Risk of Adverse Outcomes				
Risk score (0-1)	0.281	0.298	0.280	-0.018
Any risk factor	0.273	0.190	0.310	0.120

Note: Risk score shown is based a boosting predicting algorithm among the subset of enrollees who were fully-enrolled in Medicaid in the year prior to the start of their pregnancy episode. The risk factors used in the prediction are high blood pressure, diabetes, under/over weight, tobacco use, alcohol use, drug use, chronic kidney disease, autoimmune disease, HIV, urinary tract infections, asthma, thyroid, cardiovascular disease, and mental health disorders.

There are four major subcategories who could choose to be on MC (WNYLC, 2021). The first is Native Americans, but they constitute less than 1% of the pregnancy episodes and I run robustness checks removing them from the sample. The second is children with developmental disabilities who received some assistance from Medicaid (DOH, 2013). Although I cannot identify this group in my data, I can remove enrollees who are 17 years or younger at the time of delivery, since enrollees had to be at most 17 years old to qualify for this category. This group contributes about 2.4% of the sample episodes. The two remaining subcategories are adult enrollees with developmental disabilities and adult enrollees who experienced a traumatic brain injury. I cannot identify these two last subcategories in my data - however, their expected small proportions in Medicaid enrollees in general and among female Medicaid enrollees in reproductive age makes their existence less of a concern.

As for the groups that had to stay in FFS and that make up my control group,

these are Medicaid beneficiaries who hold third party insurance, recipients in the Medicaid spend-down or excess-income program who had slightly higher incomes than the Medicaid cutoffs; and beneficiaries who only qualify for limited Medicaid benefits (WNYLC, 2021). Finally, Medicare-Medicaid dual eligibles are theoretically in this group who have to stay on FFS, but they are not in the sample because their Medicare covered services are not observable in their Medicaid claims.

4 Methods

4.1 Empirical Strategy and Identification

The setup discussed so far lends itself to a simple difference-in-difference: I compare the insurer's preventative care provision for its pregnant enrollees before and after the carve-out ends. To remove secular time trends in the use of these medical technologies, I use episodes covered by FFS as a comparison group. The primary difference-in-difference specification is thus:

$$Y_{it} = \beta_0 + \beta_1 MC_{it} + \beta_2 After_t + \beta_3 MC \times After_{it} + X_{it} + u_{it} \tag{8}$$

where Y_{it} is an outcome of pregnancy episode i which ended at time t; MC_{it} is a dummy indicator for whether the enrollee with the pregnancy episode was covered by MC as opposed to FFS; and $After_t$ is a dummy for whether the end of the pregnancy episode took place in the post-period after the announcement of ending the carve-out (i.e. starting 2011). I control for the enrollee's race, ethnicity, and age using vector X_{it} with dummies. Given that enrollees in the same plan will likely experience similar policies and may have correlated error terms, I cluster the standard errors on the plan level.

The coefficient of interest on the interaction of the last two variables $MC \times After_{it}$ represents the premiums of having a pregnancy after the carve-out ended with an insurance setting that has less fragmentation. In order for this coefficient to isolate a true causal effect of fragmentation, two assumptions have to be satisfied. The first is that there are no anticipation effects — i.e., MC plans did not adjust their behavior prior to the end of the carve-out — and the second is the parallel trends assumption.

With respect to anticipation, there are four dates relevant to how early MC plans knew about the end of the carve-out: In February 2011, more than one year before the carve-out was to end, a panel hired by the governor of NY, Andrew Cuomo, voted to accept the reforms proposed by a group called the Medicaid Redesign Team, one of which was the proposal to *carve-in* very low birth weight infants into Managed Care (Kaplan, 2011).[18] NY Medicaid then went ahead and applied for federal approval in April 2011. Two months before the scheduled date, in February 2012, the state agency formally announced the change to MC plans, contingent on obtaining federal approval. Finally, in March 2012, federal approval came in and the carve-out was no longer in place by April 1st, 2012.[19] Because it is unclear on which of these four dates MC plans learnt about the change and, critically, *believed* it is taking place, I include 2011 as part of the post-period to capture any anticipatory effects in my event study design.

The event study also (indirectly) tests the second assumption: that trends in outcomes among pregnant enrollees on MC plans must have developed in parallel to those on FFS absent the treatment. Specifically, I run an event study with the following specification:

$$Y_{it} = \alpha_0 + \alpha_1 MC_{it} + \sum_{j=-3}^{4} \alpha^j MC_{it}^j + \gamma_t + u_{it} \tag{9}$$

where each α^j shows the effect of having a pregnancy covered by MC at time j relative to 2010, the year preceding the announcement. I define a time period as a calendar year in order to avoid any idiosyncratic month-to month or quarter-to-quarter noise in my outcomes.[20]

[18] It is unlikely that the proposals were circulated among MC plans much earlier than February 2011, given that the Medicaid Redesign Team which came up with the proposals was convened on January 7th, 2011 (NYSDOH, 2011c).

[19] See the timeline for the 1115 waiver amendment here under amendment #4 on Medicaid.gov. See the formal announcement on NY Medicaid's website here.

[20] Additionally, because some of these preventative care measures are rare, monthly and quarterly averages may be too small to disclose.

4.2 Preventative Care Outcomes

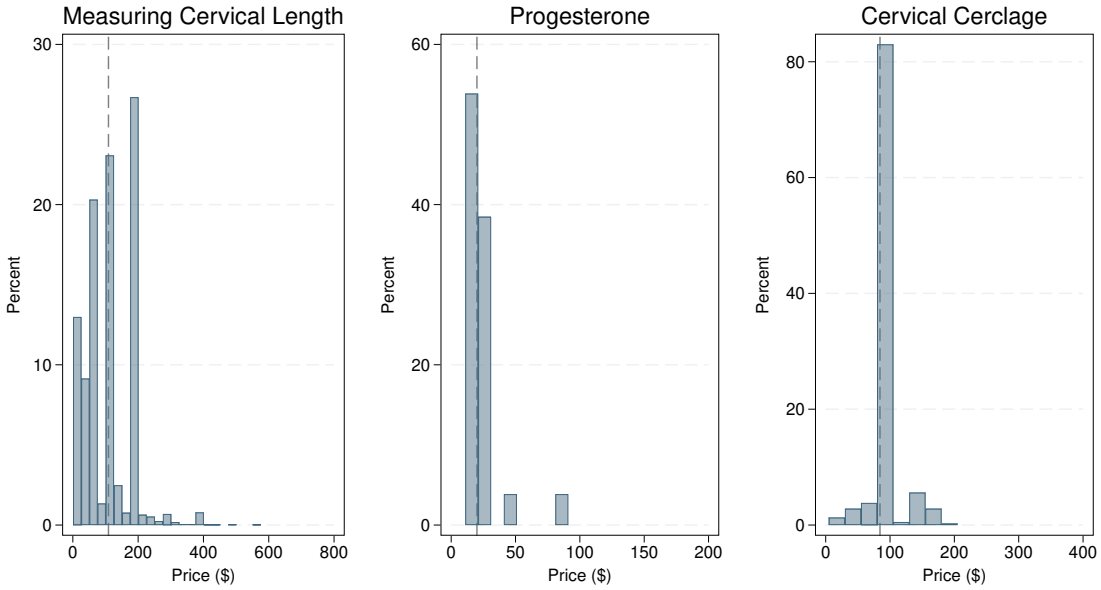
My primary outcomes are treatments that are meant to monitor or reduce the risk of a preterm birth and that can be identified in billing data. My indicator for monitoring is whether a pregnancy underwent a transvaginal ultrasound to measure cervical length. I also identify whether the pregnancy had a diagnosis of any risk factors of preterm births, including a short cervix. Other risk factors identified from the medical literature include having a placenta previa (an abnormally positioned placenta, diagnosed from a transvaginal ultrasound), vaginal bleeding or infections, urinary tract infections, gestational diabetes, high blood pressure, and blood clots (NICHD, 2017). To examine take-up of preventative treatments, I consider changes in progesterone supplementation and the cervical cerclage procedure, which are described in section 1.3.

An important question here is whether these methods of monitoring risk or prevention are costly. The more expensive they are, the less readily MC plans would use them, especially when they are not responsible for covering VLBW newborns. Figure 4 plots the distribution of reimbursements that the FFS program paid to physicians for monitoring the cervical length, using progesterone shots, and placing a cerclage.[21] As shown, paid prices are low. The average price is about $100 or lower for all three of these procedures. Although spending on progesterone will be higher given that a course of supplementation will include as many as 20 shots per pregnancy, the overall spending on the treatment would still be many orders of magnitude lower than the costs of covering preterm and LBW newborns discussed earlier. Worth noting here is that MC prices are typically undisclosed and it is possible that MC plans face higher costs than FFS for these treatments. Given the high costs of VLBW newborns, however, as well as government audit reports showing that MC plans in NY Medicaid paid only slightly more than FFS (GAO, 2014), one can safely assume that the costs of these procedures were not prohibitive.[22]

[21] Note that the progesterone prices do not include rebates - if any.

[22] As will be discussed in section 6, the price of progesterone may be an exception.

Figure 4
Costs of Preventative Care Treatments in the Baseline Period



Based on FFS prices in the pre-period. The vertical line indicates average price. Progesterone prices do not include rebates and represent the price of one shot.

4.3 Cream-skimming

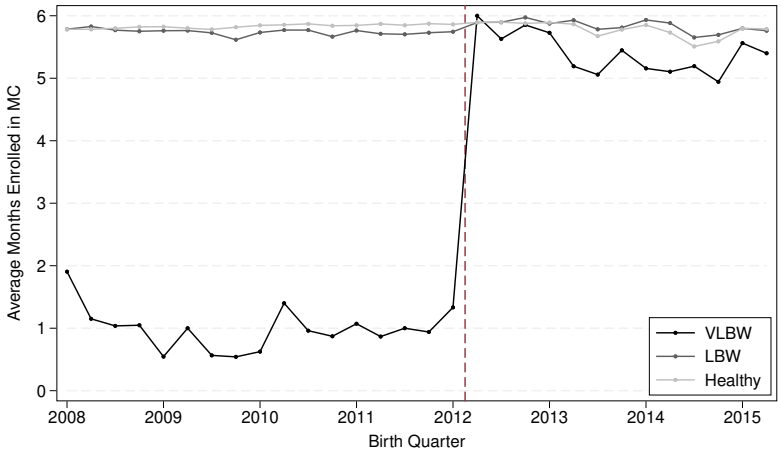
As implied throughout, an obvious way in which private plans may avoid the costs of expensive newborns is to provide preventative care to their enrollees during pregnancy. Another way to do that is to avoid enrolling high-risk pregnant women from the outset. Although "guaranteed issue" is a feature of NY Medicaid making it illegal for plans to reject enrollees based on their health or profitability, a plan can choose to enroll in some counties and not others and it can choose how intensely to advertise in different locations and to different populations based on their health levels (NY DOH, 2011c). Moreover, evidence from other state Medicaid programs suggests that MC plans sometimes dropped high-risk pregnant enrollees on the premise of failing to listen to doctor's orders or missing appointments (Kuziemko et al., 2018) – this is a scenario that could happen in NY, at least to the subset of women who enrolled because of their pregnancy and who had three months to leave their plan before they were considered "locked-in" (WNYLC, 2021). Given this possibility, I examine different proxies of plan selection like the overall risk of enrollees and their continuity

under the MC system and in the same MC plan through pregnancy.

5 Results

5.1 First Stage

Figure 5
First Stage MC Coverage in the First Six Months

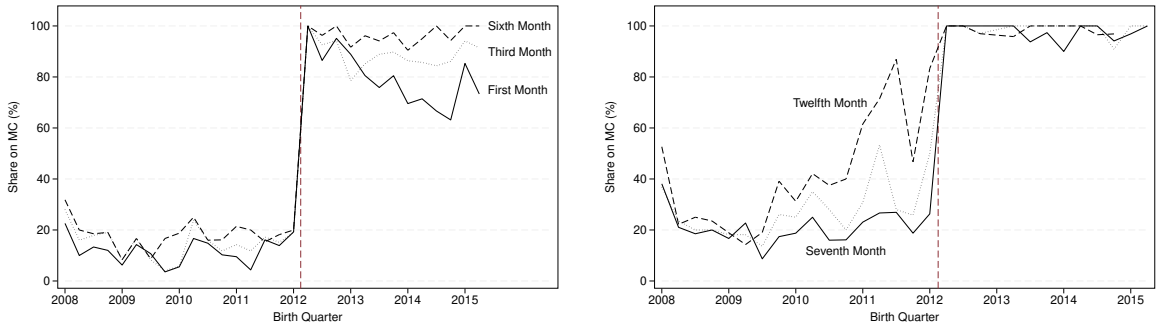


A preliminary first step is to understand whether the carve-out succeeded in placing VLBW infants born to MC mothers on MC. Figure 5 below suggests that the answer is affirmative. The figure plots the average number of months that VLBW newborns were enrolled in an MC plan during their first six months of life, conditional on their mother also being covered by an MC plan. In April 2012, this share increases discontinuously from slightly under one month to the full six months in the second quarter of 2012 when the carve-out ends.[23]

[23]Why this mean is not exactly zero before the carve-out ends is a data artifact. Specifically, the birthweight bin that is closest to the cutoff includes newborns up to 1250 grams rather than 1200 grams where the cutoff is. As such, it is possible that newborns between 1200 and 1250 grams who would have been assigned to MC raise the average to a number that is larger than zero.

Figure 6
Plans Stalling on Covering VLBWs

(a) VLBW MC Coverage, First Six Months **(b)** VLBW MC Coverage, Second Six Months



Although this discontinuity suggests a transition of VLBW newborns to MC, the effect appears to have decreased over time. Panel (a) in figure 6 plots the share of VLBW newborns with MC mothers who are covered by an MC plan in the first, third, and sixth month of their life. During the sixth month, almost 100% appear to be covered by MC throughout the post-period. However, the first month MC status decreases from 100% in 2012 to 60% by the end of 2014, suggesting that MC plans were able to postpone covering the VLBW newborns for the first few months. Although this delay implies that the bite of ending the carve-out has declined over time – especially given that the first few months after birth are the most expensive, MC coverage in this latter period remains well above 60% leaving a substantial first-stage effect.[24]

Importantly, however, this delay in coverage possibly points to plans stalling on covering VLBW newborns, taking longer and longer before taking them on. Panel (b) in figure 6, which plots the same outcome – share of VLBWs covered by MC by month – suggests another reason why this may be the case. The carve-out appeared to be more generous in practice than in theory. While VLBWs were supposed to return to MC by month seven, very few actually did. In fact, until 2011, half of VLBW newborns stayed under FFS for their entire first year. Even VLBWs born

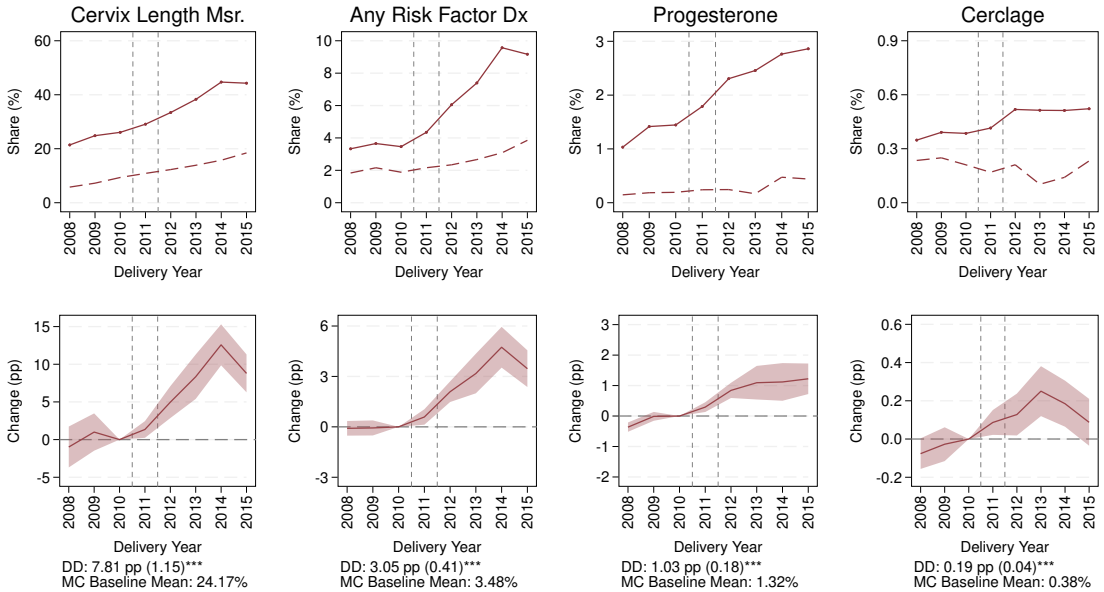
[24]How MC plans were able to postpone coverage is unclear. One theoretical reason is that they were able to use exemptions and exceptions for the MC mandate more effectively after the end of the carve-out. However, this appears to be empirically untrue at least with respect to rates of Native American descent, being dually eligible for Medicare, having third party insurance, having restricted Medicaid benefits, and being developmentally disabled (trends are not shown due to the small numerators in these attributes).

on the quarter right before the carve-out ended spent their first nine months under FFS. This deviation from policy suggests that ending the carve-out transferred the financial responsibility for potentially longer than the first six months.

5.2 Primary Result: Effects on Monitoring and Prevention

Given the first stage, I now ask whether MC plans increased their provision of preventative care. Figure 7 below shows the primary results. First, the rate of cervical length monitoring increased somewhat discontinuously among MC enrollees and not FFS enrollees starting 2011, the year when the carve-out was announced to be ending, by about 7.81 pp. This represents a 32% increase relative to the MC pre-period rates. *Because the MC and FFS group have a substantial level difference in the use of these treatments and because the behavior of the MC plans is the primary object of interest, all the relative estimates will be presented relative to the MC pre-period means.* As shown in appendix figure A11, this increase in transvaginal ultrasounds is largest in the second trimester where monitoring is more discretionary.

Figure 7
Monitoring and Prevention



Notes: Solid lines indicate trends for private plan enrollees covered by MC plans. Dashed lines indicate trends for public plan enrollees covered by FFS.

Consistent with this increase in monitoring is a similar increase in diagnoses of

immediate risk factors of preterm birth (i.e. a short cervix or a placenta previa). As shown in the second column of figure 7, the share of enrollees who received one or more diagnoses increased by about 3 pp, almost doubling the baseline mean of 3.48%. When examining which of the two risk factors increased, I find that diagnoses for both a short cervix and a placenta previa increased (by 0.74 pp or 93% and 1.49 pp or 88%, respectively) as seen in Appendix figure A12. The increase in detection of these risk factors are consistent with the monitoring results, given that a placenta previa is typically diagnosed during the second trimester and more accurately diagnosed with a transvaginal ultrasound (Rao et al., 2012). Naturally, a short cervix is detected when the cervix length is measured. I also examine additional risk factors and find large increases in diagnoses of vaginal infections, diabetes, hypertensive disorders, and urinary tract infections.[25]

Finally, I examine whether the eventual use of preventative treatments increased after the carve-out ended in the two rightmost columns of figure 7. In the pre-period when the carve-out was in place, very few women received these treatments. This, however, is consistent with the fact that not all women were at high risk of having a preterm birth and even if they were, not all qualified for the treatment (for example - women pregnant with twins or triplets do not qualify for progesterone supplementation or cerclages even though they are considered high-risk (Meis et al., 2003; Roman et al., 2020)). It is also consistent with rates of progesterone use among Medicaid enrollees in other states, which is similarly low (Orsulak et al., 2015).[26] As it pertains to the effect of ending the carve-out, the figure shows that take-up rates among MC enrollees remained low, flat, and parallel to FFS enrollees, until 2012 for both progesterone use and cerclages. The difference-in-difference coefficients suggest that the end of the carve-out has increased the take-up of these treatments by 1.03 pp (or 78%) for progesterone and another 0.19 pp (50%) for cervical cerclages.

[25] Worth noting here is that these measured increases in diagnoses do not necessarily reflect a higher rate of detecting risk factors. They could instead reflect more intensive coding, especially if MC plans aspired for larger reimbursements from risk adjustment in the post-period. However, there is the question of why plans had not tapped into this source of income in the pre-period.

[26] The study shows that Medicaid enrollees in Louisiana had a progesterone use rate of 0.7% in 2013. In 2011 and 2012, the rates were even lower.

5.2.1 Robustness and Placebo Checks and Other Control Groups

I then estimate these difference-in-difference coefficients in a variety of robustness checks, clustering on different standard errors; adding demographic controls, plan fixed effects, and county fixed effects; dropping partially treated enrollees, enrollees who could select into FFS and MC, and low-risk groups; and adding provider fixed effects.[27] As appendix figure A15 shows, the changes maintain their statistically significant levels and their large magnitudes, with the noisiest coefficients reflecting cerclage increases.

I also run two types of placebo tests to confirm that the effects are not generated by chance. The first randomizes MC assignment among the sample episodes, while maintaining the true distribution of MC & FFS in a given year. I repeat this step 1000 times for each of the four outcomes above and plot the distribution of coefficients - in a spirit similar to Fisher's randomization inference methods (Imbens and Rubin, 2015). I then report the share of coefficients generated from the randomized MC assignment that are as large or larger in absolute value to the true coefficient. For the first three outcomes; monitoring, diagnosis of risk factors, and progesterone use, the results in appendix figure A17 show that less than 0.1% of the placebo draws generated an effect that is as large or larger than the estimated effect, meaning that it is highly unlikely that the estimates was generated by chance.[28]

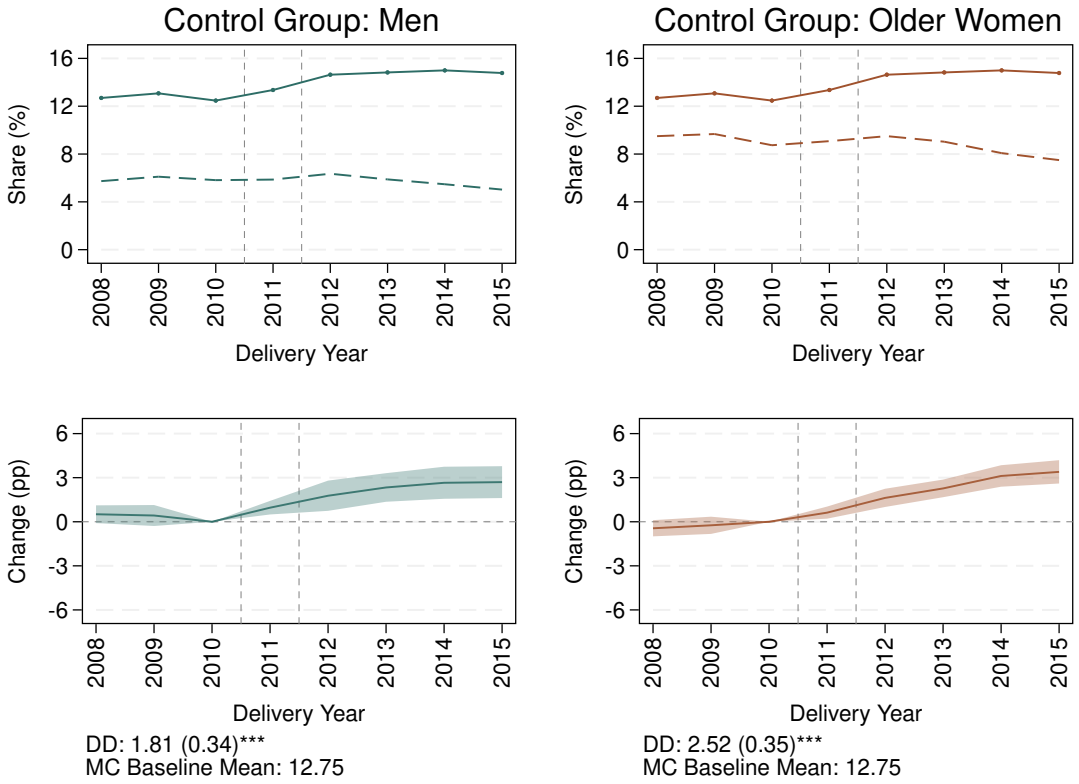
The second type of placebo check constructs a fake timing for the end of the carve-out. Specifically, I focus on the pre-period, dividing it into quarters, and assigning each quarter to be the treatment date, and then comparing outcomes before and after among MC and FFS enrollees. Any positive effects in these placebo tests would indicate the presence of secular time changes: MC enrollees were more likely to experience this type of healthcare over time relative to FFS enrollees and, as such, the difference-in-difference estimates would be over-estimated if not fully accounted for by these secular trends. The results in appendix figure A18 show a consistent pattern. Although there is evidence of some differential trends, the estimated effects

[27] To construct provider dummies, I assigned each pregnancy the modal provider with the highest number of claims, with the restriction that these claims had to take place in an office, a clinic, or an outpatient hospital setting and that the physician had to be either a primary care physician, internalist, family medicine provider, obstetrician, gynecologist, or a maternal fetal specialist.

