Subsidizing contraception: Effects on take-up of highly effective contraceptives and on unintended births*

Sarah Rosenberg†

September 8, 2022

Abstract

Since the 1990s, expansions of Medicaid programs have extended subsidized coverage of prescription contraceptives to millions of women. If such programs induce women to take up more effective contraceptives, they could reduce rates of unintended births, which account for about a third of all births in the U.S. Using data from the State Drug Utilization Program between 1992 to 2019, I show that these coverage expansions at best increased coverage of highly effective contraceptives by about 60 to 80 per 1,000 women of childbearing age per year that would have been eligible for the expansions, although the true effect on usage is likely smaller. Using both U.S. Vital Statistics data and the American Community Survey for births, I find the Affordable Care Act expansion reduced teen fertility by about 6%, but over all other groups no evidence of significant effects. These findings suggest that while subsidization of contraceptives may be an important policy lever to reduce teen births, we must look to other approaches to reduce unintended births to non-teens.

Keywords: Contraception, fertility, public health insurance

JEL: I13, J13

1 Introduction

The estimated number of unintended births in the United States has been falling continuously for more than a decade, breaking from the trend in estimated intended births, which increased for much of the same period (Buckles et al., 2022). At the same time, unintended births still account for about a third of all births. To understand how to further reduce the rate of unintended births, which include births that mothers reported as having been unwanted or severely mistimed at the time of pregnancy (Guzzo and Hayford, 2011), we need to know what policy levers may affect them.

In a mechanical sense, to reduce unintended births it is necessary to induce individuals to use more effective contraceptive methods, since contraceptive failure or inconsistent usage of contraceptives are the proximate causes of unintended births. How can policy generate this behavioral change?

An obvious candidate policy for reducing unintended births is reducing the financial barriers to highly effective contraceptives, since the most effective contraceptives are also the most expensive. Prior work on this question has primarily studied subsidization via funding of publicly supported

*

†Lund University, Department of Economics. Email: sarah.rosenberg@nek.lu.se

clinics, finding negative effects on young women's fertility when highly effective methods are encouraged and made free via this channel (Bailey, 2012; Lindo and Packham, 2017; Lindo et al., 2020; Kelly et al., 2020). However, all contraceptives are not equal, and programs to provide less effective contraceptives for free can even backfire when not coupled with appropriate education, as Buckles and Hungerman (2018) found with respect to condom distribution programs in high schools. Relatedly, for women already seeking reproductive care at Planned Parenthood clinics, discounting the up-front out-of-pocket costs of contraceptives makes women more likely to take up the most effective methods of contraceptives (Bailey et al., 2021). The results of these studies suggest the worth of asking whether these successes can be replicated at scale by eliminating out-of-pocket costs for contraceptives for a broad group of women, or whether they were a result of particularly effective and targeted programs.

In fact, this policy experiment has already been widely conducted in the United States. Over the course of the last three decades, subsidization of contraceptives for low-income women has expanded in waves across many states, through two programs: family planning waiver programs and the Affordable Care Act Medicaid expansions. Although both comprise public insurance, these programs differ by whether they provide targeted family planning coverage or full medical insurance to low-income individuals, and their income eligibility thresholds differ as well.

I estimate the effects of both family planning waiver programs and the Affordable Care Act Medicaid expansions on prescribing of different types of contraceptives as well as on fertility. In accordance with previous work, I find that subsidization via public insurance expansions decreases births for teens, but I also find that it has no measurable effect on births for non-teens.

For economists, the lack of an effect for economic incentives for non-teens may be surprising. However, descriptive and qualitative work on the usage of contraceptives provides ample reason to consider that financial barriers may not be a first-order issue for why women not actively trying to become pregnant may use less effective or inconcistent contraceptive methods. Contraceptive usage, including prescription contraceptives, are already widespread among women in the U.S., including uninsured women, among whom about half use a prescription method of contraceptives (Frost and Darroch, 2008). Ambivalence about pregnancy is moreover strongly related to whether and how consistently women use contraceptives Yoo et al. (2014). Planning and execution skills also may be a factor leading to inconsistent or ineffective usage England et al. (2016). Reducing the financial costs of contraceptives may not be sufficient to shift these factors for those not already intent on using effective forms of birth control.

I make three specific contributions relative to the existing works on the effects of expanding public health insurance on contraception and fertility (Kearney and Levine, 2009; Lindrooth and McCullough, 2007): first, whereas the prior studies look at only early expansions of family planning waiver programs, I extend the temporal window of analysis, given many additional policy changes between 2002 and the present day. Second, I use state of the art methods in estimation of difference-in-difference models to ensure that estimates reflect an average treatment effect that is useful and relevant for thinking about policy. Concretely, choices about weighting in the estimation of these models explain the main differences between the results of this paper and the prior work on family planning waiver prograams.