[28] In fact, the only draw that was equal to or lower than the true coefficient is the draw of the true coefficient itself yielding a probability of 1 out of 1000 draws for each of the four outcomes.

– for the first three outcomes – are always at least two orders of magnitude larger than the largest of these placebo effects. For the cervical cerclage outcome, the placebo effects can be as large as the estimated effects given the true timing of the end of the carve-out.

Figure 8
Using Men and Older Women on MC in NY as Control Groups



Notes: Solid lines indicate trends for pregnant enrollees in the primary sample enrolled under MC plans. Dashed lines indicate trends for MC enrollees in the two respective control groups.

Finally, I use other control groups. FFS enrollees represent a suitable control group insofar as they are low-income, Medicaid beneficiaries, living in New York, and receiving care from providers who are subject to the same regulatory environment as those treating MC enrollees. However, FFS enrollees have two weaknesses: First, they may be demographically different as described above. Second, they do not capture any co-occuring changes in the MC environment beside the end of the carve-out. Both of these points mean that even though FFS enrollees have similar trends in take-up before the carve-out ended, they may not accurately reflect how MC enrollees would

have responded had the carve-out continued (i.e. the parallel trends assumption may not have held).

To overcome any potential issues, I examine whether the results hold when using placebo MC enrollees: i.e. MC enrollees in NY who were not affected by the carve-out such as men and older women who fall outside of the typical reproductive window. While these groups are suitable control groups because they have experienced secular changes that took place in New York's MC plans, they are not suitable in the sense that they simply cannot receive pregnancy-related care and, as such, are incomparable when it comes to the four primary outcomes discussed earlier. As such, I use a more general outcome: the number of office visits which can capture any non-carve out related changes to care that MC enrollees also experienced. The trends in the top row of figure 8 shows that the average number of office visits per true MC-covered pregnancy episode increased discontinuously starting 2011 and maintained this increase throughout the post-period. The dashed lines indicating the number of office visits per 10-month episodes for men and older women did not undergo the same increase, leading to an effect size of 2.13 visit per pregnancy (or 14%) when comparing to men and 2.52 visits (20%) when comparing to older women.[29]

Although MC enrollees in other states are not subject to the same regulatory environment as those in NY, they also make a useful control group because they can capture how private Medicaid plans covered pregnancies during the study period. As such, I use information on pregnant enrollees in MC plans in nine different states to construct yet another control group. The results, shown in appendix figure A20, suggest that NY plans exhibited particularly higher prevention in the post-period. [30]

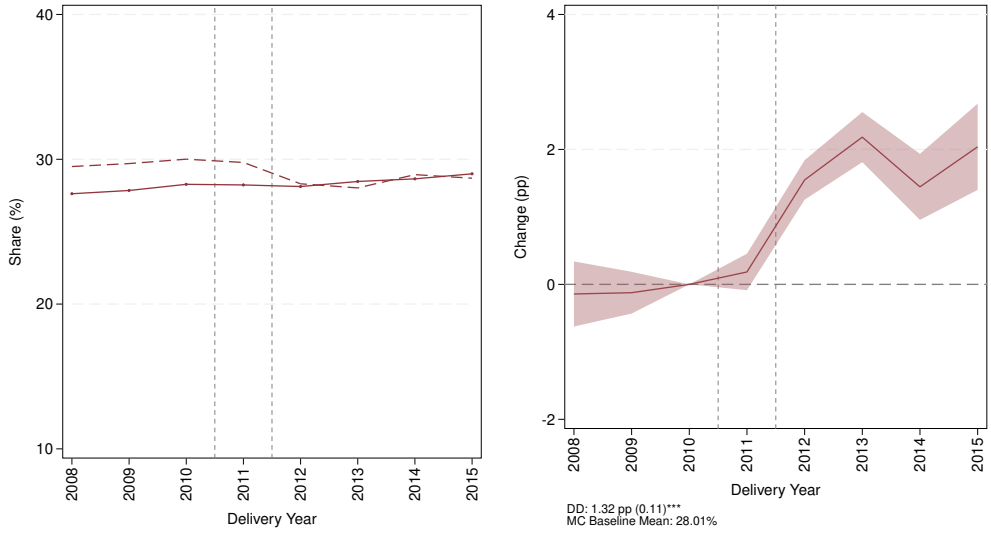
[29]To construct these placebo pregnancy episodes among men and older women, I first obtained the universe of these two groups who met the same inclusion criteria as in the true MC sample (e.g. no duals, no churn between MC and FFS, etc). Among each of the two groups, I then randomly selected the same number of true pregnancy episodes. I also kept mimicked the proportions of live births and non-birth pregnancies as in the true MC sample. For each type of placebo pregnancy (live birth vs non-live birth), I maintained the average duration found in the true sample.

[30]The nine states are Arizona, Delaware, Michigan, Minnesota, New Mexico, New Jersey, Rhode Island, and Wisconsin. The steps used to choose these states are described in the appendix in section G.

5.2.2 Ruling Out Competing Hypotheses

Prevention and Composition Changes One potential reason explaining the documented increases in monitoring and prevention is the end of the carve-out. A competing alternative is that the type of Medicaid enrollees changed differentially after the carve-out ended by treatment group. A plausible, and a concerning scenario for identification, is if MC enrollees became sicker (or FFS enrollees became healthier) meaning that any additional preventative care given in the post-period would be driven by this change in *clinical need* for preventative care rather than by the financial considerations of the insurers. This is particularly plausible given that many enrollees were slowly being transitioned from FFS to MC as part of the efforts that ended the carve-out (NYSDOH, 2011a).

Figure 9
Change in Ex-Ante Risk of Pregnant Enrollees



To examine this hypothesis, I first examine whether the predicted risk score, discussed earlier in section 3.2, deferentially changed in the post-period. Figure 9 shows that this is indeed the case.[31] However, it is *not* the case that MC enrollees became sicker and thus were more likely to require care. Instead, FFS enrollees became healthier (i.e. exhibited risk scores that were 1 to 2 pp lower on average). Although this leads to differential changes across groups over time, FFS enrollees

[31]The Appendix section H also shows the change in risk scores predicted with random forest.

becoming healthier in the post-period cannot explain why MC enrollees were more likely to experience heightened monitoring and prevention as shown by the trends in figure 7.[32]

Second, I confirm the difference-in-difference estimate is not completely driven by the differential changes in risk, I repeat the primary analysis while controlling for the risk score. I also repeat it within a subsample with no reported risk factors another time. Appendix figures A23 and A24 both show the estimates do not change substantially. As a final check, I also compare my primary outcomes among all NY enrollees, FFS and MC included, relative to Medicaid enrollees in the control states, as described above. To the extent that the effects from the main specification are driven by movement of risky enrollees from FFS to MC, NY should not see any change in aggregate outcomes relative to other states. However, the results in appendix figures A25 through A26 show that NY enrollees, in aggregate, experienced increases in prevention relative to the other control states. As such, these checks suggest that the estimates are not explained away by the underlying risk decreasing among FFS enrollees or increasing among MC ones.

Institutional Changes in Progesterone Treatment The results above showed large changes in progesterone use - however, a full understanding of the larger context of progesterone is important. Progesterone supplementation can take place through many forms, the most common of which is the weekly injection of a substance called 17-Alpha-Hydroxyprogesterone Caproate, a treatment regimen that had existed and been used since the 1970s (Johnson et al., 1975; Goldstein et al., 1989). In NY, as in many other states, a provider who prescribed progesterone would order it from a compounding pharmacy (hence, the term "compounded" in compounded progesterone), pay for it, and then bill Medicaid. In the case of FFS, Medicaid only reimbursed the provider for the drug acquisition cost, typically $15-$100 per weekly shot.

This continued to be the status quo until 2011 when a pharmaceutical company took the same chemical substance to FDA clinical trials. After showing clinical effec-

[32]There are many good reasons to be skeptical about this constructed risk score. To summarize its weaknesses from appendix section H, its ability to predict whether a pregnancy will end with an adverse outcome is only 24% higher than randomly selecting pregnancies and it largely predicts miscarriage over preterm or VLBW births. However, these two should affect both MC and FFS similarly. What could induce bias is that the model is disproportionately trained on pre-period MC rather than FFS (reflecting the skewed distribution). This may explain the decline in average FFS risk in the post-period, which the model was not trained on.

tiveness, the company sold its branded version of progesterone – called Makena – on the market with a price tag of $800 to $1,500 per shot, leading to controversy among advocacy groups and clinicians (Meis et al., 2003; Reichmann, 2012; Cohen et al., 2011). In response and in a deviation from its history of discouraging compounded drugs if approved branded alternatives were available, the FDA issued a statement encouraging the continued use of the compounded version of progesterone (Patel and Rumore, 2012).[33] Similarly, New York Medicaid also encouraged its providers to keep using compounded progesterone and encouraged MC plans to place their prior authorization requirements on Makena, although the FFS program continued to cover it and MC plans were still allowed to cover it if they wished (NY DOH, 2011a).

Despite the price tag and the blessing of both the state of New York and the FDA to continue using the compounded version of progesterone, MC plans started using Makena. Table 2 separates the earlier progesterone results into compounded progesterone in the first column and Makena on the second.[34] The results show that Makena increased from almost no use (since it was not available until 2011) to 0.17% of the population, only slightly smaller than the increase in compounded progesterone (0.23 pp).

These finding raises the question of why MC plans would pay for an expensive drug when a cheaper one is not just available but also encouraged. The majority of MC plans in NY were operated by large insurance companies and, even if profit-maximization is their not objective function, they would not want to threaten their bottom lines. One possibility may be that – unlike FFS which only reimbursed the acquisition cost – MC plans paid their providers a higher cut for administering more expensive drugs. In this case, clinicians treating MC enrollees would prefer administering Makena in order to recoup a higher revenue, a pattern that is well documented in other settings (Jacobson et al., 2010). Provider reimbursements by MC plans are generally kept private and as such, I cannot verify whether this hypothesis is empirically true. It is highly unlikely, however, that MC plans who are largely profit-maximizers and who had been involved in "managing care" for many years at that point would set up this payment system knowing that the two drugs are chemically

[33] That the FDA encouraged using a compounded drug even when a branded version was available was such a rare case in FDA history and Makena manufacturers responded by suing the FDA (Yukhananov, 2012).

[34] In this analysis, Makena is identified using the NDC while the compounded drug was identified using both the HCPCS code and the absence of a Makena NDC code.

equivalent.

Instead, a more plausible theory points to liability given the safety issues of compounded drugs that came up around the same period. Specifically, a contamination in compounded injections from a large pharmacy in Massachusetts caused a deadly outbreak in 2012 and was followed by inspections and closures of other compounding pharmacies as well as news-publicized criminal persecutions (Goodnough, 2012; U.S. Attorney's Office, 2021). In response, many large providers switched to hospital-compounded drugs or ended their use of compounded drugs where non-compounded alternatives were available (WBUR, 2012).[35] Consistent with this theory is the quick adoption of a vaginal form of progesterone, as shown in the last column in Table 2. Although vaginal progesterone was not approved by the FDA for a miscarriage/preterm indication, physicians started prescribing it off-label, after a clinical trial showed some promising results in 2011 for high-risk groups (Hitt, 2012; Hassan et al., 2011).[36] As such, it is possible that concerns over liability caused MC plans to switch to Makena despite its cost.

Table 2
Effect of Carve-Out End on Progesterone Use, by Type

	(1)	(2)	(3)	(4)
	Existing Types		New Types	
	Compounded	Oral	Branded	Vaginal
MC X Post-Period	0.23***	0.57***	0.17**	0.11***
	(0.04)	(0.14)	(0.07)	(0.02)
Baseline MC Mean (%)	0.13	1.19	0.00	0.00
Pre-Trend p-value	0.052	0.005	–	–
Observations	489,927	489,927	489,927	489,927

Note: Compounded progesterone shots refer to 17P; oral progesterone refers to branded and generic oral progesterone tablets. Both of these types had existed to the end of the carve-out. The branded progesterone refers to Makena shots, and vaginal progesterone refers to progesterone suppositories, branded as crinone or endometrin. See Appendix figure A27 for trends in each type.

[35] There is some journalistic work showing that concerns about the safety of compounded 17P predated these Massachusetts controversy, partly because the manufacturers of Makena published results showing the safety of compounded progesterone is questionable. It is unclear whether these results swayed plans or providers in their use of compounded progesterone given the partiality of the researchers (Bogdanich and Tavernise, 2012).

[36] Pre-rebate estimates from the State Drug Utilization Database suggest that an insert of vaginal progesterone cost about $250-300 per patient in 2012.

Although unexpected, the adoption of Makena in the post-period does not by itself threaten the validity of the estimated results. What *would* threaten the validity is if the introduction of Makena, which took place only a year before the end of the carve-out, rather than the end of the carve-out itself spurred prevention efforts. One possibility may be that the introduction of Makena, and the controversy surrounding its price and safety, served as a reminder for physicians to use progesterone. Consistent with this theory is the last column in Table 2 which shows that it was not just the compounded drug or Makena that increased; MC plans also adopted the use of vaginal progesterone, which had been approved for a preterm and miscarriage indications in 2011. However, it is not clear why this reminder of the availability of drugs for preterm birth risk would not impact providers caring for MC patients *as well as* FFS patients.

Another theory suggests that Makena was perceived as the safer option for progesterone and that absent changes in the carve-out, just having Makena on the market would have encouraged providers to prescribe progesterone more frequently. This theory, however, would have required that clinicians wind down the use of compounded progesterone, but as described in Table 2, the share of MC enrollees receiving it doubled. Moreover, this new-found safety would not have generated the observed increases in monitoring, in vaginal progesterone, and in cerclages. In other words, the increasing concern about the safety of compounding drugs and the availability of the arguably safer brand can potentially explain some of the Makena use after the carve-out ended, but it cannot explain all of the observed increases in prevention across all fronts.

Prevention and Plan Exits In the post-period, six plans exited at least one county or changed ownership in the included counties which may have forced some enrollees to change plans mid-way through the pregnancy.[37] This churn, in turn, may have theoretically contributed to increased preventative care. For example, if an enrollee has to change plans mid-pregnancy, she may have to switch providers, who in turn may want to repeat tests for ease of access. As such, I check whether removing enrollees whose MC plans exited impacts estimates for monitoring and take-up of preventative care. The results, shown in Appendix figure A28, suggests that this is

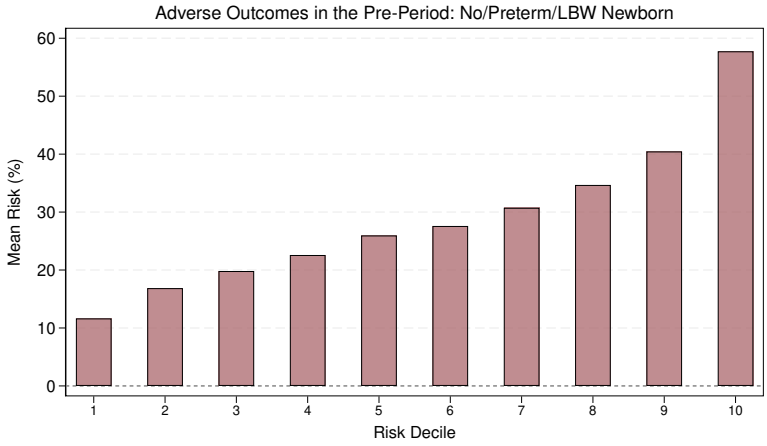
[37] Out of the 18 distinct plans that operated in the sample counties in 2010 before the Medicaid Redesign Team convened, six plans exited at least one county or experienced a change in ownership structure (e.g. were acquired by another plan). For more details see Appendix Table A6.

not the case and the results are similar in magnitude and direction when excluding these enrollees.

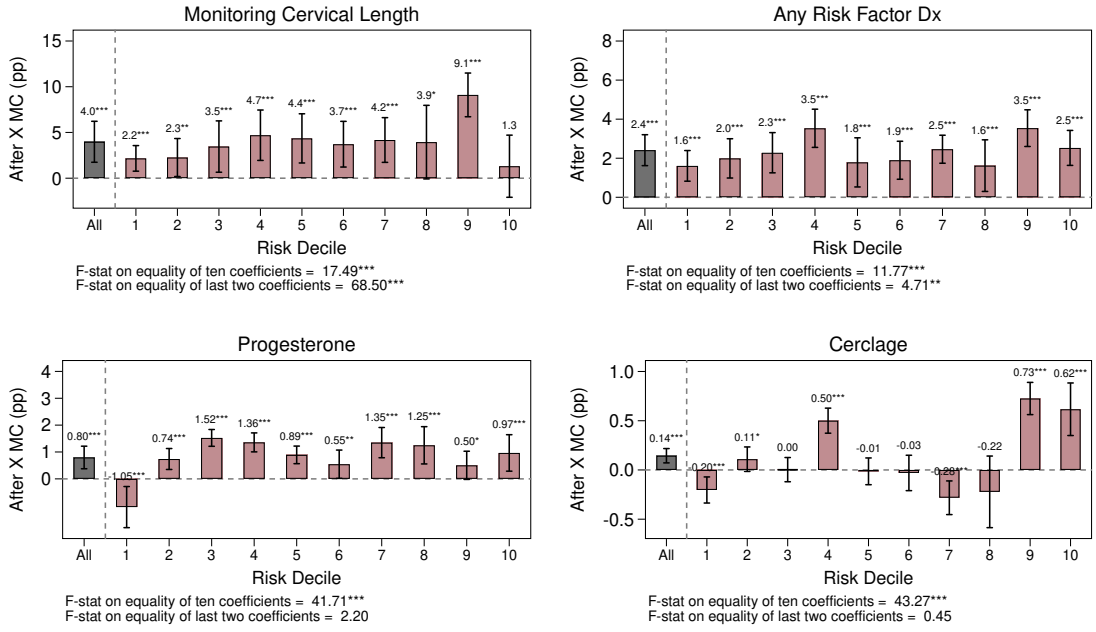
5.3 Distributional Effects: Who Benefited?

Figure 10
Preventative Care Changes by Risk

(a) Adverse Pregnancy Outcomes by Ex-Ante Risk Decile



(b) Preventative Care Changes by Ex-Ante Risk Decile



Having established that the carve-out was indeed responsible for suppressing preventative care, I go on to examine which type of enrollee was most impacted. I first focus on the mapping between the effect size and the risk of preterm birth. An ideal distributional result would be that the medium- and low-risk enrollees experienced the largest changes, that is, enrollees with the highest risk would not be impacted by whether coverage is streamlined or fragmented. Instead, ending the carve-out would only impact enrollees with marginal levels of health.

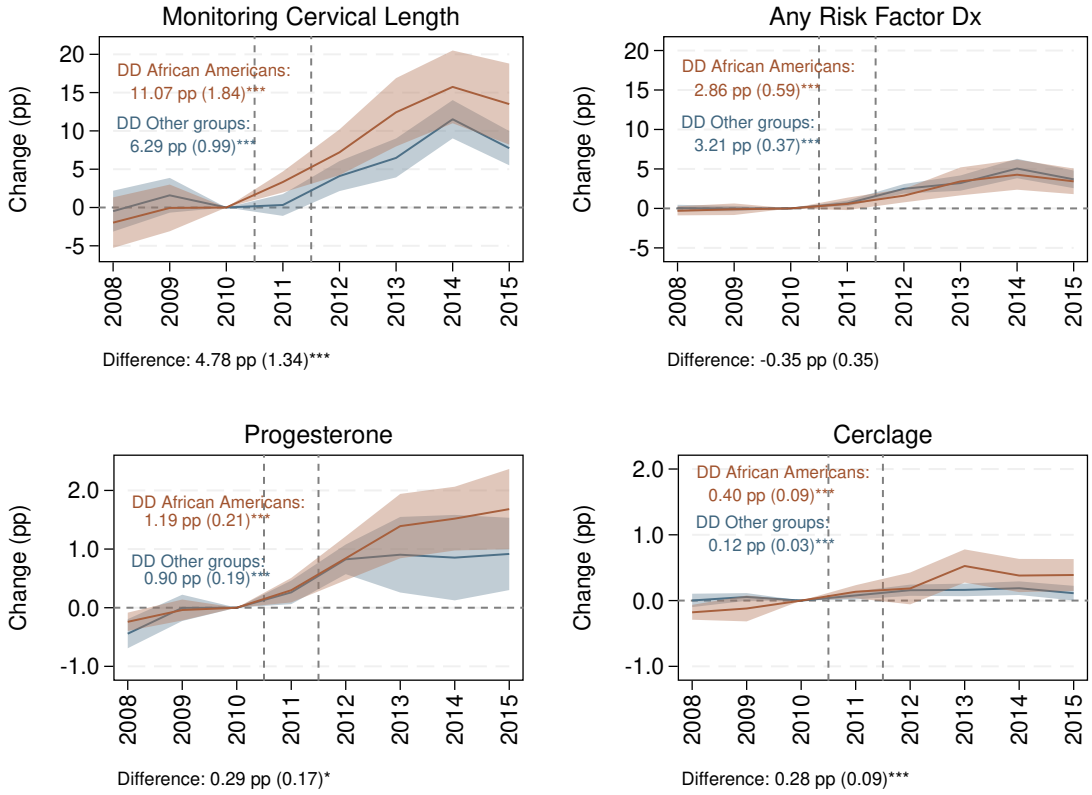
To examine whether this scenario describes the empirical context in question, I divide my sample into ten deciles based on their ex-ante risk score, which as discussed in section 3.2, is predicted using exogenous predictors of risk in the year prior to the pregnancy as well as race, age, and three-digit zip code of residence. Figure 10a shows that the rate of adverse outcomes – preterm births, low birth weight births, or still births – are largely monotonic with the risk decile. In other words, the predicted risk score captures true levels of risk.

The results in figure 10b do not imply that the best-case scenario outlined earlier took place. Neither do they imply the reverse of it: that the highest risk groups were the most impacted. Instead, the results show a largely non-monotonic relation between risk and effect size. The one clear pattern emerging, however, is with monitoring: where the effect size is correlated with ex-ante risk. One interpretation of this is that monitoring was done a non-haphazard way, based on observable risk factors. That the decile with the highest risk appears to have had limited monitoring may be because pregnancies at this level of risk do not survive until the second trimester when cervix length measurement is typically done.

One likely possibility, however, is that the variables I use to construct the risk score are not sufficient to accurately capture ex-ante risk. Providers, for example, are likely privy to more information than that recorded on claims data. As such, I examine a more observable, and arguably more exogenous, marker of risk: that related to being African American. African American women are known to have a higher risk of preterm birth and LBW infants (Burris et al., 2019). In the pre-period, for example, they are 40% more likely than white enrollees to have either or both of these adverse outcomes. Consistent with the results by risk groups, figure 11 shows that all stages of the care pipeline saw larger gains among Black and African American

enrollees than among other groups.[38]

Figure 11
Preventative Care among African American Enrollees



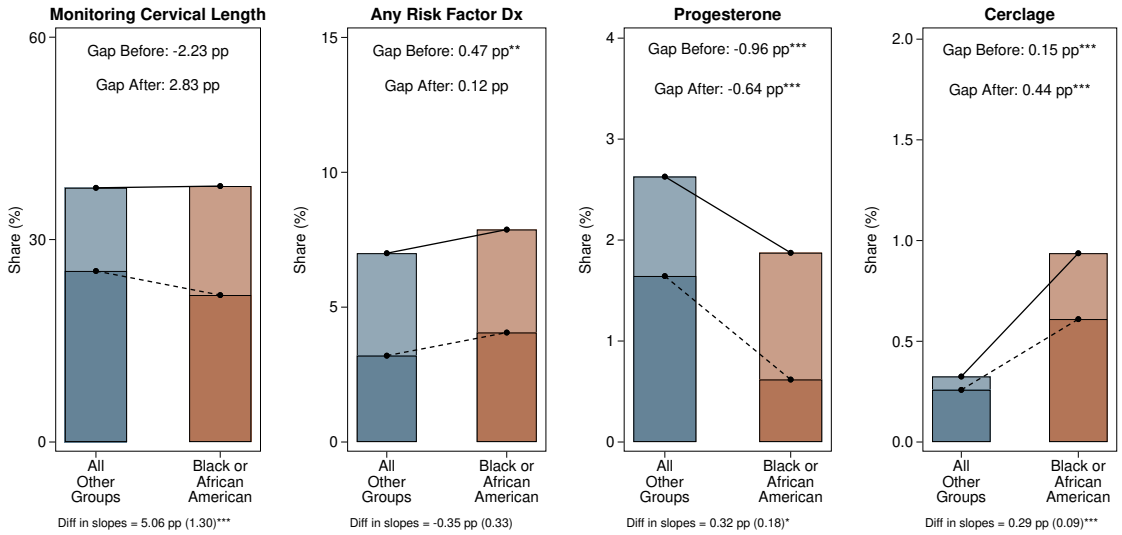
Assuming that being African American can more accurately predict the true risk of having a preterm/LBW birth, the results suggest that streamlining coverage improved allocation. Another way of examining this is by considering the gradient of preventative care with respect to risk/race. Figure 12 plots the raw levels of preventative care among MC enrollees before and after the carve-out ended and describes the differences between racial groups while differencing out FFS. In the pre-period, preventative care was negative for monitoring and progesterone use – in other words, prevention was negatively correlated with risk; the highest risk group was monitored and treated less often.[39] The ending of the carve-out increased the slope in these

[38] Appendix figure A29 shows the results across all race and ethnic groups. The largest, most precise, gains in monitoring and prevention appear to accrue to Black and African American enrollees. Changes in diagnoses appear highest among Asian enrollees.