Third, I incorporate new data to yield relevant insights. I use both administrative data on Medicaid reimbursement of prescription contraceptives from the State Drug Utilization Dataset between 1992 and 2019, and on births with the American Community Survey between 2001 and 2019, in addition to the U.S. Vitality Statistics Natality data that is commonly used in studies of fertility.

The SDUD allows for assessing the effects of subsidization on takeup of different types of contraceptives, particularly the most effective types, known as long-acting reversible contraceptives. As previously mentioned, the key to mechanically reducing unintended fertility is to incentivize use of more effective contraceptives. This data allows us to assess whether these programs in fact generate increases in prescriptions for such contraceptives. While Kearney and Levine (2009) do look at effects on any contraceptive usage from the National Survey of Family Growth, this survey is fairly small and it is not possible to distinguish even prescription contraceptives, much less the types of prescription contraceptives.

Moreover, based on estimated treatment effects for the reimbursement of different types of prescription contraceptives via these public insurance programs, I can predict plausible ranges of possible reductions in unintended fertility, which aids in planning and interpreting an analysis of the actual fertility effects.

With respect to prescription contraceptives, I find that both family planning waiver programs and full Medicaid expansions under the Affordable Care Act lead to robust and measurably significant increases in the rate of reimbursement of prescription contraceptives relative to the female population. Notably, prescriptions for short-acting hormonal contraceptives increase much more than longer-acting contraceptives, which also exhibit modest increases. Since these reflect Medicaid payment of prescriptions rather than use, it's not possible to assess whether the reflect an increase in the rate of short- versus longer-acting contraceptives. However, the most plausible extrapolations about the likely fertility effects of these increases suggest that they would decrease births by perhaps one to two births per thousand women ages 15-44 on average, relative to baseline levels of births of between 60 to 70 births per thousand women. These predictions also highlight the potential difficulty of precisely estimating such a treatment effect, and the importance of attempting to maximize power in analysis of the effects on fertility.

As with many analyses of fertility, I first study effects on aggregate fertility rates, constructed by using administrative birth counts from U.S. Vital Statistcs and dividing by population estimates for the relevant group from the U.S. Census Bureau. However, administrative birth data does not have information on income, and the U.S. Census Bureau does not provide population estimates of parents and non-parents. As a result, this analysis of aggregate fertility rates is unable to zero in on the groups that directly become eligible for insurance under these policies. I supplement the analysis using data from the ACS by using detailed individual economic characteristics in combination with information on births to study newly eligible populations specifically.

Interestingly, results using both Vital Statistics Natality data and the ACS suggest that the Affordable Care Act expansions specifically led to meaningful reductions in teen fertility of about 6%, adding to the existing evidence on the value of subsidizing contraceptives to reduce teen births. The fact that I find no such reduction for the family planning waiver programs, in contrast, could potentially owe to the fact that the use of the most effective (and expensive) contraceptives were not encouraged for adolescents and nulliparous women until 2009, just before the passage of the Affordable Care Act.

For non-teens, I find little evidence that either type of policy affected births, even when restricting only to eligible individuals within the ACS. The one exception is that I find some evidence of reductions in fertility for eligible parents under family planning waiver programs. Among non-significant estimates, confidence intervals for the estimated effects of each type of policy on births, while in most cases centered around zero, still often include potential effects consistent with between 3 to 5% reductions or increases in births. This range is however substantially smaller in magnitude than the roughly 9 to 10% reductions in births that Kearney and Levine (2009) estimate for newly eligible individuals, and smaller than the extrapolated estimates of Bailey et al. (2021), who estimate a comparable reduction in births would occur from only eliminating out-of-pocket costs to Title X patients. These results imply that large-scale effects of reducing the financial costs of contraceptives may be more modest than prior work on more targeted programs has suggested.

In a broader sense, this paper contributes to attempts to understand what causes and what affects unintended births by helping to answer whether further efforts in both research and policy should be made in regard to financial barriers to contraceptives, or whether we would be better off studying other aspects and policies that may affect unintended births in terms of behavioral and social factors. The results support the conclusion that subsidization of even highly effective contraceptives has modest if any effects on fertility for non-teens, and takeup of subsidized contraceptives likely represents primarily a financial transfer to low-income individuals. In turn, these findings point to the importance of exploring other possible explanations and policy levers to affect unintended births. In particular, compared the study of financial incentives, behavioral or social factors that may contribute to inconsistent or ineffective use of contraceptives has received little attention in economics.