[39] Although the difference is not statistically significant for monitoring when comparing outcomes

three outcomes. In the post-period, African American enrollees were equally likely to have their cervical length monitored and more likely have a cervical cerclage relative to other groups. To the extent that African American and Black enrollees face higher risk, this result suggested that the allocation of prevention improved.

Figure 12
Preventative Care by Risk, as Proxied by African American and Black Race



5.4 Effects on Cream-Skimming

As discussed in the introduction, one potential reason behind the carve-out is to reduce insurer incentives to cream-skim, i.e. to avoid the coverage of enrollees who may have an expensive newborn. This section asks whether ending the carve-out re-introduced these incentives.

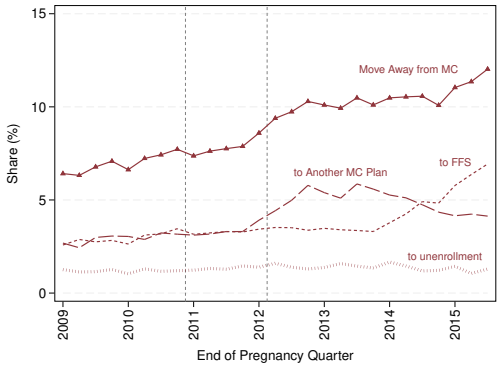
If it were the case that plans successfully pursued cream-skimming strategies, the primary evidence would be that covered pregnancies are healthier in the post-period. However, as shown and discussed in section 5.2.2 on composition change, the pregnancies covered by MC plans did not exhibit a substantially lower risk after the carve-out ended. But just because the covered pregnancies did not exhibit higher risk does not indicate that MC plans did not attempt to engage in cream-skimming. It

among African American and Black enrollees against all other groups, they are statistically significant and negative when comparing against white enrollees. See appendix figure A30 for more details.

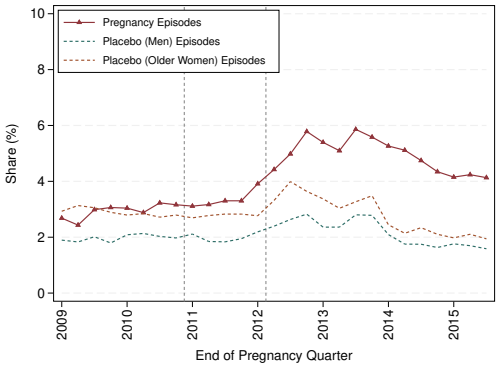
could as well mean that they did but failed to influence overall risk. I test for these attempts by examining the continuity of coverage.

Figure 13
Continuity of Coverage

(a) Churn across Three Margins



(b) Within MC Churn



Recall that one of the inclusion criteria for the sample was that the enrollees had to have either FFS or MC Medicaid continuously during the episode. This analysis describes whether this enrollment continuity (or lack thereof) could have been an outcome of ending the carve-out. The top line in figure 13a shows that indeed churn increased starting the first quarter of 2012, right before the carve-out was scheduled to end.[40]

The bottom three lines break down this churn across three margins: 1) enrollment in Medicaid in general; 2) enrollment in an MC plan; and 3) enrollment in the same MC plan for all nine months. Theoretically, MC plans can have leverage on the first two of these margins if they help enrollees re-verify their eligibility for Medicaid or if they advertise eligibility for their plans via Medicaid to unenrolled, but eligible populations. They can also have leverage on whether enrollees are continuously enrolled in an MC plan by (illegally) encouraging enrollees to leave, especially if they had just enrolled in Medicaid for their pregnancy and were still under the first three months before the lock-in period starts.

As obvious from figure 13a, churning across the enrollment margin does not seem to be affected. Churn from MC to FFS appears to have increased but only for enrollees

[40]Note that this analysis excludes any enrollees whose plans exited throughout the study period.

whose pregnancies ended in 2014, likely because of the start of the Affordable Care Act. The largest, most relevant change appears to in churn be across MC plans. Figure 13b compares this churn with the placebo pregnancy episodes described and used earlier in section.[41] Although these placebo male and older female enrollees appear to also have exhibited an increase in churn across MC plans, its magnitude is much smaller than those experienced by truly pregnant women. Appendix figure A31 shows the event study for these two comparisons and suggests the effect size is about 1.3-1.6 pp (about 43-53% of the baseline churn rate). In other words, if a primary goal of the carve-out was to avoid cream-skimming, then removing it may have reversed this effect.

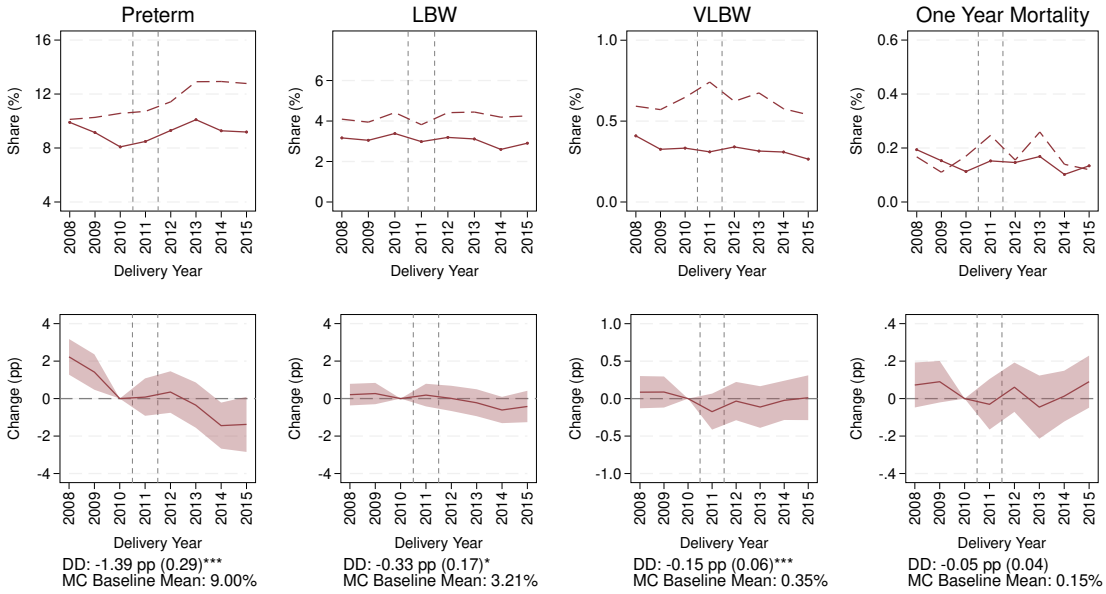
5.5 Effects on Newborn Outcomes

I then turn to my sample of linked mother and newborns. Given the increases in preventative care, were infants from MC-covered pregnancies more likely to have better birth outcomes after the carve-out ended than infants from FFS-covered pregnancies? Figure 14 shows the challenges with answering this question. When examining VLBW rates, which are typically rare, I find that rates among the smaller-sized FFS group were too noisy, rendering the confidence intervals on the yearly coefficient estimates too wide.

Preterm and LBW rates, which are slightly less rare than VLBW births or mortality, are less noisy in the FFS group but exhibit increases in the post-period that are hard to explain in the policy setting or rely on as the true counterfactual of birth outcomes for newborns of MC mothers. Although both of these outcomes show negative difference-in-difference coefficients, they are either tainted by pre-trends – such as in the case of preterm births – or statistically insignificant or only marginally significant – such as the case of LBW. Moreover, even though one-year mortality among infants may have declined, the effects are largely concentrated among newborns born in 2014 and 2015, potentially reflecting changes from the Affordable Care Act. Worth noting here is that unlike the previous set of newborn outcomes, the effects for mortality combine the effect of healthcare during the mother's pregnancy *as well as* the newborn's first year of life which is now more likely to be covered by MC plans rather than FFS.

[41] See section 5.2.1.

Figure 14
Newborn Health Outcomes



Notes: Solid lines indicate trends for newborns born to MC enrollees. Dashed lines indicate trends for newborns born to FFS enrollees.

5.6 Mechanisms

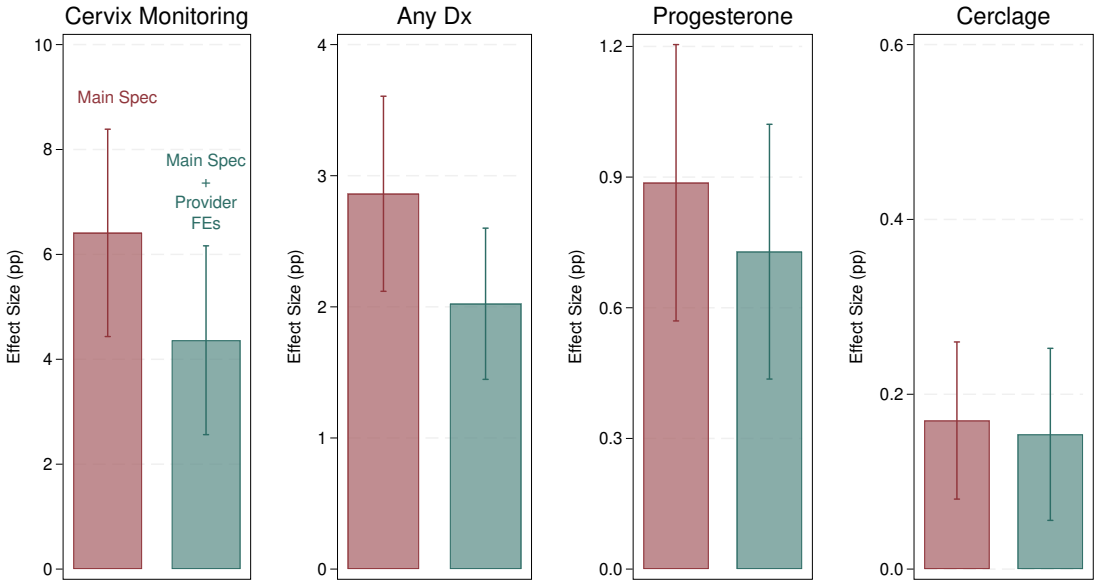
Physicians make clinical decisions and yet the results above suggest that insurers can and do influence these decisions. The question is then how. There are three potential mechanisms that can explain how the financial considerations of insurers seeped into clinical decision-making and improved the take-up of monitoring and preventative procedures: these are expanded networks, better provider reimbursements, and less restrictive prior authorizations. Because of data availability limitations, this section tests the first channel and discusses the other two.

Figure 15
Changes in Networks for Relevant and Placebo Specialties

Insurers can change their networks to include providers who are more likely to pay attention to risk and/or use preventative procedures more often. I conduct two empirical exercises to test this possibility. The first exercise examines the numbers of providers who are in-network for a given MC plan as a share of all providers serving sample enrollees. I define "in-network" for a given plan to be the number of

providers that were seen by at least one enrollee.[42] I calculate this in-network share, by specialty - marking obstetrics and gynaecology, maternal fetal medicine specialists, nurse practitioners, and midwives as relevant specialties. I also mark four placebo specialities - placebo in the sense that they are not specialties that are regularly seen by pregnant individuals. Figure 15 compares this share among relevant and placebo specialties and shows no differential changes: there is no evidence that MC plans expanded their provider networks when it came to pregnancy related care.

Figure 16
Estimates After Controlling for the Primary Physician



The red bars represent the main specification in equation 8 run on the subsample of enrollees for which the primary provider is identified but without controlling for the primary provider. The emerald bars uses the same sample, the same baseline specification, and adds provider fixed effects.

The second exercise is less direct. Instead of testing whether MC enrollees are seeing more providers, it tests whether they are seeing different ones, even if these new providers were in-network before the carve-out ended and continued to be in network after it did. Figure 16 shows the primary effect sizes with and without

[42] One can use network data, reported by MC plans, to assess whether MC plans expanded their network breadth such that they are more suitable for high-risk pregnancies. However, many recent studies have suggested that these reported networks do not reflect true measures of access — Ludomirsky et al. (2022), for example, finds that in four state Medicaid programs 25% of the specialists listed as in-network provided 86% of the care, raising the concern over "phantom providers" in Medicaid Managed Care who are included in plan networks on paper but do not accept Medicaid patients in reality (Zhu et al., 2022).

controlling for the modal provider.[43] When controlling for the primary care provider, the effects of ending the carve-out maintain 68% of their original size in the case of monitoring cervical length, 71% for risk factor diagnoses, 82% for progesterone, and 91% for cerclages. This suggests that at least two thirds of the effect sizes are driven by providers changing their behavior rather than insurers diverting enrollees to different providers.

Other Mechanisms: Determining exactly how providers change their behavior remains challenging, and although they do not fully answer the question, I explore some potential avenues descriptively. One hypothesis is that MC plans might be loosening their prior authorization requirements. Prior authorization is a key tool for MC plans to control spending by denying coverage for medically unnecessary or potentially wasteful care. Typically, providers submit a form explaining why a procedure is needed, which is then reviewed by the insurer. If the request is denied, the provider may appeal or provide more documentation, spurring several rounds of communication (Kyle and Keating, 2024). Whether a procedure or drug requires prior authorization is typically noted in online platforms accessible only to insurers and in-network providers, making it difficult to confirm whether MC plans have actually reduced prior authorization restrictions in the NY setting in question.

As such, I resort to a limited litmus test – specifically, I examine the time elapsed between a procedure and the last office visit that preceded it. For example, the interval between a cervical cerclage and the prior office visit can serve as a proxy for whether prior authorization delays have shortened. Descriptively, I do not find evidence of a reduction in this "wait time" measure (see Appendix Figure A34). However, it is important to note here that this test only measures the intensive margin of wait time and not the extensive margin of the total number of procedures that might not have occurred in the pre-period. If prior authorization restrictions were relaxed and more moderately urgent procedures were performed, the wait time for these new procedures might actually be longer than for the more urgent, inframarginal ones.

Another possibility is that MC plans may have increased reimbursement rates to providers for these services. Like prior authorization rules, verifying this hypothesis is difficult because these prices are not publicly available. One exception, though

[43] Note that to keep the sample constant, only the subsample of enrollees for whom modal provider information was available were used. Figure A32 in the appendix show the same results as above *and* the primary results estimated in the primary sample.

limited in scope, is the pre-rebate prices for branded drugs such as Makena (i.e. progesterone shots) and Crinone (vaginal progesterone). According to the Medicaid Drug Database, between 2012 and 2015, their cost per unit to MC plans increased by 35% and 72%, respectively.[44] However, since both drugs became available only in the post-period, these figures do not capture broader trends in prices paid for preventive services. A final possibility is that MC plans may have directly encouraged providers to use these preventive services and procedures without high-powered incentives, such as increased payments, or constraints like prior authorization. This might take the form of the insurer sharing new clinical research, updated guidelines, or educational material.[45]

In conclusion, while various avenues could explain how providers' behaviors have shifted in response to policy changes in MC plans, definitive conclusions remain elusive. The loosening of prior authorization requirements, although a plausible explanation, does not appear to manifest in reduced wait times for procedures. Similarly, while reimbursement increases may have occurred for some services, the available data on drug prices do not fully represent the broader trends across preventive services. Additionally, the possibility of MC plans encouraging providers through non-monetary means, such as sharing new clinical guidelines, remains a feasible yet unverified pathway. Ultimately, these findings suggest that while managed care plans may influence provider behavior through multiple channels, further research is needed to pinpoint the exact mechanisms at play.

6 Concluding Remarks

Using quasi-experimental variation, this paper showed that fragmenting insurance coverage can cause both an underprovision and misallocation of preventative care. What this paper has *not* showed and where it stands short are also worth discussing. The first limitation is a general one. This study has assumed that transitioning from a fragmented insurance regime to a streamlined one would have reverse but identical effects to transitioning the opposite way. In other words, it assumes that

[44] These figures do not fully reflect the sample as they include all counties in New York, and FFS data was suppressed due to the small number of their enrollees using these drugs.

[45] An example of this type of communication from FFS, albeit outside the pregnancy context, can be found here and is reproduced in Appendix Figure A35.

had a fictional NY state instated the VLBW carve-out for the first time in 2012, we would have observed insurers *reducing* preventative care during pregnancy – the opposite of what this study documents. Whether insurers would follow this profit-maximizing strategy (provided they are allowed) in that scenario is an assumption. As such, should there be reasons to question this assumption, one should change the interpretation of the documented results to *the benefits of streamlining coverage* rather than *the costs of fragmenting it.*

Another limitation is specific to the context of carve-outs. The paper found some evidence of churn that is consistent with cream-skimming after transitioning from a fragmented regime to a streamlined insurer regime. Given the large kick-payment in this setting and the presence of guaranteed issue laws, it is unclear why this churn may have happened. However, future work focusing on carve-out policies should aim to understand whether their role in suppressing cream-skimming outweigh their role in reducing prevention.

Overall, in this case, where preventive care was largely ineffective, streamlining coverage may have resulted in a net loss. Although the population received more preventive services, those interventions did not work, and enrollees may have been exposed to disruptions in coverage. From a public finance perspective, the kick-payment system designed to compensate plans was costly and may have failed to prevent cream-skimming. Government audits from this period also show that it may have induced plans to commit fraud misclassifying healthier newborns as VLBW and collecting over $12 million in overpayments (DiNapoli, 2014).

Despite these limitations and the idiosyncrasies of this specific carve-out, there are important policy lessons to be drawn about the costs of fragmenting coverage, even when this fragmentation is well-intentioned. The NY Medicaid carve-out aimed to protect a vulnerable population of high-risk, low-income enrollees by reducing cream-skimming and ensuring access to care for newborns in need. However, if the available preventive technology had been effective, the results of this paper suggest that at least with respect to preventative care, the carve-out would have significantly harmed this same population, with the most severe impacts felt by the most vulnerable.

Other well-intentioned carve-outs exist where preventive technologies are more well-established. For instance, HIV has been preventable with Pre-Exposure Prophylaxis (PrEP) since 2012, yet by 2022, HIV treatments were carved out of at least

two state Medicaid programs (Dawson et al., 2023). Similarly, the major drivers of end-stage renal disease (ESRD) are poorly managed diabetes and hypertension, but Medicaid often leaves the management of these conditions to private insurers while Medicare assumes responsibility for treating ESRD. In both cases, financial fragmentation is unsurprising, given the high treatment costs and concerns about rationing. However, the findings of this paper suggest that future research should weigh any reductions in rationing against the potential for reduced prevention.

Outside the context of carve-outs, the American health insurance system is replete with fragmentation. To list just a few, consider the almost-universal transition to Medicare at age 65, that Americans who lose their jobs lose their employer-sponsored insurance, that Americans who gain jobs after unemployment could lose their Medicaid, and that people who switch jobs have to switch insurers. While some of this fragmentation has been the product of historical accident, some is by design: for example, the availability of multiple plans within any of these systems can and does create competition, the benefits of which have been studied (Curto et al., 2021). Future work should consider identifying policies that either remedy the impacts on preventative care provision or reduce fragmentation where unnecessary.

References

Ables, Kelsey, "Sole drug for preterm birth pulled after FDA panel found it ineffective," *Washington Post*, March 2023.

Auty, Samantha G, Megan B Cole, and Jacob Wallace, "Association Between Medicaid Managed Care Coverage of Substance Use Services and Treatment Utilization," *JAMA Health Forum*, 2022, *3* (8), e222812–e222812.

_ , Paul R Shafer, Stacie B Dusetzina, and Kevin N Griffith, "Association of Medicaid managed care drug carve outs with hepatitis C virus prescription use," *JAMA Health Forum*, 2021, *2* (8), e212285–e212285.

Berghella, Vincenzo, "Universal cervical length screening for prediction and prevention of preterm birth," *Obstetrical & gynecological survey*, 2012, *67* (10), 653–657.

_ , Timothy J Rafael, Jeff M Szychowski, Orion A Rust, and John Owen, "Cerclage for short cervix on ultrasonography in women with singleton gestations and previous preterm birth: a meta-analysis," *Obstetrics & Gynecology*, 2011, *117* (3), 663–671.

Bertini, Ayleen, Rodrigo Salas, Steren Chabert, Luis Sobrevia, and Fabián Pardo, "Using machine learning to predict complications in pregnancy: a systematic review," *Frontiers in bioengineering and biotechnology*, 2022, *9*, 780389.

Bogdanich, Walt and Sabrina Tavernise, "U.S. Concern Over Compounders Predates Outbreak of Meningitis," *New York Times*, 2012.

Borsky, Amanda, Chunliu Zhan, Therese Miller, Quyen Ngo-Metzger, Arlene S Bierman, and David Meyers, "Few Americans receive all high-priority, appropriate clinical preventive services," *Health Affairs*, 2018, *37* (6), 925–928.

Brot-Goldberg, Zarek C, Amitabh Chandra, Benjamin R Handel, and Jonathan T Kolstad, "What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics," *The Quarterly Journal of Economics*, 2017, *132* (3), 1261–1318.

Burris, Heather H, Scott A Lorch, Haresh Kirpalani, DeWayne M Pursley, Michal A Elovitz, and Jane E Clougherty, "Racial disparities in preterm birth in USA: a biosensor of physical and social environmental exposures," *Archives of disease in childhood*, 2019, *104* (10), 931–935.

Cabral, Marika and Neale Mahoney, "Externalities and taxation of supplemental insurance: A study of Medicare and Medigap," *American Economic Journal: Applied Economics*, 2019, *11* (2), 37–73.

Cebul, Randall D, James B Rebitzer, Lowell J Taylor, and Mark E Votruba, "Organizational fragmentation and care quality in the US healthcare system," *Journal of Economic Perspectives*, 2008, *22* (4), 93–113.

Chandra, Amitabh, Jonathan Gruber, and Robin McKnight, "Patient cost-sharing and hospitalization offsets in the elderly," *American Economic Review*, 2010, *100* (1), 193–213.

Chernew, Michael E, Allison B Rosen, and A Mark Fendrick, "Value-Based Insurance Design: By abandoning the archaic principle that all services must cost the same for all patients, we can move to a high-value health system.," *Health Affairs*, 2007, *26* (Suppl2), w195–w203.

CMS, "MAX Validation Reports," 2019.

Cohen, Arnold W, Joshua A Copel, George A Macones, M Kathryn Menard, Laura Riley, and George R Saade, "Unjustified increase in cost of care resulting from US Food and Drug Administration approval of Makena (17α-hydroxyprogesterone caproate)," *Obstetrics & Gynecology*, 2011, *117* (6), 1408–1412.

Curto, Vilsa, Liran Einav, Jonathan Levin, and Jay Bhattacharya, "Can health insurance competition work? evidence from medicare advantage," *Journal of Political Economy*, 2021, *129* (2), 570–606.

Cutland, Clare L, Eve M Lackritz, Tamala Mallett-Moore, Azucena Bardají, Ravichandran Chandrasekaran, Chandrakant Lahariya, Muhammed Imran Nisar, Milagritos D Tapia, Jayani Pathirana, Sonali Kochhar et al., "Low birth weight: Case definition & guidelines for data collection, analysis, and presentation of maternal immunization safety data," *Vaccine*, 2017, *35* (48Part A), 6492.

Dawson, Lindsey, Jennifer Kates, Tatyana Roberts, and Priya Chidambaram, "Medicaid and People with HIV," Technical Report, KFF 2023.

DiNapoli, Thomas P., "Medicaid Program – Overpayments to Managed Care Organizations and Hospitals for Low Birth Weight Newborns," Technical Report, New York State Office of the State Comptroller 2014.

DOH, NY, "Care at Home Medicaid Waiver for Developmentally Disabled Children (Office for People with Developmental Disabilities [OPWDD])," 2013.

Einav, Liran and Amy Finkelstein, "We're Already Paying for Universal Health Care. Why Don't We Have It?," *New York Times*, July 2023.

Ellis, Randall P and Willard G Manning, "Optimal health insurance for prevention and treatment," *Journal of Health Economics*, 2007, *26* (6), 1128–1150.

Fang, Hanming and Alessandro Gavazza, "Dynamic inefficiencies in an employment-based health insurance system: Theory and evidence," *American Economic Review*, 2011, *101* (7), 3047–77.

_ **and Yang Wang**, "Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions," *International Economic Review*, 2015, *56* (2), 565–596.