Moreover, while the focus in the contraceptive literature has largely been from the perspective of what may affect the choices of the patient, there is also another perspective that may be important that has been less studied: the influence of provider behavior. If, for example, patients' ambivalence about pregnancy or concerns about side effects are relevant to their choices of contraceptives, then having providers who are more or less effective at encouraging women to use more effective contraceptives may also play an important role that would be worth exploring in further research.

To provide background for the subsequent analysis, Sec. 4 characterizes the different types of prescription contraceptives, followed by an explanation of the institutional setting and policy changes of interest for the paper in Sec. 3. Sec. 5 discusses the identification strategy and assumptions associated with different estimators, as well as why it is beneficial to compare results from more than one estimator. Sec. 6 builds upon the discussion of identification to describe how each of the estimators is implemented, and implications for bias and efficiency. Finally, Sec. 7 and Sec. 8 proceed to the analysis of the estimated effects on contraceptive reimbursements and births respectively, characterizing in turn the data used, the implementation of estimation, results, and a discussion of their implications. Sec. 9 concludes.

2 Background on contraceptives

Contraceptive technologies have evolved substantially over the last half-century. Whereas behavioral methods and barrier methods of contraception have long been available (including condoms, fertility awareness, and withdrawal), the advent of the birth control pill in the U.S. in the 1960s marked a new phase of increased reliability and control for women over their own reproductive outcomes.

By the 1990s, the earliest point of the period studied in this paper, numerous additional types of increasingly effective prescription sources of birth control were available. Throughout the paper, I primarily focus on three categories of prescription contraceptives:

- **Short-acting hormonal contraceptives.** These includes various formulations of "the pill", which must be taken daily; "the patch", which must be reapplied weekly; and "the ring", which must be replaced monthly. In many states, pharmacists are authorized to prescribe short-acting hormonal contraceptives.
- **Depot medroxyprogesterone acetate**, colloquially known as an injectable contraceptive (and hereafter referred to as DMPA). This contraceptive must be given as a shot by a medical professional and provides protection for 12 weeks.
- **Long-acting reversible contraceptives**, also known as LARCs. These include the copper intra-uterine device, the hormonal intra-uterine devices, and the hormonal birth control implant. They provide highly effective protection that can last for several years, but must be inserted (and subsequently removed) by a specialist at a medical visit.

These birth control methods differ substantially in their effectiveness in preventing pregnancy, how easy they are to access, and how costly they are. For example, birth control pills are often estimated to have typical usage failure rates of between 5-9% over a year compared to less than 1% for long-acting reversible contraceptives. At the same time, birth control pills can often be purchased out-of-pocket for 15−20 for a month's package, while the out-of-pocket cost for long-acting reversible contraceptives can be hundreds of dollars. The most effective methods like DMPA and LARCs thus have the highest up-front costs, both in terms of direct expense and the necessity to visit a medical professional.

While long-acting reversible contraceptives have been around since the 1990s, they were long only recommended for use by women who had given birth. In 2009, the American College of Obstetricians and Gynecologists began recommending that they should be offered both to nulliparous women and adolescents as well (**?**), a decision that has been identified as leading to increased prescription and use of long-acting reversible contraceptives (Buckles et al., 2022).

3 Institutional Background

To understand how the expansion of public insurance might affect take-up of contraceptives and fertility, it's necessary first to understand what options women have for accessing and paying for prescription contraceptives and how related policies have changed over time. This section describes how women can receive subsidized contraception either via insurance coverage or publicly supported clinics, and how policies for public insurance coverage have changed in recent decades.

3.1 Access to subsidized prescription contraceptives in the United States

Women can access prescription contraceptives in the United States in two primary ways: either by visiting a doctor or by going to a publicly supported clinic. To make a visit to a doctor, individuals generally need to have insurance coverage or pay out of pocket, which is often expensive. Many women who do not have insurance coverage or who do not have a usual source of primary care instead seek sexual and reproductive health care at publicly supported clinics. Close to half of uninsured individuals and about a third of individuals covered by Medicaid (the U.S. public health insurance program for low-income individuals) seek contraceptive care at publicly supported clinics (**?**).

Publicly supported clinics generally collect fees for services on a sliding scale, providing entirely free services for many low income individuals. Many publicly supported clinics are funded in part by the federal Title X program, a grant program that provides support for family planning services via existing organizations.

Public policies to fund subsidized contraception can therefore come in two primary forms: those which extend individual insurance coverage or those which increase funding to publicly supported clinics. The focus of this paper is on studying the former, particularly the expansion of public health insurance for low-income individuals. There are two important and substantive differences in the practical implications of funding one or the other worth noting.