Fonseca, Eduardo B, Ebru Celik, Mauro Parra, Mandeep Singh, and Kypros H Nicolaides, "Progesterone and the risk of preterm birth among women with a short cervix," *New England Journal of Medicine*, 2007, *357* (5), 462–469.

Fox, Nathan S, Shari E Gelber, Robin B Kalish, and Stephen T Chasen, "History-indicated cerclage: practice patterns of maternal-fetal medicine specialists in the USA," *Journal of perinatal medicine*, 2008, *36* (6), 513–517.

Frank, Richard G and Rachel L Garfield, "Managed behavioral health care carve-outs: past performance and future prospects," *Annu. Rev. Public Health*, 2007, *28*, 303–320.

_ **and Thomas G McGuire**, "The economic functions of carve-outs in managed care," *American Journal of Managed Care*, 1998, *4* (SP), SP31–SP39.

_ **, Haiden A Huskamp, Thomas G McGuire, and Joseph P Newhouse**, "Some economics of mental health 'carve-outs'," *Archives of General Psychiatry*, 1996, *53* (10), 933–937.

GAO, "Medicaid Payment: Comparisons of Selected Services under Fee-for-Service, Managed Care, and Private Insurance," Technical Report 14-533 2014.

Goldstein, Peter, Jayne Berrier, Scott Rosen, Henry S Sacks, and Thomas C Chalmers, "A meta-analysis of randomized control trials of progestational agents in pregnancy," *BJOG: An International Journal of Obstetrics & Gynaecology*, 1989, *96* (3), 265–274.

Goodnough, Abby, "Massachusetts Orders Another Pharmacy to Close," *NYT*, 2012.

Grazier, Kyle L and Laura L Eselius, "Mental health carve-outs: Effects and implications," *Medical Care Research and Review*, 1999, *56* (2_suppl), 33–59.

Hassan, SS, R Romero, D Vidyadhari, S Fusey, JK Baxter, M Khandelwal, J Vijayaraghavan, Y Trivedi, P Soma-Pillay, P Sambarey et al., "Vaginal progesterone reduces the rate of preterm birth in women with a sonographic short cervix: a multicenter, randomized, double-blind, placebo-controlled trial," *Ultrasound in Obstetrics & Gynecology*, 2011, *38* (1), 18–31.

Hitt, Emma, "FDA Panel Votes Down 8

Imbens, Guido and Donald Rubin, *Causal Inference in Statistics, Social, and Biomedical Sciences* 2015.

Jacobson, Mireille, Craig C Earle, Mary Price, and Joseph P Newhouse, "How Medicare's payment cuts for cancer chemotherapy drugs changed patterns of treatment," *Health Affairs*, 2010, *29* (7), 1391–1399.

Johnson, John WC, Karl L Austin, Georgeanna S Jones, George H Davis, and Theodore M King, "Efficacy of 17α-hydroxyprogesterone caproate in the prevention of premature labor," *New England Journal of Medicine*, 1975, *293* (14), 675–680.

Kaplan, Thomas, "New York Medicaid Panel Backs Cuts," February 2011. July 3, 2024.

KFF-Peterson, "Treatable and preventable deaths," 2024.

Kuziemko, Ilyana, Katherine Meckel, and Maya Rossin-Slater, "Does managed care widen infant health disparities? Evidence from Texas Medicaid," *American Economic Journal: Economic Policy*, 2018, *10* (3), 255–83.

Kyle, Michael Anne and Nancy L Keating, "Prior Authorization and Association With Delayed or Discontinued Prescription Fills," *Journal of Clinical Oncology*, 2024, *42* (8), 951–960.

Layton, Timothy J, Alice Ndikumana, and Mark Shepard, "Health Plan Payment in Medicaid Managed care: A Hybrid Model of Regulated Competition," in "Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets," Elsevier, 2018, pp. 523–561.

Lee, Ajin, "How do hospitals respond to managed care? Evidence from at-risk newborns," *Journal of Public Economics*, 2020, *184*, 104130.

Li, Anran, "Commitment, Competition and Preventive Care Provision." PhD dissertation, Northwestern 2023.

Liu, Sze Yan and Sungwoo Lim, "Difference in Hospital Utilization Within the First 12 Months Among Low-Birth-Weight Infants in Medicaid Managed Care Versus Fee-for-Service: A Regression Discontinuity Study," *Maternal and Child Health Journal*, 2021, *25* (9), 1410–1419.

Ludomirsky, Avital B, William L Schpero, Jacob Wallace, Anthony Lollo, Susannah Bernheim, Joseph S Ross, and Chima D Ndumele, "In Medicaid Managed Care Networks, Care Is Highly Concentrated Among A Small Percentage Of Physicians: Study examines the availability of physicians in Medicaid managed care networks.," *Health Affairs*, 2022, *41* (5), 760–768.

Macambira, Danil Agafiev, Michael Geruso, Anthony Lollo, Chima D Ndumele, and Jacob Wallace, "The private provision of public services: Evidence from random assignment in medicaid," Technical Report, National Bureau of Economic Research 2022.

MacDonald, Sarah C, Jacqueline M Cohen, Alice Panchaud, Thomas F McElrath, Krista F Huybrechts, and Sonia Hernández-Díaz, "Identifying pregnancies in insurance claims data: methods and application to retinoid teratogenic surveillance," *Pharmacoepidemiology and drug safety*, 2019, *28* (9), 1211–1221.

Marton, James, Aaron Yelowitz, and Jeffery C Talbert, "Medicaid program choice, inertia and adverse selection," *Journal of health economics*, 2017, *56*, 292–316.

Meis, Paul J, Mark Klebanoff, Elizabeth Thom, Mitchell P Dombrowski, Baha Sibai, Atef H Moawad, Catherine Y Spong, John C Hauth, Menachem Miodovnik, Michael W Varner et al., "Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate," *New England Journal of Medicine*, 2003, *348* (24), 2379–2385.

NICHD, "What are the risk factors for preterm labor and birth?," Electronically 2017.

NY DOH, "Coverage and Billing Guidelines for 17 Alpha-Hydroxyprogesterone Caproate (17P)," August 2011.

_ , "Section 6.11: Newborn Enrollment," in "Medicaid Advantage Model Contract" 2011.

_ , "Sections 6 and 8: Enrollment and Disenrollment Requirements," in "Medicaid Advantage Model Contract" 2011.

_ , "New York State Medicaid Update: Infants with Low Birth Weight and Infants with Disabilities under Six Months of Age in Medicaid Managed Care," February 2012.

_ , "New York State Medicaid Update: Enrollment of Newborns When a Mother Is Enrolled in Medicaid Managed Care or Family Health Plus," January 2013.

NY Medicaid, "New York State Medicaid Program: Managed Care Manual - Stoploss Policy and Procedures," Technical Report, NY State 2022.

NYSDOH, "Approved MRT proposals," Technical Report 2011.

_ , "Descriptions of Recommendations," Technical Report 2011.

_ , "MRT Meeting," February 2011.

Orsulak, Mary Kathryn, Dana Block-Abraham, and Rebekah E Gee, "17α-hydroxyprogesterone caproate access in the Louisiana Medicaid population," *Clinical therapeutics*, 2015, *37* (4), 727–732.

Parente, Stephen T, David S Salkever, and Joan DaVanzo, "The role of consumer knowledge of insurance benefits in the demand for preventive health care among the elderly," *Health Economics*, 2005, *14* (1), 25–38.

Parys, Jessica Van, "How do managed care plans reduce healthcare costs," *Job market paper, Columbia University*, 2015.

Patel, Yesha and Martha M Rumore, "Hydroxyprogesterone caproate injection (Makena) one year later: to compound or not to compound that is the question," *Pharmacy and Therapeutics*, 2012, *37* (7), 405.

Rao, Kiran Prabhaker, Victoria Belogolovkin, Jerome Yankowitz, Joseph A Spinnato et al., "Abnormal placentation: evidence-based diagnosis and management of placenta previa, placenta accreta, and vasa previa," *Obstetrical & gynecological survey*, 2012, *67* (8), 503–519.

Reichmann, James P, "Makena or Compounded 17P?," *Pharmacy and Therapeutics*, 2012, *37* (9), 487.

Ressel, Genevieve, "ACOG issues recommendations on assessment of risk factors for preterm birth," *American family physician*, 2002, *65* (3), 509–510.

Ressel, Genevieve W, "ACOG releases bulletin on managing cervical insufficiency," *American family physician*, 2004, *69* (2), 436–439.

Ringel, Jeanne S, Susan D Hosek, Ben A Vollaard, and Sergej Mahnovski, "The elasticity of demand for health care: A review of the literature and its application to the military health system," Technical Report, RAND@articlefrank1996some, title=Some economics of mental health'carve-outs', author=Frank, Richard G and Huskamp, Haiden A and McGuire, Thomas G and Newhouse, Joseph P, journal=Archives of General Psychiatry, volume=53, number=10, pages=933–937, year=1996, publisher=American Medical Association 2002.

Rockefeller Institute, "Evaluation of New York's Federal-State Health Reform Partnership (F-SHRP) Medicaid Section 1115 Demonstration," Technical Report 2015.

Roman, Amanda, Noelia Zork, Sina Haeri, Corina N Schoen, Gabriele Saccone, Sarah Colihan, Craig Zelig, Alexis C Gimovsky, Neil S Seligman, Fulvio Zullo et al., "Physical examination–indicated cerclage in twin pregnancy: a randomized controlled trial," *American Journal of Obstetrics and Gynecology*, 2020, *223* (6), 902–e1.

Society for Maternal-Fetal Medicine and Vincenzo Berghella, "Progesterone and preterm birth prevention: translating clinical trials data into clinical practice," *American journal of obstetrics and gynecology*, 2012, *206* (5), 376–386.

Stankaitis, Joseph A, Howard R Brill, and Darlene M Walker, "Reduction in neonatal intensive care unit admission rates in a Medicaid managed care program," *Am J Manag Care*, 2005, *11* (3), 166–172.

Starc, Amanda and Robert J Town, "Externalities and benefit design in health insurance," *The Review of Economic Studies*, 2020, *87* (6), 2827–2858.

Stewart, Amanda and Ernest Graham, "Preterm birth: An overview of risk factors and obstetrical management," *Developmental disabilities research reviews*, 2010, *16* (4), 285–288.

Surendiran, R, R Aarthi, M Thangamani, S Sugavanam, and R Sarumathy, "A Systematic Review Using Machine Learning Algorithms for Predicting Preterm Birth," *International Journal of Engineering Trends and Technology*, 2022, *70* (5), 46–59.

U.S. Attorney's Office, "Former Owner of Defunct New England Compounding Center Resentenced to 14 Years in Prison in Connection with 2012 Fungal Meningitis Outbreak," July 2021.

WBUR, "Hospital Pharmacies Step Up Compounding Efforts In Wake Of Meningitis Outbreak," online, radio November 2012.

Winkelman, Ross and Syed Mehmud, "Risk Adjustment and Reinsurance under the ACA: New York State Recommendations," Technical Report, Wakely Consulting Group 2012.

WNYLC, "Who Does Not Have to Join a Managed Care Plan? Exemptions and Exclusions," 2021.

Wolfberg, Adam, "Measuring Cervical Length: A Key to Ending Preterm Births," *The Atlantic*, 2012.

Yukhananov, Anna, "K-V sues FDA over Makena in fight for survival," July 2012.

Zhu, Jane M, Christina J Charlesworth, Daniel Polsky, and K John McConnell, "Phantom Networks: Discrepancies Between Reported And Realized Mental Health Care Access In Oregon Medicaid: Study examines phantom networks of mental health care providers in Oregon Medicaid.," *Health Affairs*, 2022, *41* (7), 1013–1022.

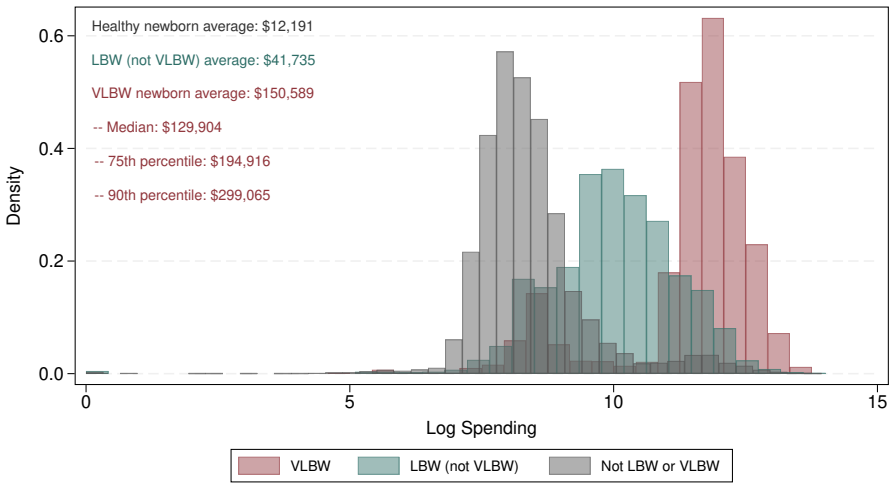
Appendix

A Background

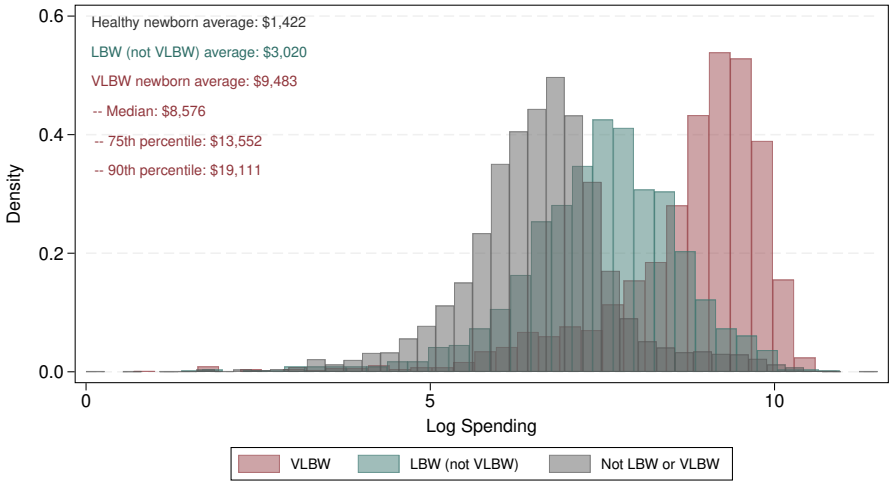
Spending on Newborns by Type of Care

Figure A1
Spending on FFS Newborns In Their First Six Months, by Care Type

(a) Inpatient



(b) Outpatient



Notes: Graph reflects spending in the pre-period of this study from 2009 through March 2012.

Studies on the Effectiveness of Preventative Treatment

Table A1
Studies Available Before the End of the Carve-Out on Progesterone

Study	Study Type	Population	Birthweight Effects	Gestational Age Effects	Other Effects
(A) Progesterone Shots					
Meis et al. (2003)	RCT, N = 463	a history of spontaneous preterm delivery	=> Birthweight less than 2500 gm: RR: 0.66 (0.51-0.87) => Birthweight less than 1500 gm: RR: 0.62 (0.36-1.07)	=> less than 37 weeks: RR: 0.66 (0.54-0.81) => less than 35 weeks: RR :0.67 (0.48 - 0.93) => less than 32 weeks: RR: 0.58 (0.37 - 0.91)	
Facchinetti et al. (2007)	RCT, N = 90	Women admitted for threatened preterm labor between 25 and 33 + 6 days of gestation	–	=> Preterm delivery: . The relative risk was 0.15 (95% CI, 0.04-0.58).	=> reduction in the risk of cervical shortening of ≥4 mm (odds ratio, 0.18; 95% CI, 0.04-0.66).
Berghella et al. (2010)	Observational within a cerclage trial, N = 300	Prior spontaneous preterm birth and short cervix length	–	No effects on women randomized to receiving a cerclage. Women with no cerclage: => Previable births < 24 weeks: 0.08 (0.01–0.60)	Women with no cerclage: Perinatal deaths: 0.14 (0.03–0.61)
Bastek et al. (2011)	Observational, N=15,421	women with a singleton pregnancy and a history of spontaneous PTB of a singleton infant between 20 and 36 6/7 weeks gestational age (at an urban academic medical center)	–	No reduction in preterm births. Gestational age shifted from early preterm birth to late preterm birth	

Table A2
Studies Available Before the End of the Carve-Out on Progesterone (Continued)

Study	Study Type	Population	Birthweight Effects	Gestational Age Effects	Other Effects
			(B) Vaginal Progesterone		
O'Brien et al. (2007)	RCT, N = 659	Women with a history of prior spontaneous singleton preterm birth at between 20 + 0 and 35 + 0 weeks	=> No effect on birthweight	=> No effect on preterm birth	=> No effect on NICU admission
Fonseca et al. (2003)	RCT, N=413	high risk for preterm delivery: the presence of at least one previous spontaneous preterm birth, prophylactic cervical cerclage, and uterine malformation.	–	Increases in gestation age at delivery, especially between 29 weeks to 32 weeks.	
Hassan et al. (2011)	RCT, N = 458	Women with a sonographic short cervix in the mid-trimester	=> birthweight < 1500 g: RR, 0.47; 95% CI, 0.26–0.85; P = 0.01). => Birthweight < 2500 gm: RR, 0.80 (0.57–1.13), P= 0.210	=> Before 28 weeks (RR, 0.50; 95% CI, 0.25–0.97; P = 0.04) => Before 33 weeks: (RR, 0.55; 95% CI, 0.33–0.92; P = 0.02) => Before 35 weeks: RR, 0.62; 95% CI, 0.42–0.92; P = 0.02)	
Cetingoz (2011)	RCT, N = 150	Women with prior spontaneous preterm birth, twin pregnancy, and uterine malformation.	--	=> Delivery <37 weeks: OR: 2 (1.04–3.83) (Placebo to Progesterone) => Delivery <34 weeks: OR: 3.35 (1.3–8.63) Among women with prior preterm birth: => Delivery <37 weeks: OR: 3.11 (1.13–8.53) (Placebo to Progesterone) => Delivery <34 weeks: OR: 6.3 (1.25–31.7)	
Fonseca et al. (2007)	RCT	Women with asymptomatic short cervix mid-gestation	=> No effect on birthweight	=> less than 34 weeks: RR: 0.56 (0.32-0.91)	=> neonatal morbidity (8.1% vs. 13.8%; relative risk, 0.59; 95% CI, 0.26 to 1.25; P=0.17).

Table A3
Studies Available Before the End of the Carve-Out on Progesterone and Cerclages

Study	Study Type	Population	Birthweight Effects	Gestational Age Effects	Other Effects
			(C) Oral Progesterone		
Rai (2009)	RCT, N = 150	Women with at least one previous spontaneous singleton delivery between 20 and 36 weeks plus 6 days	=> birthweight (2400 vs 1890 g, P < 0.001)	=> Mean gestational age (36.1 vs 34.0 weeks, P < 0.001). => preterm births between 28 and 31 weeks plus 6 days (RR 0.20; 95% CI, 0.05–0.73, P < 0.001). => Neonatal age at delivery (34 vs 32 weeks, P < 0.001)	=> NICU stay (> 24 h, P < 0.001) => fewer neonatal deaths occurred (3 vs 7, P = 0.190).
			(D) Multiple Types of Progesterone		
Rode et al. (2009)	Meta-analysis on randomized trials (different types of progesterone)	=> Six studies on singleton pregnancies with previous preterm birth. => short cervix at 23 weeks => women with preterm labor	–	=> 6/6 studies showed that receiving progesterone was associated with a significant reduction of delivery before 32 weeks and of perinatal mortality.	=> 6/6 studies showed that receiving progesterone was associated with a significant reduction of delivery before 32 weeks and of perinatal mortality.
			(E) Cervical Cerclages		
Berghella et al. (2011)	Meta-analysis of five clinical trials	women with singleton gestations, previous spontaneous preterm birth, a short cervical length in the second trimester enrolled in five trials	–	Preterm births: RR: 0.70, 95% CI: [0.55-0.89]	
Drakeley, Roberts, and Alfirevic (2003)	Meta-analysis of six trials		–	=> No statistically significant reduction in preterm delivery rates => Reduction in delivery before 33 weeks in the largest trial (relative risk [RR] 0.75; 95% confidence interval [CI] 0.58 to 0.98).	

Simulation Details

This section describes the exact steps to operationalize the simulation. First, however, to clarify terminology, I define three categories ("buckets") of newborns: inframarginal newborns with birthweights under 1000 grams; marginal newborns with birthweights between 1000 and 1500 grams; and healthier newborns over 1500 grams. As described in the main text, newborns in the inframarginal and marginal buckets can be impacted by the treatment but only marginals can move across the cutoff. Within each category, I observe different birthweight bins: these are the 200 gram birthweight bins that I observe in my data.

A.0.1 Step 1: Simulating MC spending

I simulate 100 newborns covered by a MC plan. I assign 62% of them to be marginals and the rest (38%) to be inframarginals to mimic MC newborns in the pre-period as shown in Table A4. (I do not simulate the third bucket given that they are not impacted by the carve-out). I then assign each bucket the first-six-months inpatient and outpatient spending distribution that FFS newborns in the same bucket faced. This is done by fitting a distribution to the FFS spending data. The fitting uses hand-tuning techniques to identify the parameters that produce a good fit – see figure A2a for the resulting distributions. Given the resulting theoretical distributions, I draw inpatient and outpatient spending data for each newborn and add them to produce total spending.

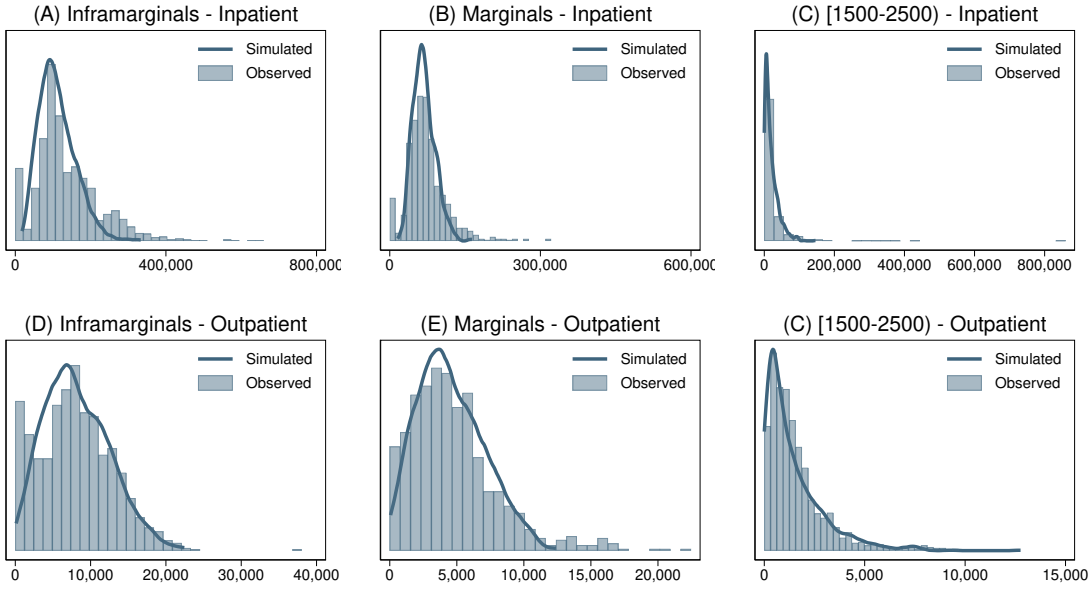
Table A4
Fraction of Marginal and Inframarginal Pregnancies In the Pre-Period

	(1)	(2)	(3)
	Fraction Inframarginal	Fraction Marginal	Number
All Pre-period	0.38	0.62	1005
2008	0.42	0.58	321
2009	0.38	0.62	292
2010	0.34	0.66	340

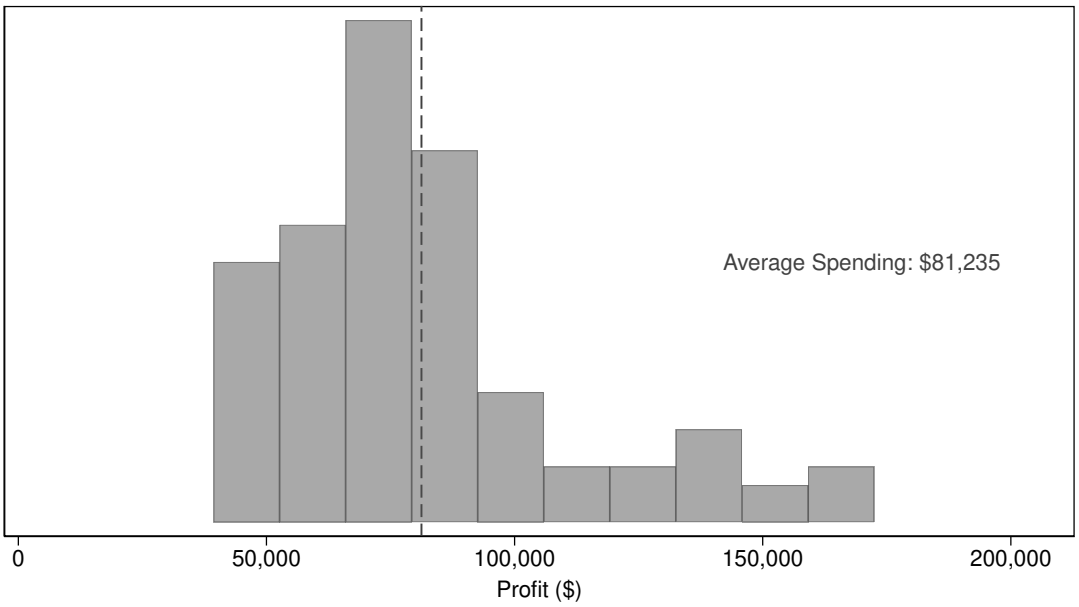
Notes: Based off MC pregnancies in 2008 through December 2010 that resulted in newborns under 1,500 gm. Marginal newborns are defined as having a birthweight between 1,000 and 1,500 gm and inframarginals under 1,000 gm. Note that newborns with unspecified birthweight were removed.

Figure A2
Simulation Steps

(a) Step 1: Simulate MC Spending, by Newborn Type and Spending Category



(b) Step 2: Combine Spending and Shares of Infra-marginal and Marginal Newborns



This process makes the assumption that within a bucket (marginal or infra-marginal), newborns covered under MC and FFS would have a similar spending dis-

tribution - this, of course, may be violated if MC plans engaged in cream-skimming of any sort. I also only include the spending of newborns who do not die within the first six months, with the assumption that MC plans do not count on a fraction of their newborn enrollees dying. Given these two factors and the general finding that MC plans typically spend less than FFS on any given enrollee, I deflate the entire distribution by 30%, following estimates from the literature on MC spending differences.[46]

Step 2: Simulating MC profits

Given these counterfactual distribution of counterfactual MC spending, I then construct the corresponding distribution of MC profits from newborns under 1500 grams. I do this by adding the plan revenues from the monthly capitation and risk adjustment kick-payments. The resulting distribution describes counterfactual MC profits when the carve-out ends and when plans do not adjust their levels of preventative care.

Step 3: Operationalizing the kick-payment effect and the cost-savings effect

When plans provide preventative care to these pregnancies, two opposing effects happen. The first is that some pregnancies benefit from the treatment and their resulting newborns cross the VLBW cutoff. This transition costs plans the kick-payment they would have received had they not increased preventative care. The second is that pregnancies become healthier and newborns are born with a higher birthweight, which, as shown earlier, reduces plan spending. This cost-saving effect shifts the spending to the right. Whether the net effect is positive or negative would determine the plan incentives to use preventative care.

To operationalize the kick-payment and the cost-saving effects, I need to specify the fraction of inframarginal and marginal newborns that are impacted by preventa-

[46] Lee (2020) who studies the same carve-out finds that hospital spending at birth was 20-30% lower in MC relative to FFS. Other studies estimate that MC spending is between 10% to 22% lower (Macambira et al., 2022; Van Parys, 2015; Marton et al., 2017). I use the largest of these estimates to be conservative against finding a risk transfer to MC plans.

tive care. I also need to know the magnitude of the impact: e.g., how much birthweight these newborns gain, which in turn translates into how much cost-savings and forgone kick-payments they generate. While the clinical studies available at the time showed that progesterone and cerclages increase birthweight, they did not specify their effects on the *distribution* of birthweight (i.e., which parts of the lower end of the distribution lost mass and which parts gained mass). Given this ambiguity, I make two assumptions:

1. The first concerns the magnitude of the impact: specifically, I assume that a a pregnancy that is impacted by preventative care would yield a newborn with a birthweight that is in the heavier birthweight bucket than the one it would have landed in had it been untreated. For example, a newborn that is inframarginal absent treatment would be marginal with treatment. A newborn that is marginal absent treatment would fall in the 1500-2500 gram bucket. This assumption allows me to assign each bucket of newborns a counterfactual distribution of spending.

2. As for the fraction that are impacted, I first consider marginal newborns. Specifically, I rely on the 38% reduction in the share of 1500 grams that is reported in the primary clinical trial for progesterone (Meis et al. (2003)). Although Meis et al. (2003) do not qualify where in the distribution of newborns under 1500 grams that this decline comes from, I assume it is concentrated solely among the marginals (birthweight between 1000 and 1500 grams). Although the the effects could be driven from the low-tail of the birthweight distribution, I choose this assumption because it is more plausible that preventative care moved those who were just under 1500 gram to a birthweight above 1500 gram.[47] I also assume that the reduction is uniform across the two bins of newborns within that bucket (1000 to 1250 and 1250 to 1500 gm). Adjusting for the fact that marginals only make up 61% of newborns under 1500 grams among the average plan, the effect among marginals is thus 62% (i.e. $0.62 * 0.61 = 38\%$). This is the fraction of marginal newborns that are impacted by preventative care.[48] Given that this

[47] Another reason is I make this assumption is because it is more conservative against finding a risk transfer to MC plans: spending on marginals is lower than that among inframarginals, as shown by figure 2. As such, if only the marginals are changing to a non-VLBW birthweight, then the implied cost-savings will be lower.

[48] Note that that using this 38% estimate assumes there was no preventative care used in the pre-period. If the baseline rate is higher, then the effect of moving to universal preventative care

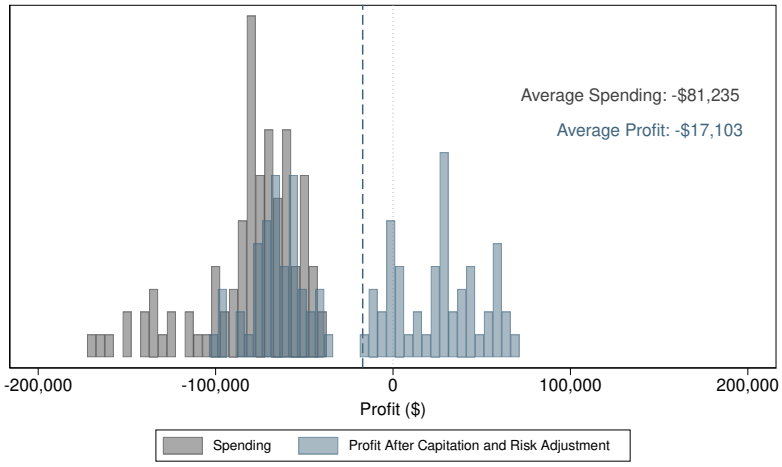
38% reduction (or 62% reduction among marginals) is a rather large effect, I also examine how the results change if MC plans discounted this effect reported in the literature. Specifically, I define the parameter s that takes varies from $[0, +\infty]$ focusing on cases where $s = 1$ (i.e., when plans believed the effectiveness of preventative technology is exactly as reported in the clinical trial) and when $s = 0$ (i.e., when plans believed the effectiveness is half as that reported in the clinical trial). Note that this discounting may also arise because provides may not know which pregnancies would produce which newborns. As such, even if they observe X number of newborns under 1500 grams in the pre-period and they treat X pregnancies in the post-period, they may not be the "correct" pregnancies, leading to lower effectiveness among the truly inframarginal and marginal pregnancies.

3. So far, I specified, that 62% of the marginals will move in birthweight to the next bucket if treated. I have also specified that a share of the inframarginals will move but do not have information to calibrate it. (Recall that inframarginals may have been impacted in the progesterone clinical trial but that share is not reported). As such, I examine the effect at two meaningful cases, when marginals and inframarginals benefit equally from preventative care and when only marginals benefit from it. That is, when the k parameter which describes the relation the effect of preventative care on inframarginals relative to marginals (described in section 2) equals to 0 and when equals to 1.

among newborns under 1500 grams should be lower. However, as reported in the main text, the rate of progesterone use among MC plans in the pre-period is around 0.04%. Similarly, for cerclage, the rate is about 0.5%.

Figure A3
Simulation Steps (Continued)

(a) Step 3: Turn Spending into Profits by Adding Capitation and Risk Adjustment



(b) Step 4: Examine Simulated Counterfactual Profits for the Average Plan with Pre-Period Level of Preventative Care

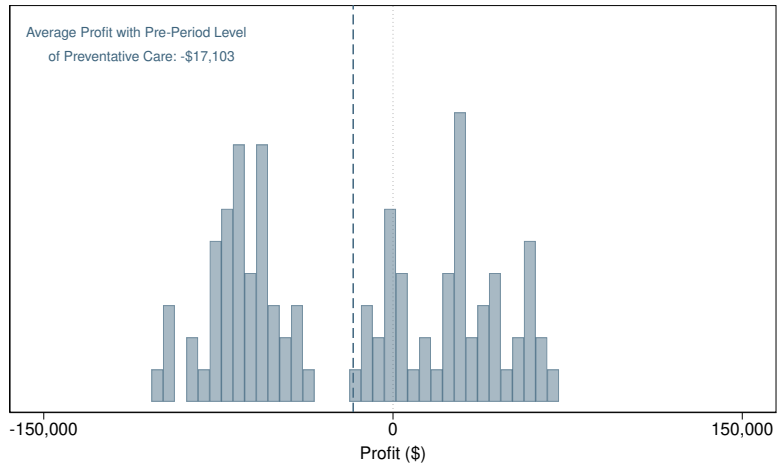
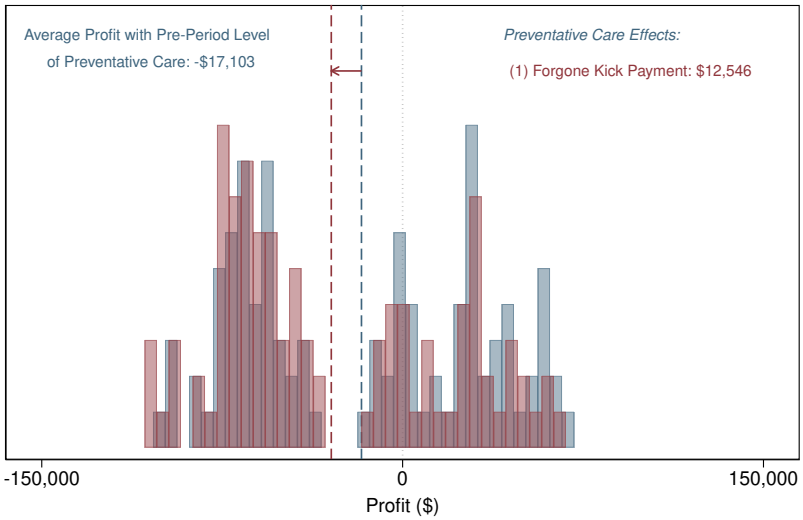


Figure A4
Simulation Steps (Continued)

(a) Step 5: Add Kick-Payment Effect, Holding Cost-Savings Constant



(b) Step 5: Add Cost-Savings, Holding the Kick-Payment Constant

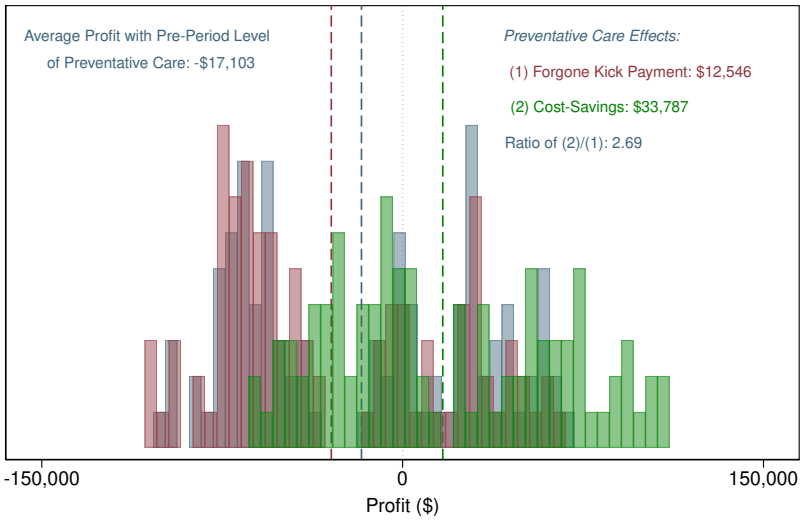
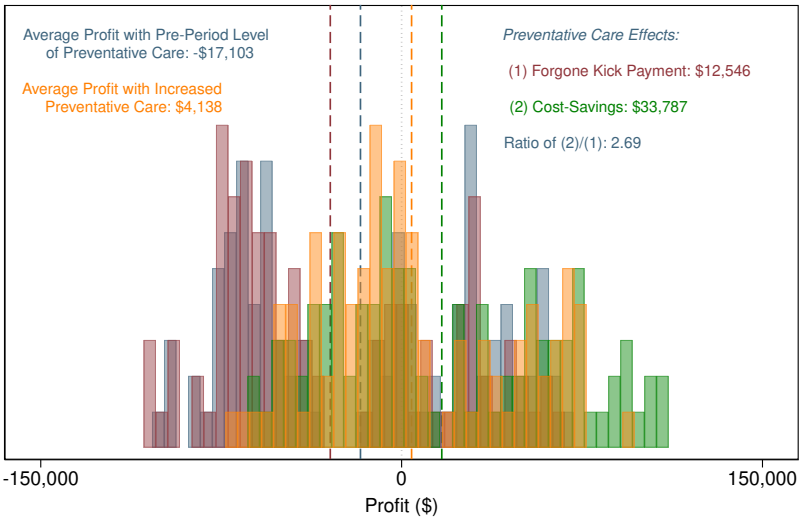


Figure A5
Simulation Steps (Continued)

(a) Step 7: Examine Net Effect of Increasing Preventative Care



Step 4: Addressing uncertainty:

To take into account that plans face a large amount of uncertainty in their newborn spending, I run 300 draws out of the theoretical MC spending distributions reached in step 1. I then repeat steps 2 through 4, each time obtaining a different net effect on profit when achieving universal treatment level for newborns under 1500 grams. I then obtain the mean for these effects and the standard deviation across the resulting 1000 effects. This process assumes that any uncertainty in the simulated effects is generated from the variation in necessary spending on newborns under the theoretical distribution rather than any uncertainty in the parameters of this distribution.

Decomposing the net change into the cost-sharing effect and the kick-payment effect

Table A5
Simulated Profit Changes from Preventative Care Use

	(1)	(2)	(3)	(4)
	Conservative Assumptions	Less Conservative Assumptions		
Effect on inframarginals relative to marginals	$k = 0$	$k = 0$	$k = 1$	$k = 1$
Perceived effectiveness	$s = 0.5$	$s = 1$	$s = 0.5$	$s = 1$
Baseline Profit	-14,889	-14,889	-14,889	-14,889
	(2,144)	(2,144)	(2,144)	(2,144)
Profit After Increased Preventative Care	-12,260	-9,308	-9,604	-3,631
	(2,811)	(2,956)	(2,786)	(2,960)
Difference	2,629	5,581	5,285	11,258
	(2,208)	(2,518)	(2,689)	(3,203)
Forgone kick payment	-7,918	-7,681	-15,268	-15,951
	(2,270)	(2,277)	(2,584)	(2,263)
Cost-Savings	10,547	13,262	20,553	27,209
	(2,240)	(2,467)	(2,678)	(3,226)

Note: Estimates are means and standard errors from 300 draws. Figure 3 in the text shows the results for additional combinations of s and k beyond those shown.

B Constructing the Primary Sample

I start by using the MAX inpatient and outpatient files to identify codes that mark an end of a pregnancy episode (EOP). I use ICD-9 diagnosis and procedure codes, CPT codes, DRGs, and HCPCS codes, following MacDonald et al. (2019) who categorize EOPs into live births, still births, mixed births, ectopic pregnancies, miscarriages, and elective terminations.[49] Creating a monthly level list of EOPs for any enrollee, I then remove any consecutive number of months that point to the same delivery (or non-delivery) EOP outcome - retaining the first month as the month when the pregnancy ended. This step ensures that any complications or follow-up care relating to a prior month's EOP are bundled with the EOP, rather than coded as a new EOP.

The NYS Department of Health publishes yearly numbers of live births and abortions on Medicaid. As such, I compare these two outcomes against the records from the claims data in figure A6. The figure shows that live births appear to follow the reported numbers, except with some relatively small underestimations in the latter years of the sample period. Abortions, on the other hand, appear largely underestimated - with one quarter to one third of them missing from the claims data annually. The one exception is 2011, the year of the policy announcement.

If one were to only rely on the MAX data, a possible conclusion would be that the abortion increase in 2011 is a result of ending the carve-out. As a form of cream-skimming, private insurers may have encouraged unhealthy pregnant enrollees to pursue an abortion in 2011. However, this is unlikely for at least two reasons. First, the increase is followed by a regression to the pre-period means, meaning that it was not the case that sensitive abortion data were under-reported and that this under-reporting was fixed in 2011. Moreover, under-reporting is unlikely because, as panel (b) in figure A6 shows, the same pattern appears in NYC, where abortion is allowed and covered regardless of medical necessity. Instead, it is more likely that these are claims with incorrectly coded dates. Second, because the increase is quite high in both absolute and relative terms – towering over 25% increase relative to 2010 , the hike would have attracted attention. To my knowledge, no media or scholarly outlets have corroborated this change.

Given this likely incorrect coding of abortion timing, I remove all pregnancy

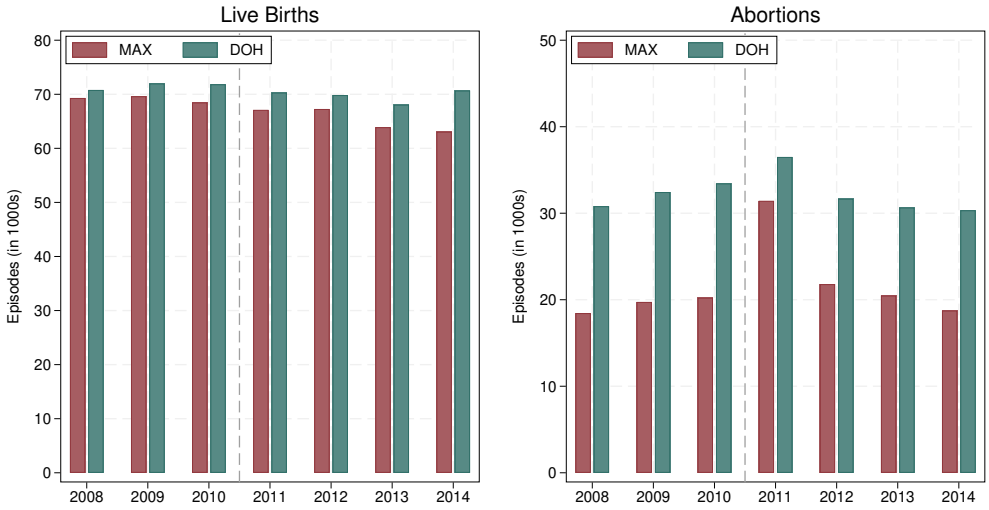
[49] I follow the codes listed in e-Appendix A of that paper.

episodes that ended with an induced abortion. In addition to avoiding the miscoding, this decision also allows me to include only "wanted" pregnancies in the sample, which are those that ended in a live birth on one end or a miscarriage, an ectopic pregnancy, or a still birth on the other end. Pregnancies given an "unspecific abortion" or "unspecific delivery" were dropped. The same applies to episodes with codes indicating both a live birth and a miscarriage.

Figure A6
Comparison of End-Of-Pregnancy Codes from MAX Data and Public Records from NY Department of Health

(a) NY State

(b) NY City

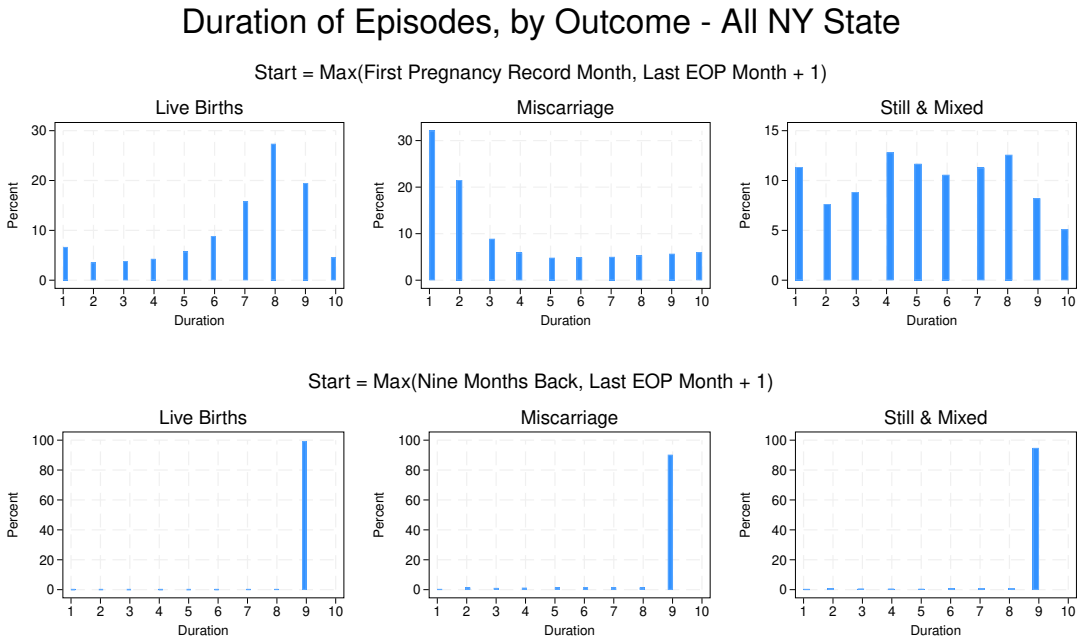


Notes: Year 2015 was removed given that the data only spans to September 2015. NYC is shown separately in the bottom panel because unlike upstate NY, abortion is covered by Medicaid there regardless of medical necessity.

I then link each of these EOP codes to the earliest pregnancy-related claim within

a 10-month range. I code this as the first month of a pregnancy and assign a unique ID number to each episode. The top panel in figure A7 shows the duration of the episodes, by their end outcome. The bottom panel shows an alternative duration assigned by brute forcing each EOP to be either nine months long or, if there were a prior episode within nine months, to the month following the end of the prior episode.

Figure A7
Duration of Pregnancy Episodes



I link each enrollee with the MAX Person-Summary files to obtain enrollment information: whether the person was enrolled throughout the pregnancy, whether she was enrolled on an MC plan or FFS, and if on MC, which plan. This information forms the basis of the inclusion criteria. So does the person's county: only episodes that took place in counties that transitioned to mandatory MC by 2004 were included.[50]. Information on enrollees' dual status were also obtained at this step in order to exclude any pregnant dual enrollees.

[50]There were two additional transition waves in that started in 2007 and 2011. However, these were not included because they overlapped with the study. The included counties are Chautauqua, Cattaraugus, Erie, Genesee, Niagara, Orleans, Monroe, Livingston, Ontario, Oswego, Onondaga, Oneida, Herkimer, Broome, Saratoga, Albany, Rensselaer, Greene, Columbia, Rockland, Westchester, Nassau, Suffolk, and the five boroughs of NYC. The identification of counties and the timing of their transition to MC was based off of the enrollment reports published by the Department of Health

Finally, I link each enrollee with any outpatient claims she had during an episode prior to day before the delivery date. Out of 988,075 episodes, 4% (42,819) did not have any non-EOP records, including any office visits. These were assigned zero treatment for all variables of interest.

C Predicting the Risk of Adverse Pregnancy Outcome

The goal behind predicting risk is to identify enrollees who present with a high risk pregnancy to assess if the gains in monitoring and preventative care were targeted towards them. In this section, I address (a) what high risk means; (b) who to include in this exercise, (c) what to predict with, and (d) how to identify high risk pregnancies.

C.1 Definition of High Risk

The thought experiment behind this exercise is to ask whether treatment effects were concentrated at pregnancies that would have otherwise led to VLBW newborns. As such, the ideal definition of "risk" here is risk of a VLBW newborn.

However, two factors make this definition both theoretically unsuitable and functionally impractical. First, the earlier definition abstracts away from the fact that birth weight cannot be predicted with such level of granularity, especially at the second trimester when monitoring and preventative care need to be started if deemed necessary. Second, predicting such a narrowly-defined outcome comes with the empirical challenges of predicting an "unbalanced" outcome – which is a question that is often discussed and studied in the machine learning literature.

Given these reasons, I use a wider definition of "risk" related to adverse pregnancy outcomes. Specifically, I define risk as the probability a preterm birth, a low birth weight newborn, a miscarriage (or a still birth), or any combination of these three outcomes. Unlike the VLBW outcome, this definition is less rare - in fact, it makes 28% of the preperiod pregnancies. Moreover, it acknowledges that from the ex-ante perspective all three of these outcomes are highly correlated and could be thought of as different points in a spectrum where a VLBW outcome sit in the middle. For example, a pregnancy that miscarries could, under treatment, also be a pregnancy

that leads to a live birth, albeit a low (or very low) birthweight one. Alternatively, a pregnancy that leads to a LBW newborn, under less necessary treatment, could end early and lead to a VLBW newborn.