First, even if an individual qualifies for free services at a publicly supported clinics, all approved methods of contraception may not be available to them for free. Clinics have finite resources, and as such must ration certain types of care. For example, only about 60% of Title X clinics provide long-acting reversible contraceptives, since they are the most expensive type of contraceptive (Kelly et al., 2020). In contrast, under public insurance programs, all FDA-approved methods are fully reimbursed to patients.

Second, funding of clinics can be used for other purposes than the direct provision of services, such as outreach programs and staff training. This benefit of funding clinics can have potentially have effects beyond what can be achieved by simply making all types of contraceptives free for individuals through insurance coverage, which is the primary effect of extending insurance coverage.

3.2 Public health insurance coverage for prescription contraceptives

Low-income individuals can gain insurance coverage for prescription contraception in the U.S. via two types of programs: one is via eligibility for full coverage under Medicaid, the public health insurance program. The second is through specific, targeted expansions to state Medicaid programs to cover family planning services, including contraception. These programs are known as family planning waiver programs, as states have to make waiver requests for expansion to the federal government.

Prior to 2010, states provided full Medicaid coverage only to limited low-income groups that fulfilled additional criteria for disadvantage, including children, pregnant women, individuals with disabilities, and parents, although the eligibility thresholds particularly for parents varied widely across states. However, between 1997 and 2009, 20 added family planning waivers based on income (there are related programs also often referred to as family planning waiver programs that offer "duration waivers," whereby postpartum women losing full Medicaid coverage under pregnancy eligibility thresholds can continue to be eligible for family planning services). The income thresholds for family planning waiver programs were largely similar across states. All but three of 20 chose eligibility thresholds for income of either 185% or 200% of the Federal Poverty Line: Virginia and Alabama set theirs at 133% and Vermont at 300%.

In 2010, the Patient Protection and Affordable Care Act (hereafter ACA) was passed. This law legislated an expansion of full Medicaid to 138% of the federal poverty line, which was intended to be implemented in 2014, although states were also allowed the option to expand early. Following litigation, the Supreme Court determined that expansion was optional for states. In order to implement the expansion, states had to pass legislation to prepare for it, and their governors had to approve it. Since support for Medicaid expansion was highly partisan, the states that chose to expand initially in 2014 were more likely to be those that had Democrat-controlled legislatures at the time the ACA passed in 2010. Over a dozen states have since implemented Medicaid expansion after 2014, including in Republican-leaning states where state legislatures and governors have at times been overridden by direct ballot measures.

With respect to coverage for reproductive care under Medicaid, the ACA Medicaid expansion extended access to fully subsidized contraception, but did not affect coverage of other reproductive health care for women. Eligibility thresholds for pregnant women were already more generous across all states than the new ACA thresholds, so the expansion did not affect the direct costs of pregnancy or birth. Likewise, in the states that supplement their Medicaid programs to cover abortion (federal funds cannot be used for this service), eligibility for Medicaid for accessing abortion goes under pregnancy eligibility rules. No states cover infertility treatments under Medicaid for the purpose of becoming pregnant, so this aspect is unaffected by Medicaid expansion.

With respect to the analysis, I split the study period into before and after 2010 to accomodate separate analysis of family planning waiver programs and the ACA Medicaid expansions. The vast majority of the family planning waiver program expansions occurred prior to the passage of the Affordable Care Act in 2010 (22 prior; 9 after). Moreover, since nearly all of these programs exhibited a shift from a zero eligibility threshold to a threshold of between approximately 185%-200% of the Federal Poverty Line, these expansions can reasonably be characterized as a binary treatment. Admittedly, there were varying levels of parental eligibility for full Medicaid during this period, which would also allow for coverage of contraceptives and may even reduce the anticipated cost of becoming a parent, but since most states did not change parental eligbility levels between 2000 and 2009, the role of parental coverage can be assessed both by explicit controls and excluding the states that changed eligibility from the sample.

The ACA Medicaid expansions present a less straightforward setting. First of all, about half of states at this point had family planning waiver programs, meaning that they were effectively "already treated" for coverage of prescription contraceptives for low-income individuals. These states will be excluded in baseline estimates.

Secondly, states that expanded according to the new law raised eligibility levels for both parents and non-parents to 138% of the Federal Poverty Line. Non-parents were previously entirely ineligible, however parents had varying levels of prior eligibility across states. As a result, expansion states with differing levels of prior parental eligibility in fact experienced different intensities of treatment, with some states having substantially larger or smaller fractions of the population that became newly eligible. This setting raises the question of whether the treatment can truly be treated as binary and whether non-expansion states can serve as appropriate controls, although in the literature on the ACA Medicaid expansion this is common practice.