C.2 Training Sample

To train my prediction model, I apply the following criteria:

- Much like my primary sample, I only use pregnancies with one payer; pregnancies in counties that transitioned to Managed Care before the pre-period starts in 2008, and pregnancies with no Medicare coverage (i.e. no duals).
- I use pre-period pregnancies (i.e. pregnancies that ended before 2011) so that they are not affected by marginal preventative care and monitoring.
- I use pregnancies from both FFS and MC to account for any variation in billing patterns.
- I restrict to pregnancies that were covered for one year prior to the start of pregnancy. This is because any predictors I use cannot come from claims during the pregnancy given that these would be endogenous to the coverage. I describe the predictors in more details in the next section. Additionally, I define the start of the pregnancy as either the first month when a pregnancy-related claim was recorded or nine month prior to the end of pregnancy. If there was another pregnancy that ended within the prior nine months, I use the month after this prior pregnancy.

After applying these criteria, the sample would ideally be further restricted to pregnancies that survived the first trimester. This is because pregnancies that end earlier are not really amenable to monitoring or preventative care. However, as discussed earlier, the data only include the first time the enrollee sought pregnancy-related care, rather than the date of her pregnancy starting. Finally, I use a 70% random sample of the resulting pregnancies, stratified by MC status, as training data and keep the remaining 30% as testing data.

C.3 Predictors of Risk

The guiding principles of which predictors to use is that (1) they must be observable to the physicians and possibly to the plans, (2) they must be related to adverse pregnancy outcomes, and (3) they must be exogenous to the coverage and the resulting monitoring and preventative care. An example of a predictor that would defy this last principle is having a short cervix diagnosis for example. Even though a short cervix is predictive of a preterm birth, the diagnosis is made when there is more monitoring, which is a function of the insurer's incentive.

Given these principles, I predict the likelihood of an adverse outcome using two sets of variables:

1. Demographics: race, eligibility code on the the first month of pregnancy, three-digit zip codes, age dummies, whether age at the end of pregnancy is considered risky (below 18 or 35 and above)
2. Exogenous risk factors (i.e. exogenous because they are diagnosed during the year prior to the pregnancy): hypertension, diabetes, obesity, being underweight, use of tobacco, drugs, or alcohol, chronic kidney disease, autoimmune disorders, urinary tract infections, HIV, asthma, thyroid disorders, cardiovascular disease, and mental health disorders.

C.4 Prediction Algorithm

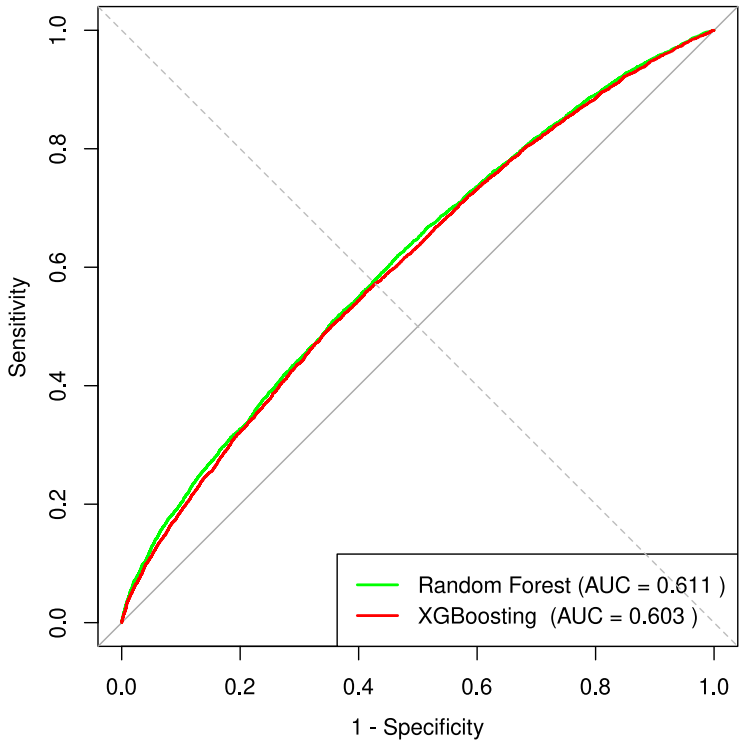
I test two different machine learning algorithms: a random Forest and XGBoosting. Both of these algorithms offer more flexibility than a traditional logit, for example, as they do not try to fit the data to a functional form – instead, they derive the appropriate groupings and cuts of predictors that lead to better prediction (e.g., enrollees with ages between 28 and 32 who are African American and have had a prior preterm birth.)

To test these two methods, I use the training data to fine tune the parameter of each method with a 5-fold cross-validation strategy. Figure A8 shows the ROC Curve for each of the three methods.[51] An ROC curve stands for receiver operating char-

[51]The number of rounds was selected based on a 5-fold cross-validation method.

acteristic curve and is typically used, along with the AUC statistic (Area under the curve) to assess the performance of an algorithm. The curve can be interpreted as the trade-off between sensitivity (share of true "risky" cases that are classified as such) and 1 minus specificity (or more tangibly in this setting, the share of safe pregnancies that are erroneously classified as "risky").[52] Better algorithms have ROC curves that approach the top left of the graph and consequently have high AUC statistics. As shown, the random Forest performs the best, although the three methods perform fairly similarly and fairly poorly – about 12 pp (or 24%) better than choosing randomly, which is depicted by the 45-line.[53]

Figure A8
ROC for Random Forest and XGBoosting



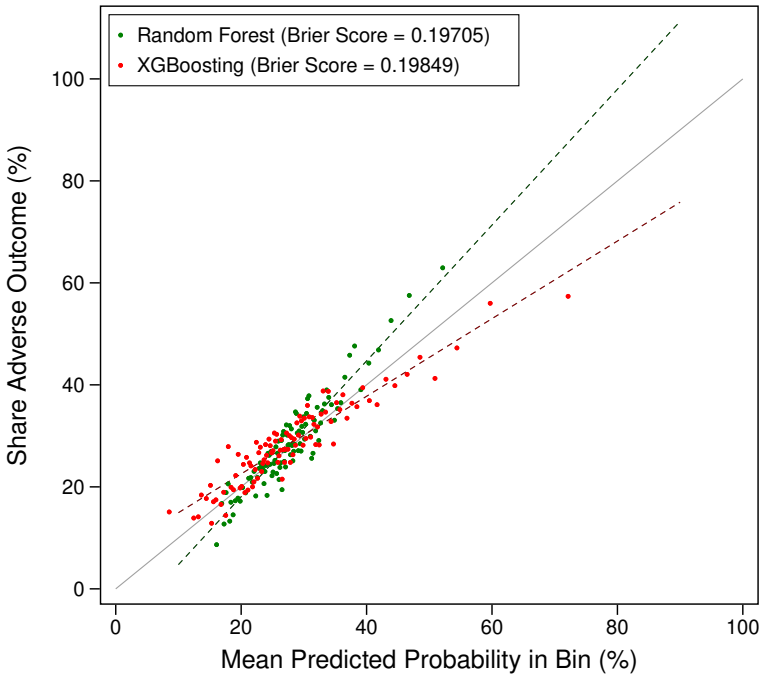
[52]For example, at the bottom left, using a high threshold of predicted probability such that none of the truly risky pregnancies are classified as such (sensitivity = 0) also means that the share of false negatives is zero (e.g. none of the safe pregnancies are labelled as risky).

[53]Worth noting here is that there are studies that have a much higher AUC at predicting preterm births – see (Surendiran et al., 2022; Bertini et al., 2022) for two literature reviews. However, these studies often use electronic health data, including results from cervical length measurements in ultrasounds. More generally, they use predictors from the pregnancy which is ideal for prediction but unsuitable here due to the endogeneity of increased diagnoses during pregnancy related to the carve-out.

While this exercise speaks to the limits of these predictive algorithms with claims data, it is worth recalling here that the goal of using a predictive algorithm was *not* to "classify" enrollees into an either safe bucket or a "risky" one, but rather to assign each enrollee with a continuous measure of predicted risk that reflects their ex-ante risk which was observable to the provider and to the plan. As such, I evaluate the performance of the three algorithms through two different metrics that are more relevant to the task at hand: (1) how the predicted risk measure deviates from the true risk and (2) monotonicity of risk prediction (a pregnancy rated as higher risk must have a higher true risk than a pregnancy predicted as a lower risk)

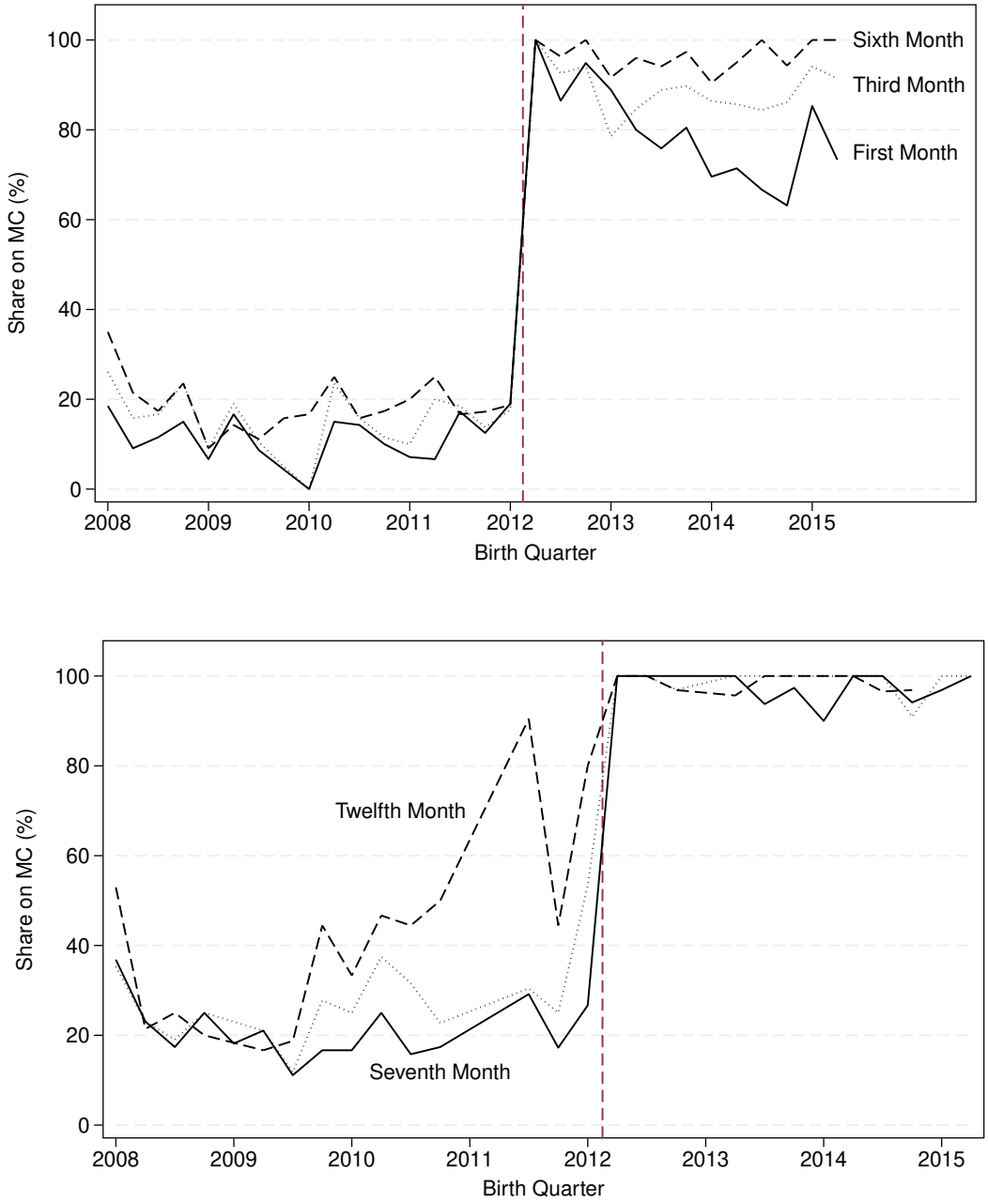
Given that I do not have a measure of true risk, I bin pregnancies by their predicted risk into 100 bins, and then calculate the share of pregnancies in each bin that did lead to a an adverse outcome. Figure A9 plots the true risk in a bin against the mean predicted risk in a bin. The results suggest that the true risk is highly correlated with the risk predicted from all three algorithms. The Brier score listed for each method indicates the average squared deviations of the predicted risk from the true risk. The results here suggest that again the three methods are almost identical, although the boosting mechanism performs slightly better than the other two.

Figure A9
Brier Scores



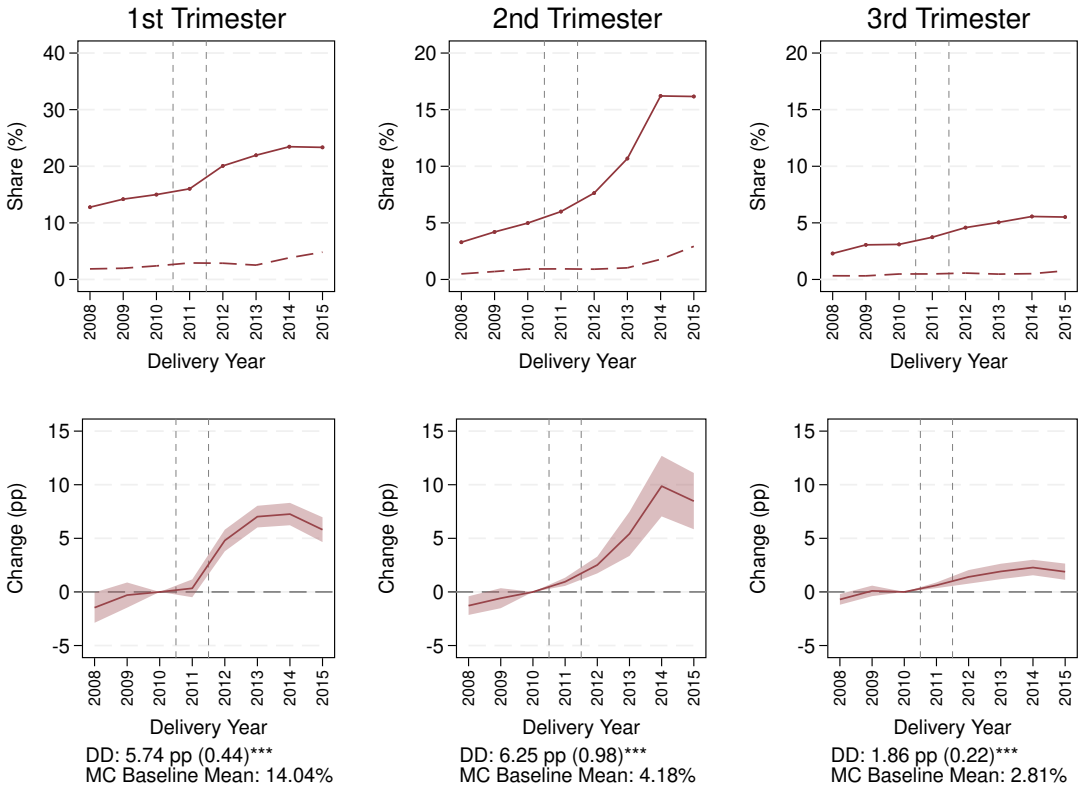
D First Stage

Figure A10
MC Coverage for VLBW Newborns - Removing Exiting Plans



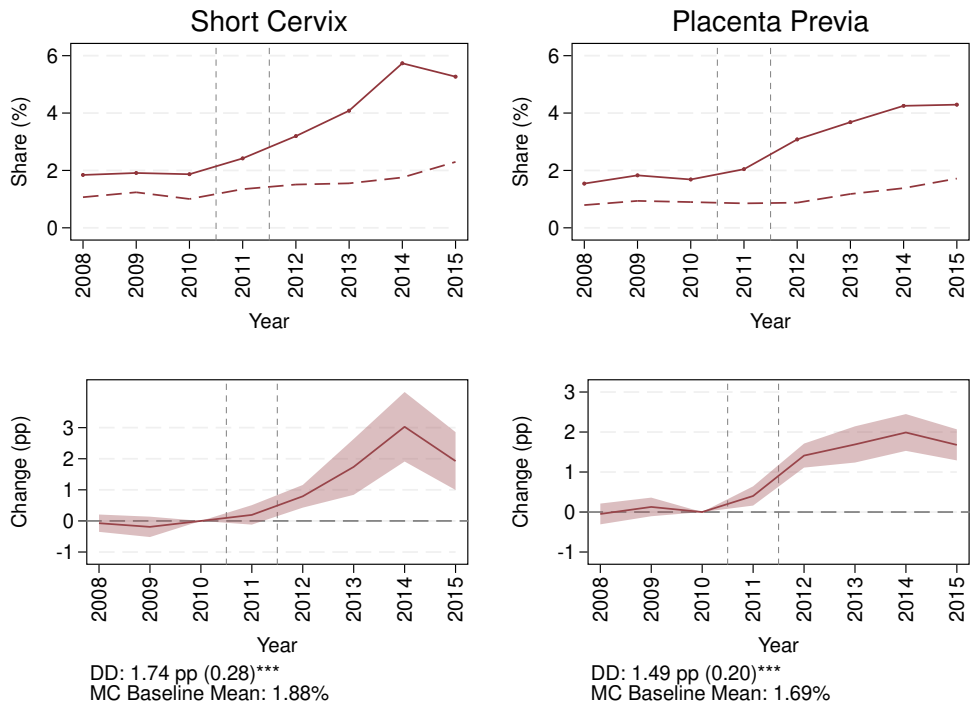
E Prevention Outcomes

Figure A11
Discretionary versus Routine Monitoring



Note: Monitoring of cervix length typically takes place once in the first trimester and is not typically included in standard prenatal packages. Monitoring of cervix length in the second and third trimester is usually discretionary.

Figure A12
Immediate Risk Factor Diagnoses



Note: Short cervix and placenta previa are two diagnoses that are made when the cervix length is monitored using transvaginal ultrasound.

Figure A13
Additional Risk Factor Diagnoses

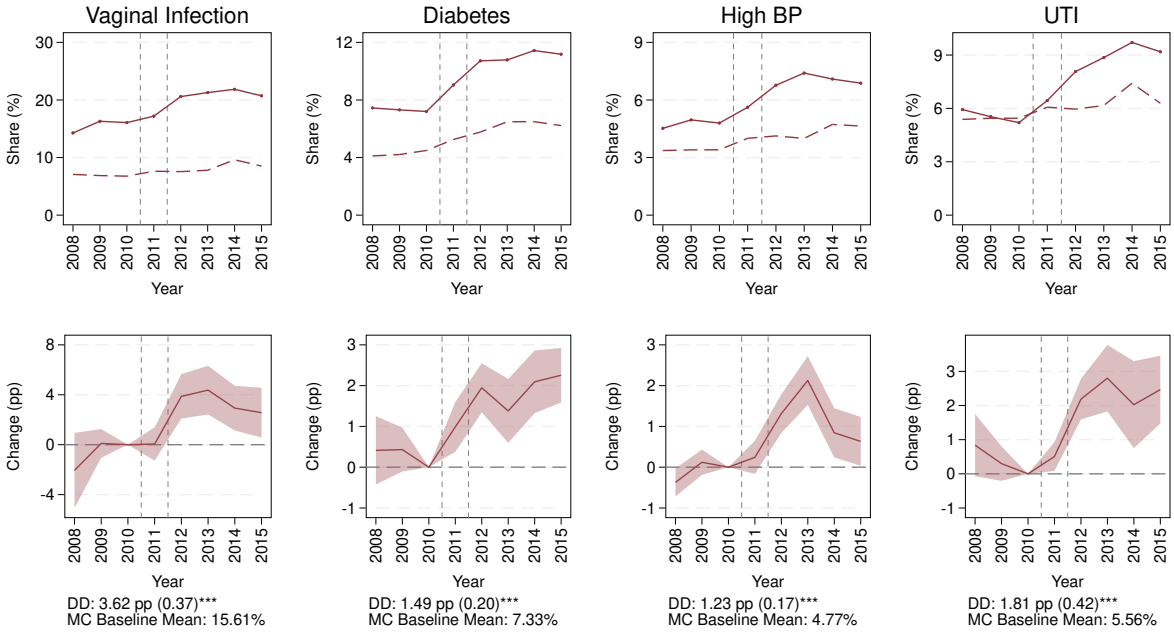
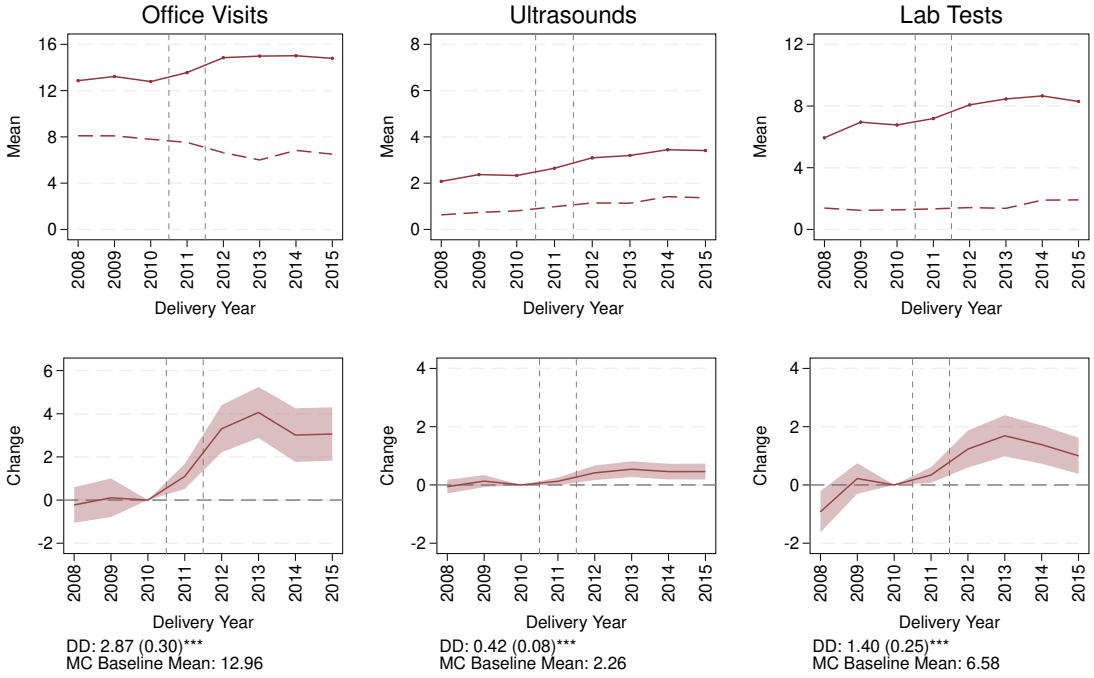


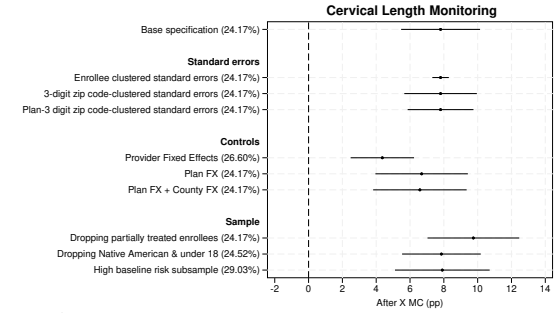
Figure A14
General Care Patterns



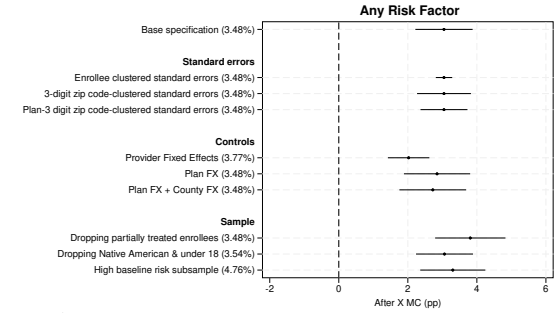
Notes: Solid lines indicate trends for private plan enrollees covered by MC plans. Dashed lines indicate trends for public plan enrollees covered by FFS.

F Robustness and Placebo Checks

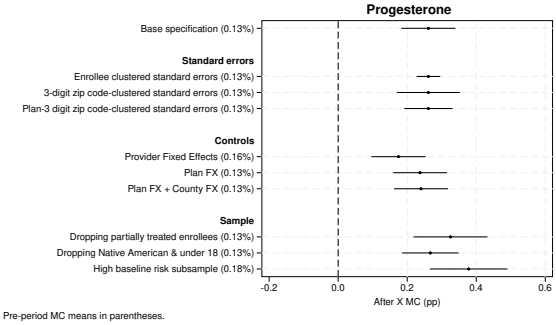
Figure A15
Robustness for Preventative Care Measures



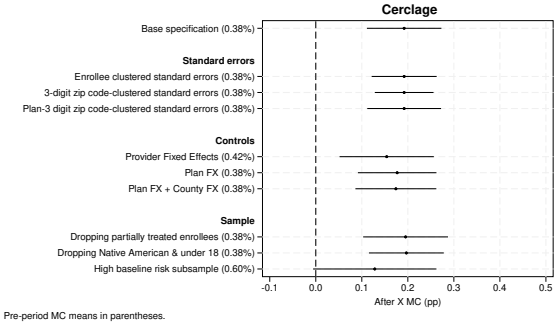
Pre-period MC means in parentheses.



Pre-period MC means in parentheses.



Pre-period MC means in parentheses.



Pre-period MC means in parentheses.

Figure A16
Changing the Reference Year in The Primary Specification

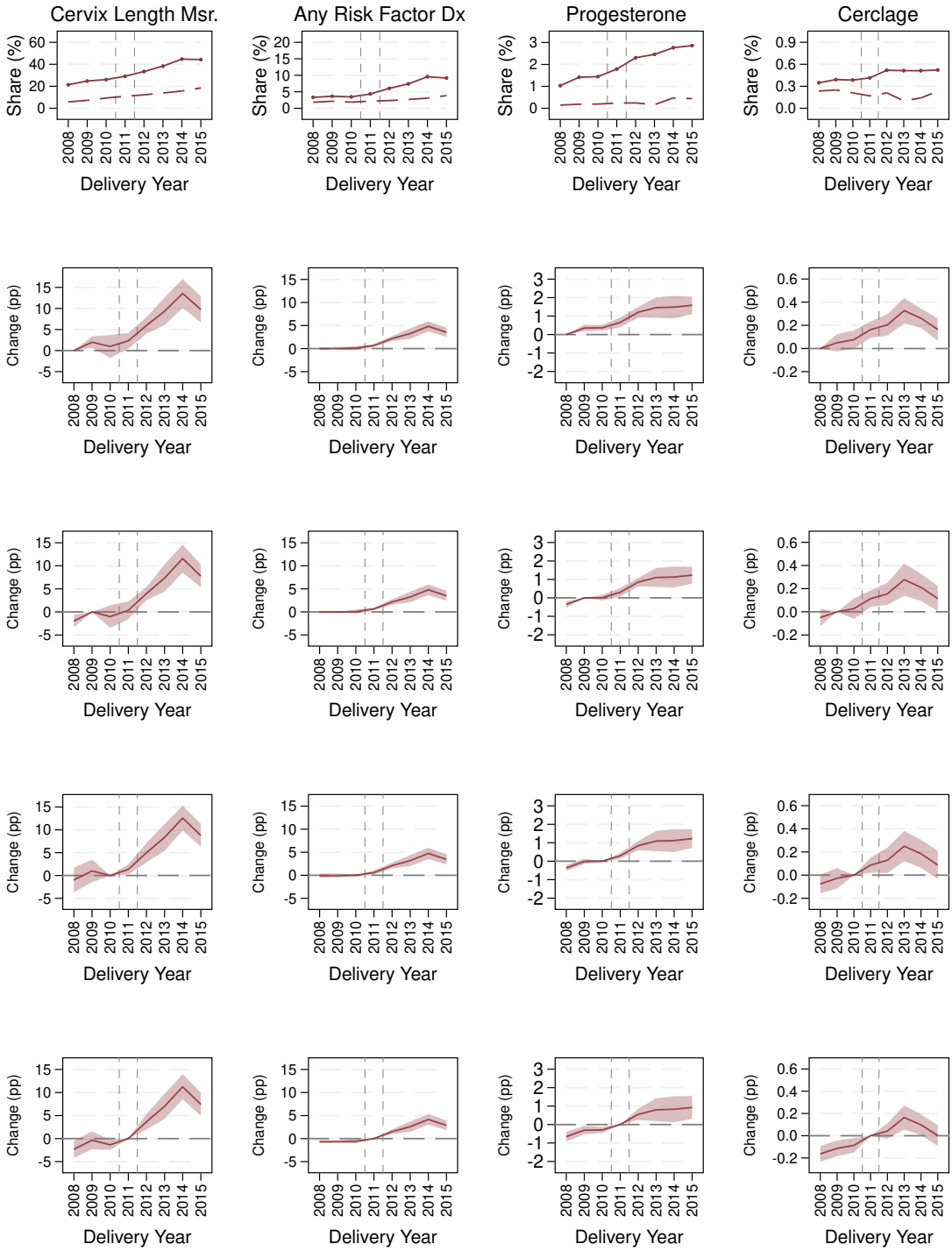


Figure A17

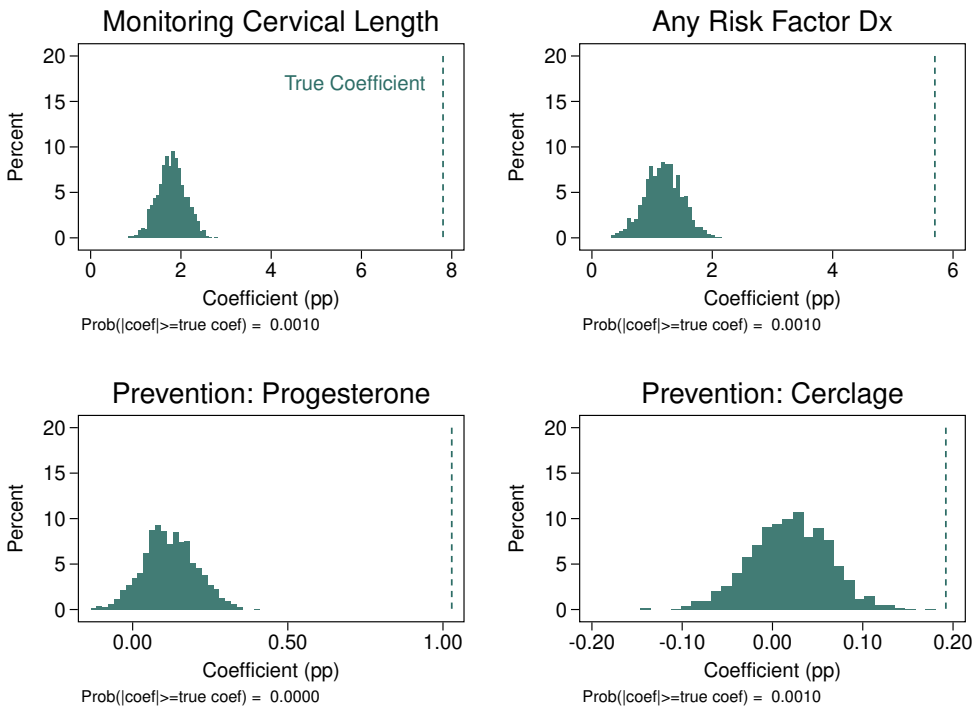

Monitoring Cervical Length
True Coefficient
Percent
Coefficient (pp)
Prob(|coef|>=true coef) = 0.0010
Any Risk Factor Dx
Percent
Coefficient (pp)
Prob(|coef|>=true coef) = 0.0010
Prevention: Progesterone
Percent
Coefficient (pp)
Prob(|coef|>=true coef) = 0.0000
Prevention: Cerclage
Percent
Coefficient (pp)
Prob(|coef|>=true coef) = 0.0010

Figure A18

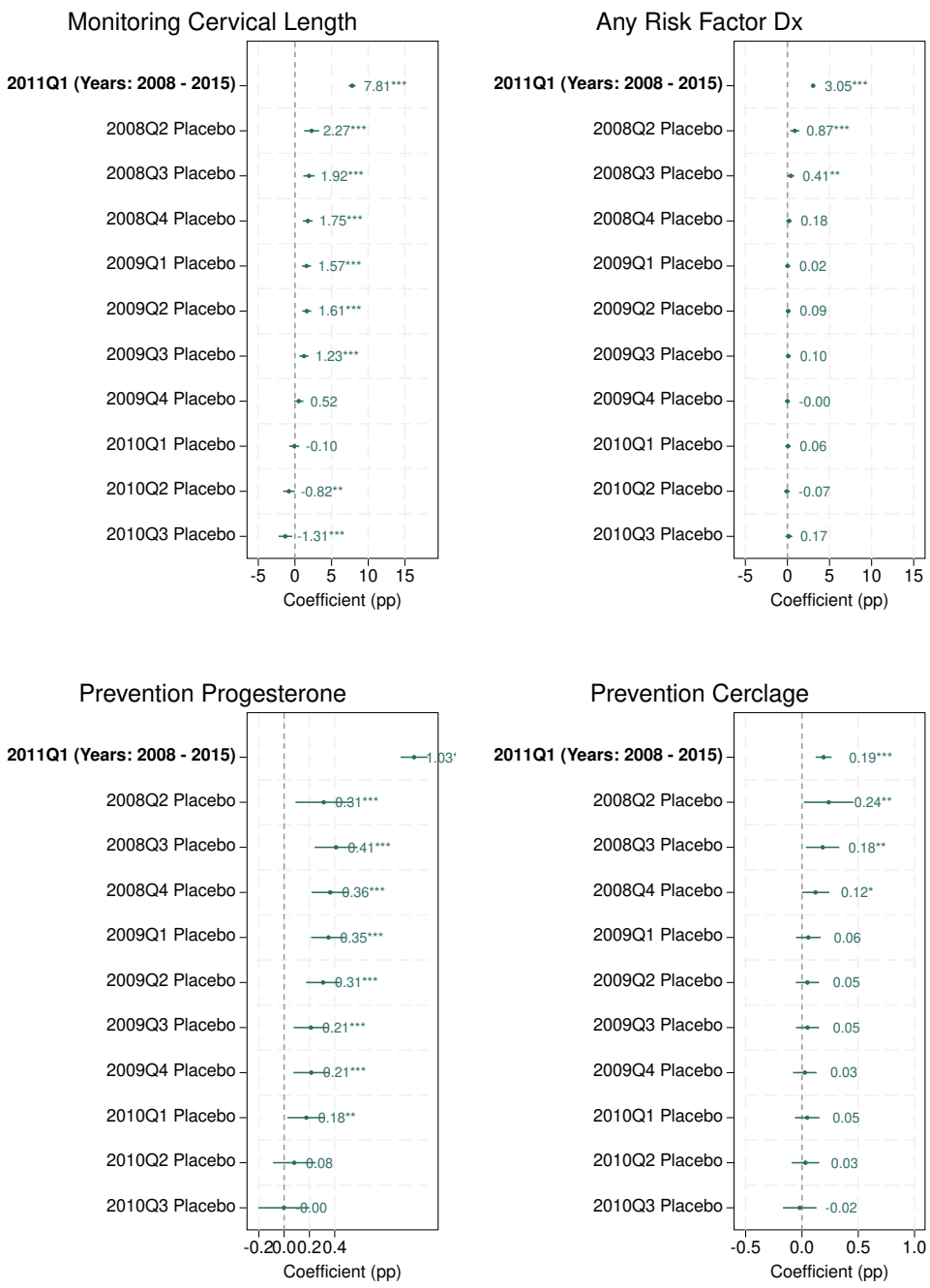


Figure A19
Robustness for Newborn Birth Outcomes

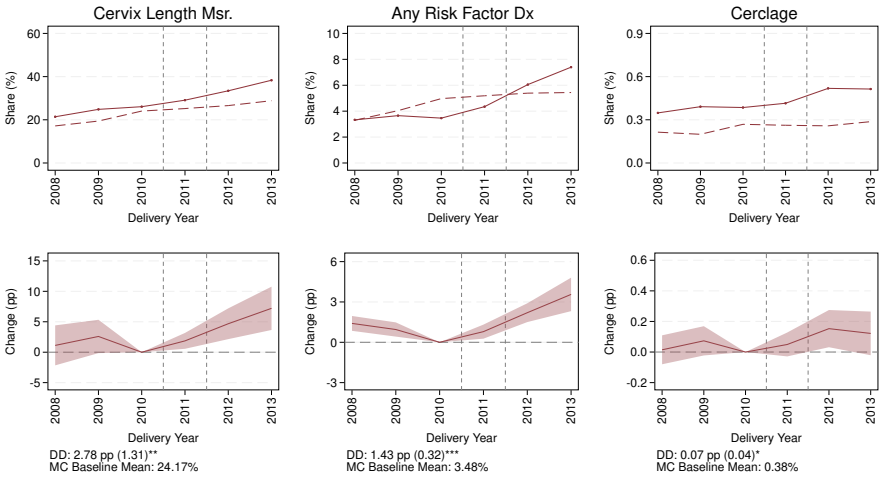
G Other Control Groups

I test the primary set of results using an additional control group: MC enrollees in other states. To select which states qualify to become a control group, I use two criteria. First, a control state needs to have a large MC program, much like NY does. Instead of applying my own benchmarks, I used the categories developed by a GAO report - specifically, I used states in groups 2 through 4 which are states that stirred away from primary care management models and leaned more heavily towards MC. Second, a control state must have MAX data available for the 6 years of the study.

Second, using the resulting control group, I plot my outcomes of interest over time to eliminate any states with MC reporting that started within the pre-period. Including these states would allow large and discontinuous changes in these outcomes that are a product of the data availability rather than actual changes on part of the MC plans. Finally, I eliminated California given concerns about the quality of MAX data from this state (CMS, 2019). After applying these criteria, there are nine remaining states: New Jersey, Minnesota, Michigan, Arizona, Delaware, New Mexico, Wisconsin, Rhode Island, and Maryland.

Figure A20
Using MC Enrollees in Nine Other States

(a) Control: Nine States With MAX Data up to 2013: NJ, MN, MI, AZ, DE, NM, WI, RI, MD



(b) Control: Five States With MAX Data up to 2014: NJ, MN, MI, AZ

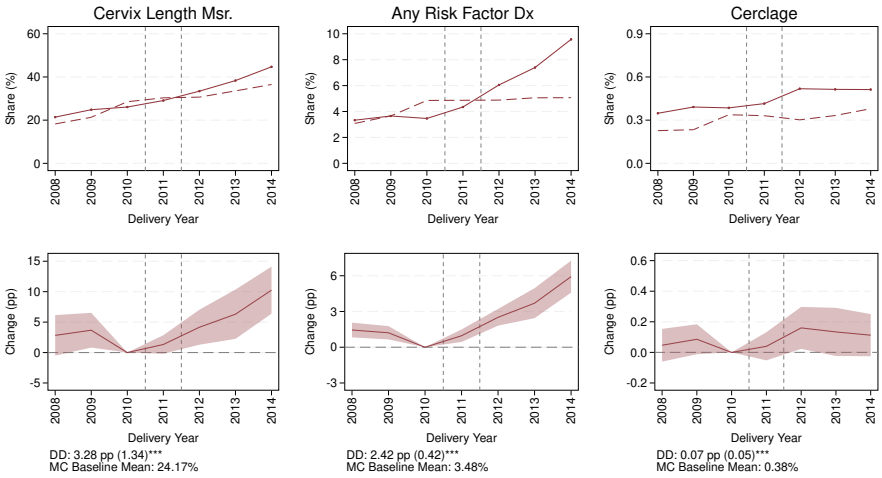
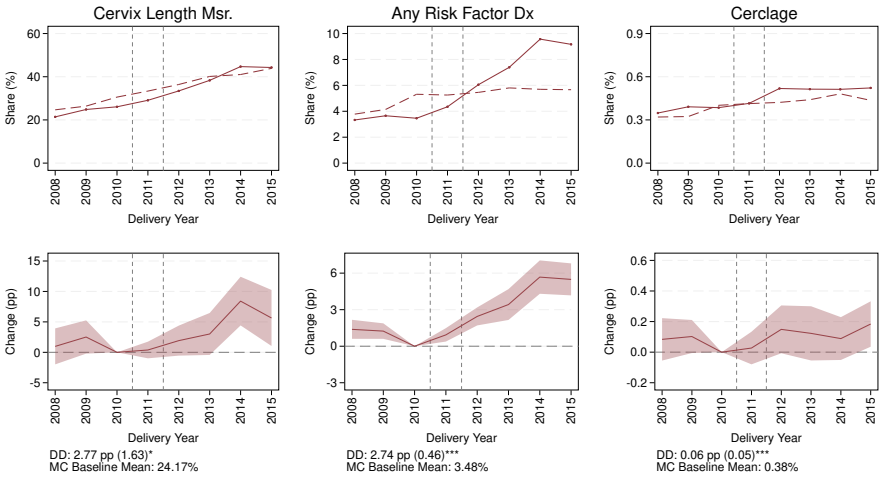


Figure A21
Continued: Using MC Enrollees in Nine Other States

(a) Control: Three States With MAX Data up to 2015: NJ, MN, MI



H Compositional Change

Figure A22
Changes in Predicted Risk of Covered Enrollees - Random Forest-Based Score

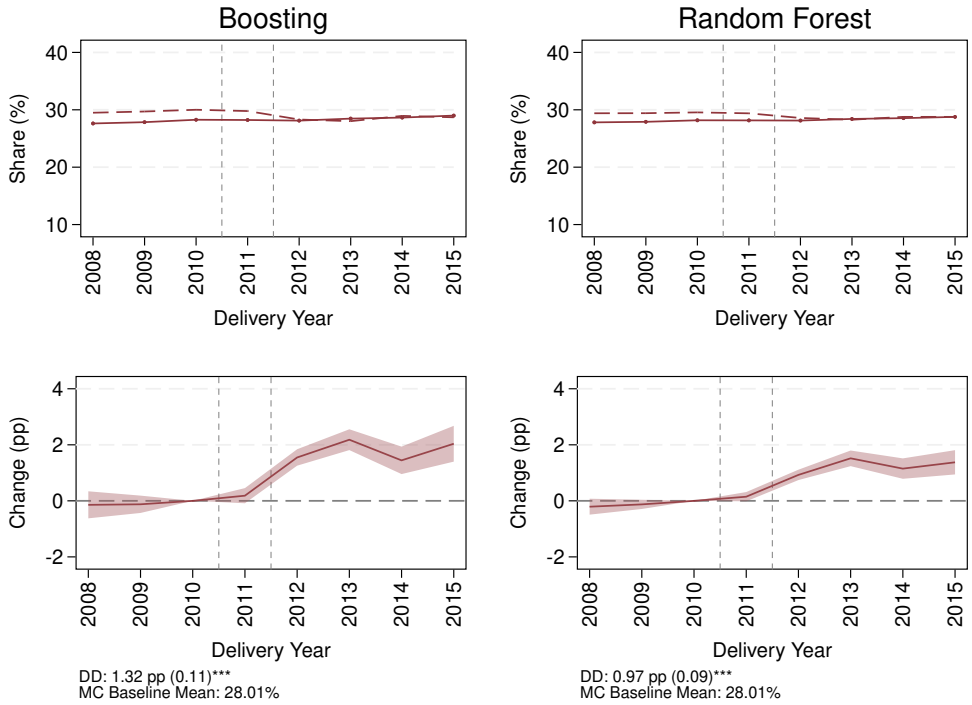


Figure A23
Primary Results, Controlling for Risk Score

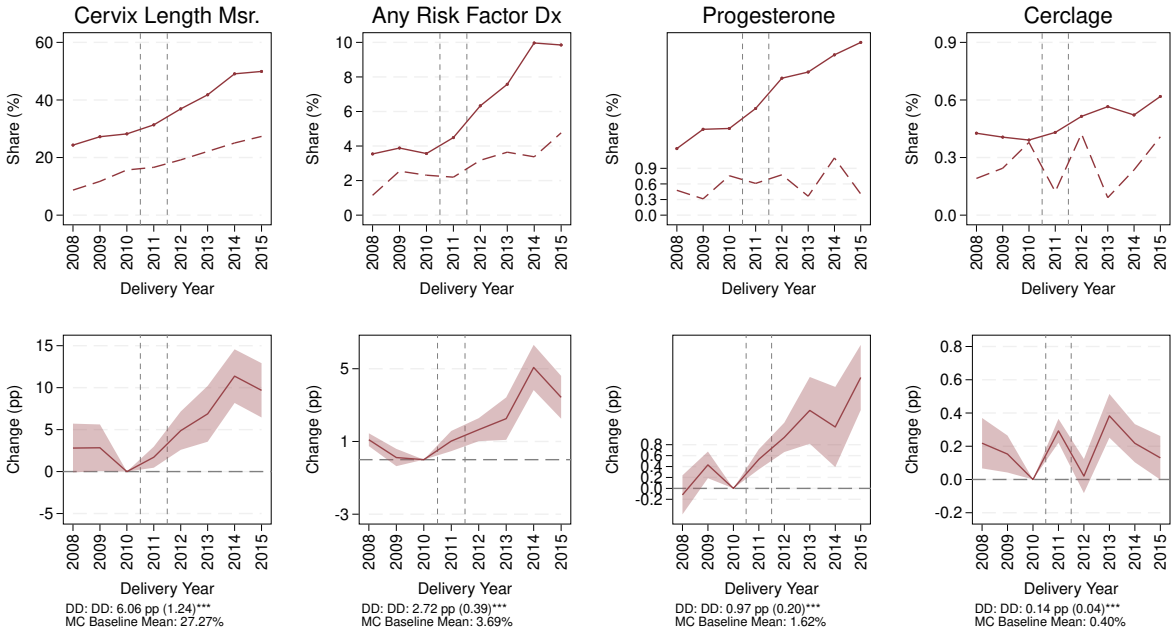


Figure A24
Primary Results in Subsample with No Risk Predictors

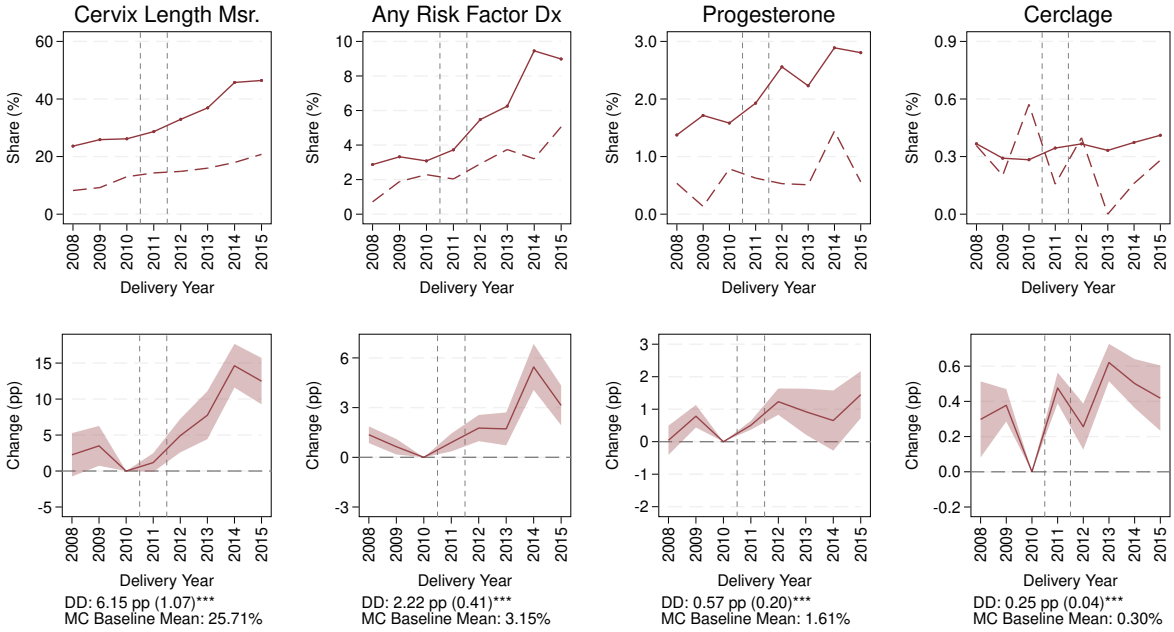
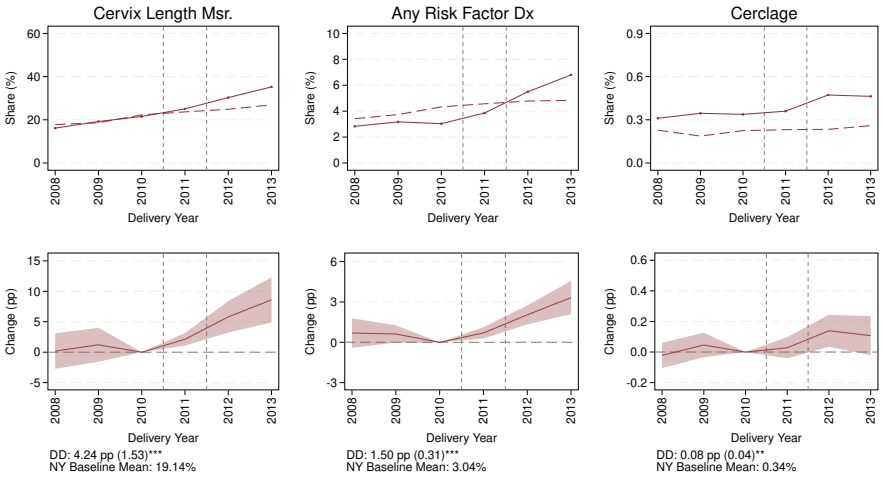


Figure A25
Comparing to Other States with MC and FFS Included

(a) Control: Nine States With MAX Data up to 2013: NJ, MN, MI, AZ, DE, NM, WI, RI, MD



(b) Control: Five States With MAX Data up to 2014: NJ, MN, MI, AZ

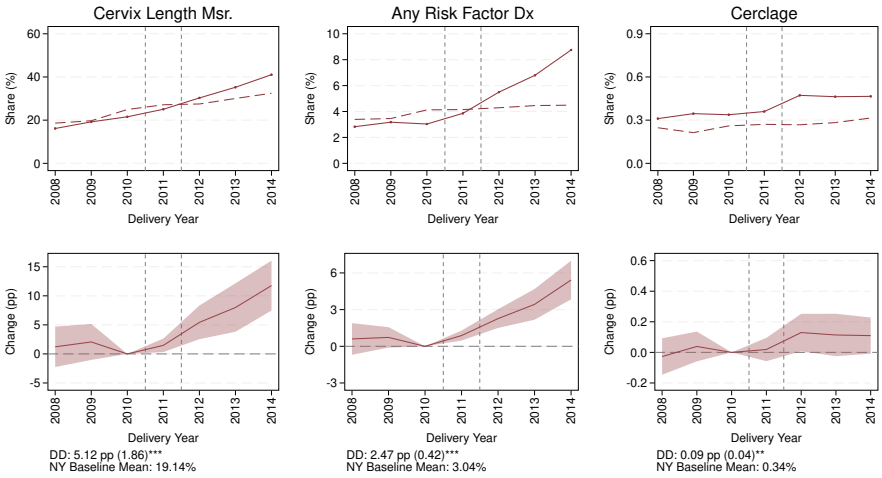
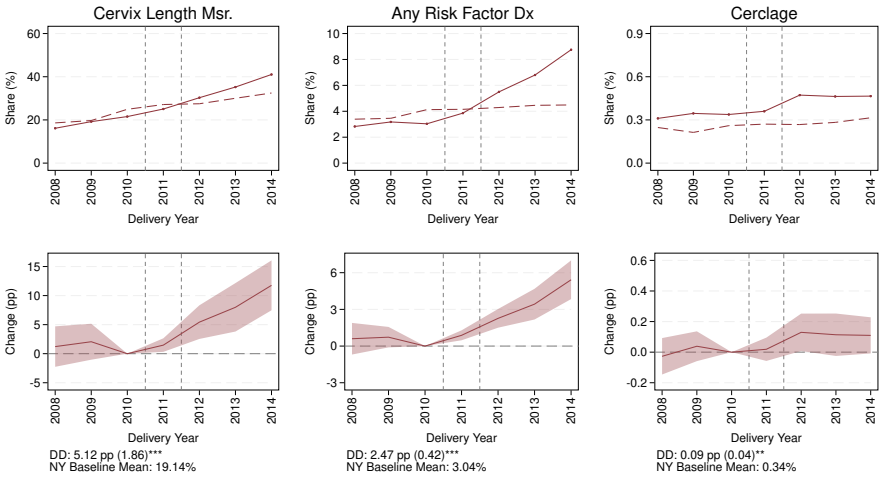


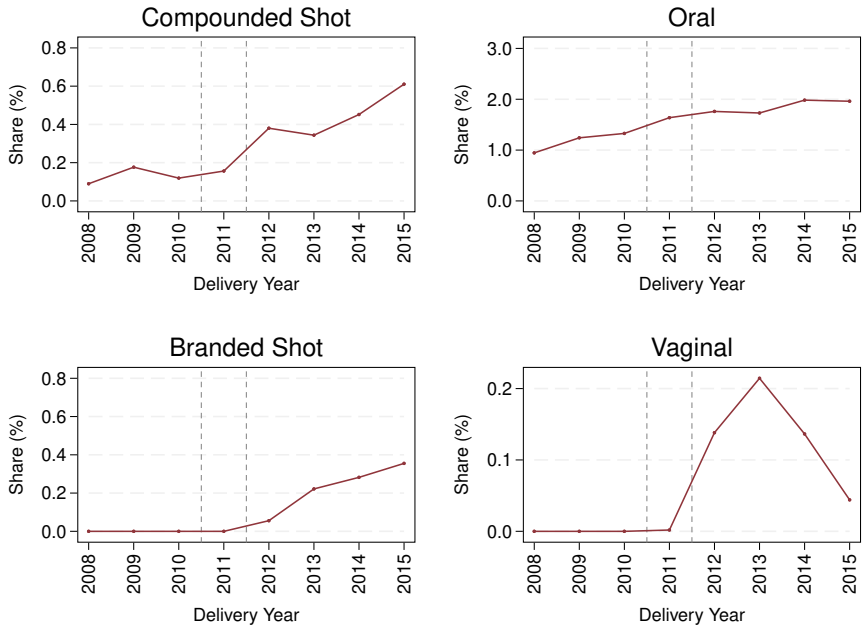
Figure A26
Comparing to Other States with MC and FFS Included

(a) Control: Three States With MAX Data up to 2015: NJ, MN, MI



I Progesterone Types

Figure A27
Changes in Old and New Progesterone Types



FFS trends are not shown due to small cell sizes for non-oral progesterone types.

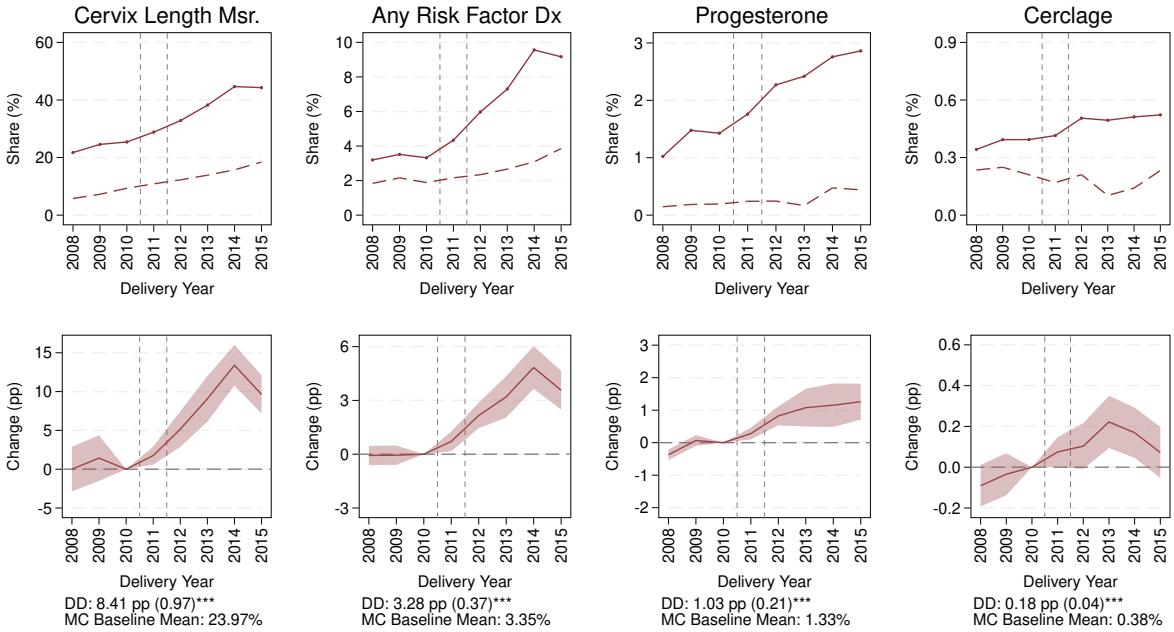
J Plan Exits

Table A6
Changes in the MC Landscape in the Post-Period

	(1)	(2)	(3)	(4)
Plan	Exiting Counties	Exiting from all service area	Acquired	Sources
Healthnow	Genesee and Niagara	No, remained in other counties	No	Enrollment Reports
Healthplus	Nassau	No, remained in NYC	No	Enrollment Reports
Excellus	Oswego and Onondaga	No, remained in other counties	No	Enrollment Reports
SCHC Total Care	Oswego and Onondaga	Yes	Acquired by Universal American Corp and run by Today's Option, which NY Medicaid lists as another plan but only in Onondaga and not in Oswego.	Source
Neighborhood	NYC	Yes	Acquired by HealthFirst	Source
NY Presbyterian Select Health Plan SN	NYC	Yes	Acquired by VNS	Source

Note: The primary source is the NY Medicaid Enrollment reports published online by the NY Department of Health. There were 18 distinct plans that were operating in the included counties in December 2010. The plans listed above are those that had stopped operating in any county or changed ownership by mid-2015.

Figure A28
Prevention without Churn from Plan Exits



K Distribution

Figure A29
Preventative Care, by Race and Ethnicity

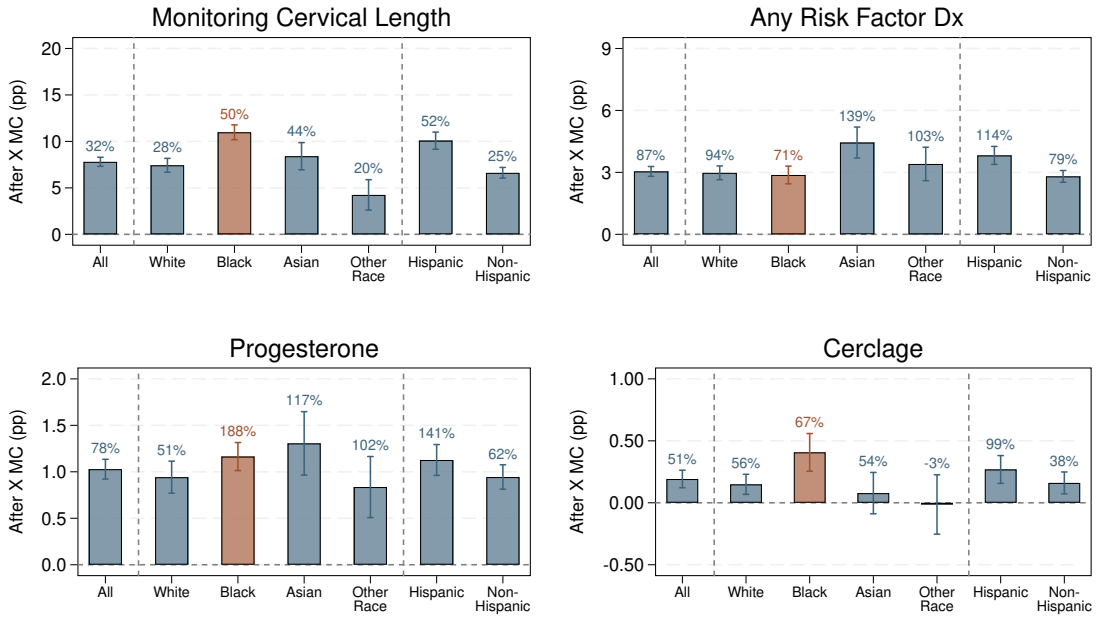
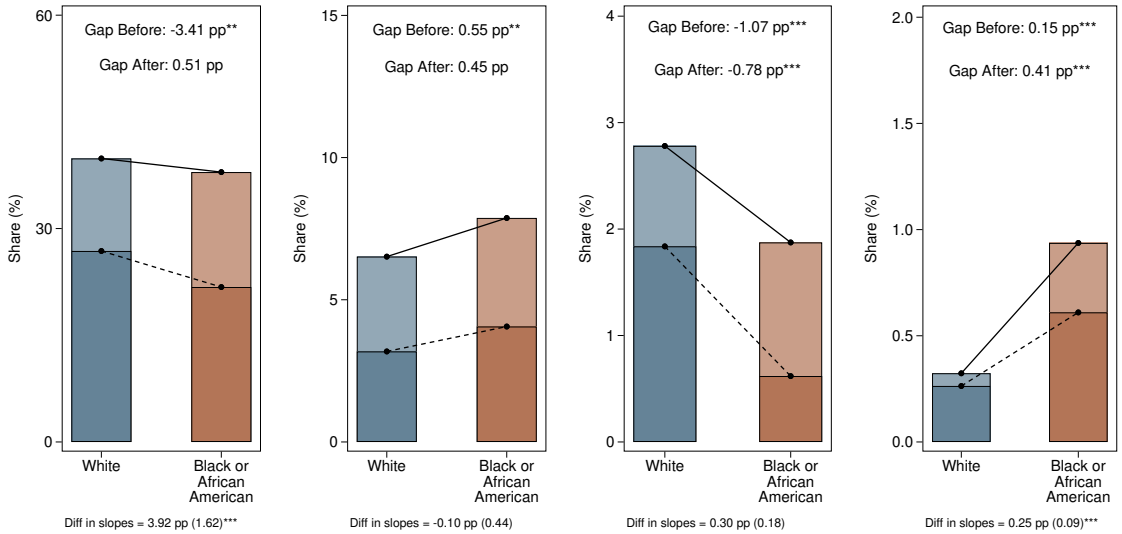
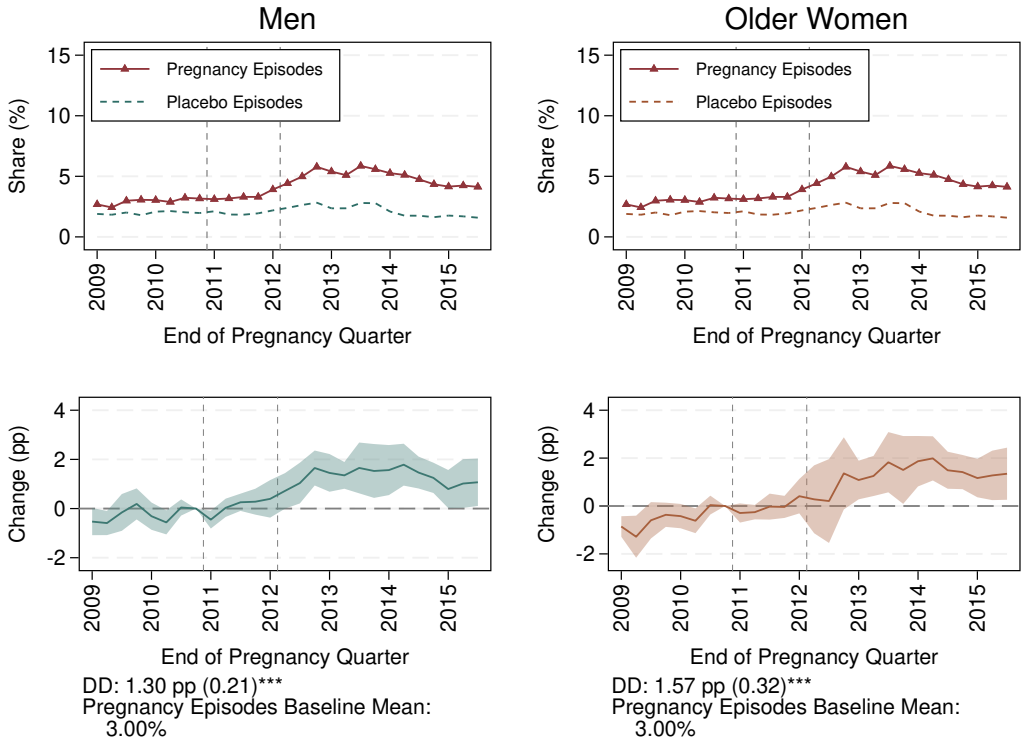


Figure A30
Preventative Care among African American Enrollees Relative to White Enrollees



L Cream-Skimming

Figure A31
Continuity of Coverage, Compared to Placebo Episodes



M Networks and Providers

Figure A32
Change in MC Provider Networks - by Specialty

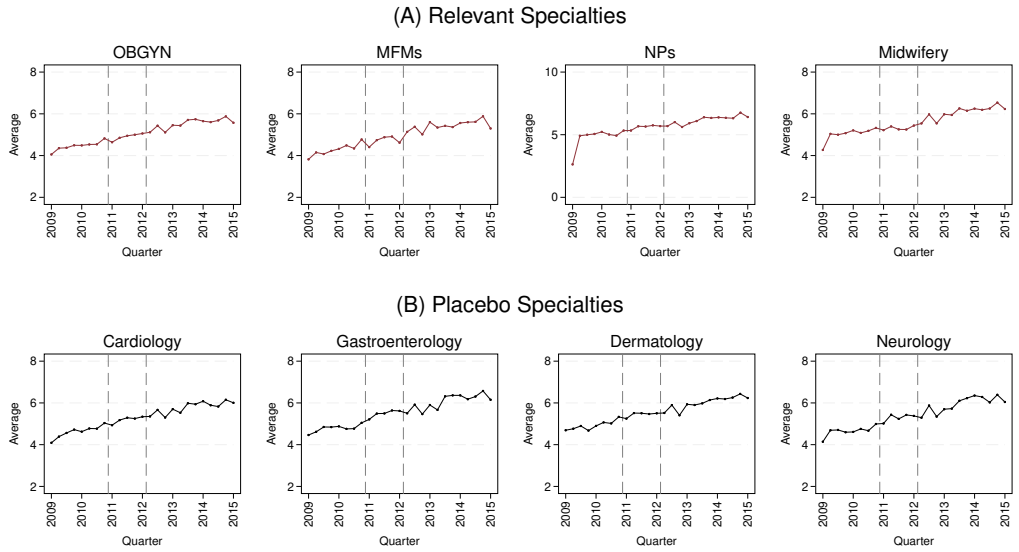


Figure A33
Estimates After Controlling for the Primary Physician

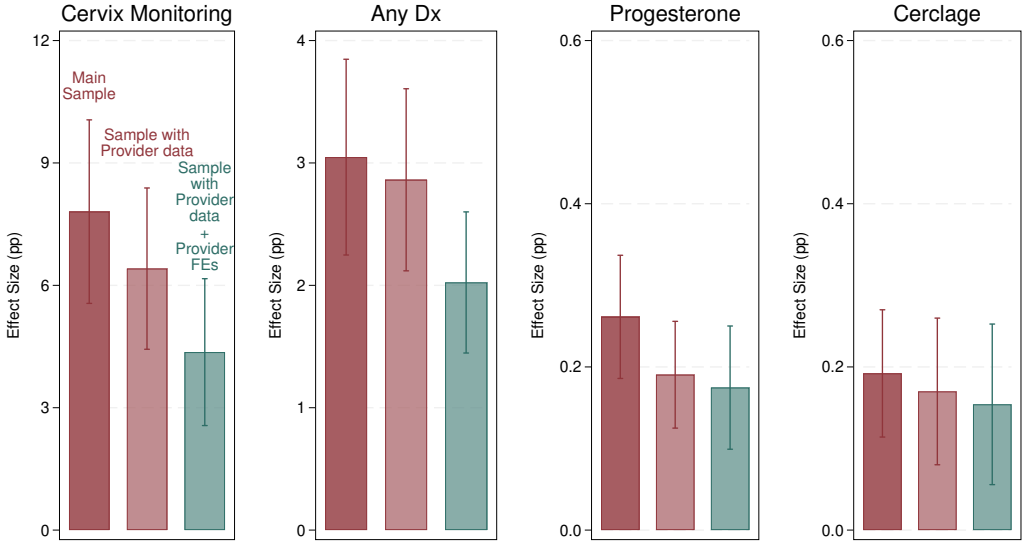
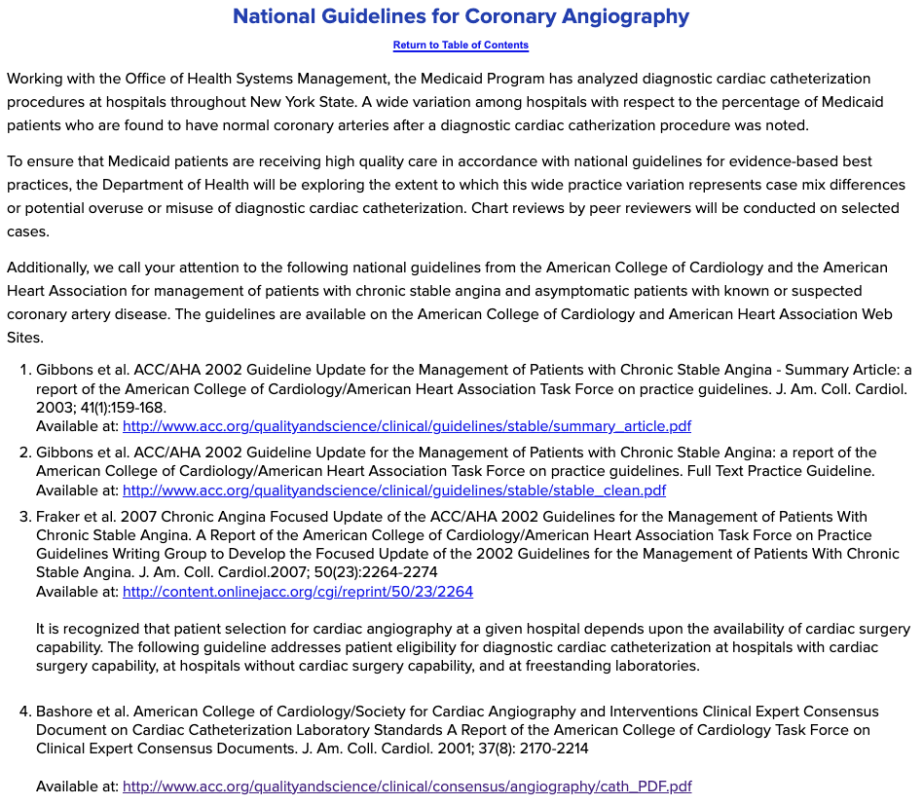


Figure A34
MC Wait Times, As a Proxy for Prior Authorization

The progesterone refers here to compounded 17P and Makena. I do not include oral progesterone or vaginal progesterone since both of these were considered prescription benefits which MC plans were not responsible for covering until October 2011.

Figure A35
Example of FFS Communication to Providers



National Guidelines for Coronary Angiography

Return to Table of Contents

Working with the Office of Health Systems Management, the Medicaid Program has analyzed diagnostic cardiac catheterization procedures at hospitals throughout New York State. A wide variation among hospitals with respect to the percentage of Medicaid patients who are found to have normal coronary arteries after a diagnostic cardiac catherization procedure was noted.

To ensure that Medicaid patients are receiving high quality care in accordance with national guidelines for evidence-based best practices, the Department of Health will be exploring the extent to which this wide practice variation represents case mix differences or potential overuse or misuse of diagnostic cardiac catheterization. Chart reviews by peer reviewers will be conducted on selected cases.

Additionally, we call your attention to the following national guidelines from the American College of Cardiology and the American Heart Association for management of patients with chronic stable angina and asymptomatic patients with known or suspected coronary artery disease. The guidelines are available on the American College of Cardiology and American Heart Association Web Sites.

1. Gibbons et al. ACC/AHA 2002 Guideline Update for the Management of Patients with Chronic Stable Angina - Summary Article: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. J. Am. Coll. Cardiol. 2003; 41(1):159-168.
   Available at: http://www.acc.org/qualityandscience/clinical/guidelines/stable/summary_article.pdf
2. Gibbons et al. ACC/AHA 2002 Guideline Update for the Management of Patients with Chronic Stable Angina: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. Full Text Practice Guideline.
   Available at: http://www.acc.org/qualityandscience/clinical/guidelines/stable/stable_clean.pdf
3. Fraker et al. 2007 Chronic Angina Focused Update of the ACC/AHA 2002 Guidelines for the Management of Patients With Chronic Stable Angina. A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines Writing Group to Develop the Focused Update of the 2002 Guidelines for the Management of Patients With Chronic Stable Angina. J. Am. Coll. Cardiol.2007; 50(23):2264-2274
   Available at: http://content.onlinejacc.org/cgi/reprint/50/23/2264

   It is recognized that patient selection for cardiac angiography at a given hospital depends upon the availability of cardiac surgery capability. The following guideline addresses patient eligibility for diagnostic cardiac catheterization at hospitals with cardiac surgery capability, at hospitals without cardiac surgery capability, and at freestanding laboratories.

4. Bashore et al. American College of Cardiology/Society for Cardiac Angiography and Interventions Clinical Expert Consensus Document on Cardiac Catheterization Laboratory Standards A Report of the American College of Cardiology Task Force on Clinical Expert Consensus Documents. J. Am. Coll. Cardiol. 2001; 37(8): 2170-2214

   Available at: http://www.acc.org/qualityandscience/clinical/consensus/angiography/cath_PDF.pdf

Image captured from NY Medicaid webpage published in February, 2008. As discussed in section 5.6, the communication does not include any high-powered incentives or supply side constraints.