In principle, a primary concern would be that estimates would be biased towards zero, as a binary

Table 1: Timing of policy implementation by state

	Family Planning Waiver Program			**ACA Medicaid Expansion**	
State	Year	Quarter	Threshold	Year	Quarter
Alaska				2016	3
Alabama	2000	4	133		
Arkansas	1997	3	200	2014	1
Arizona				2014	1
California	1997	1	200	2014	1
Colorado				2014	1
Connecticut[2]	2012	1	250	2014	1
District of Columbia[1]				2014	1
Delaware[1]				2014	1
Florida					
Georgia[2]	2011	4	200		
Hawaii				2014	1
Iowa	2006	1	300	2014	1
Idaho				2020	1
Illinois	2007		200	2014	1
Indiana[2]	2012	4	133	2015	
Kansas					
Kentucky				2014	1
Louisiana	2006	3	200	2017	3
Massachusetts[1]				2014	1
Maryland[2]	2012	1	200	2014	1
Maine[2,3]	2016	3	209	2019	4
Michigan	2006	3	185	2014	2
Minnesota	2006	3	200	2014	1
Missouri	2008	3	185		
Mississippi	2003	4	185		
Montana[2]	2012	2	211	2016	
North Carolina	2005	4	185		
North Dakota				2014	1
Nebraska				2020	4
New Hampshire[2]	2013	3	185	2015	3
New Jersey[3]	2019	4	200	2014	1
New Mexico	1998	3	185	2014	1
Nevada				2014	1
New York[1]	2002	4	200	2014	1
Ohio[2]	2012	1		2014	1
Oklahoma[3]	2005	2	185	2021	3
Oregon	1999	1	185	2014	1
Pennsylvania	2007	2	185	2015	
Rhode Island				2014	1
South Carolina	1997	3	185		
South Dakota					
Tennessee					
Texas	2007	1	185		
Utah[3]				2020	1
Virginia[3]	2007	1	133	2019	1
Vermont[1]				2014	1
Washington	2001	3	200	2014	1
Wisconsin	2003	1	185		
West Virginia				2014	1
Wyoming					

The table lists the year and quarter of implementation of income-based family planning waiver programs and ACA Medicaid expansion, as well as the eligibility threshold for the family planning waiver program relative to the Federal Poverty Line. See online appendix for a complete accounting of sources. Additional notes: (1) States that implemented partial early expansions under the ACA; these are excluded from the ACA analysis in accordance with Miller et al. (2021). (2) States that implemented family planning waivers after 2009 are considered as never-treated for the main family planning waiver analysis, and excluded from the main ACA analysis due to being already treated prior to ACA expansion. (3) States implementing either policy during or after 2019 are considered to be never-treated for fertility analyses, since these would impact only births in 2020 and later, whereas our data covers only births through 2019.

categorization implies that newly treated states were previously untreated, and never-treated states remain so, while both may in fact have partial treatment via parental eligibility. In the analysis of fertility, we can partially assess this concern by distinguishing estimates of effects on parents compared to non-parents, the latter for whom the binary treatment categorization is in fact perfectly appropriate.

4 Existing research on effects of subsidizing contraceptives

Paralleling the institutional setting where subsidies for contraceptives can be provided either by either extending individual insurance coverage or increasing funding for clinics, research on the effects of subsidized contraceptives can be divided along the same lines. Work on funding via clinics or institutions shows that these interventions primarily affect abortion and birth outcomes for teens Lindo and Packham (2017); Lindo et al. (2020); Kelly et al. (2020); Buckles et al. (2022). The limited evidence on the effects of extending public health insurance coverage likewise suggest reductions in fertility primarily among teens, although they find smaller reductions in fertility for older adult women as well (Kearney and Levine, 2009; Lindrooth and McCullough, 2007).

4.1 Subsidized contraceptives via clinics and institutions

Subsidized contraceptives via publicly supported clinics have been around since the 1960s, when the Title X program was first implemented to provide expanded and reduced-cost access to family planning services. Bailey (2012) found that these programs cause delays in the age at first childbirth and reductions in completed fertility, particularly for low-income women. This time period coincided with a period when birth control was first becoming legally available in many states, particularly to non-married women, as well as when legal access to abortion began rolling out, and as such the effects on fertility may not owe solely to the reductions in financial barriers to contraceptives (Myers, 2017).

More recently, Colorado implemented a program providing a large injection of funding to its Title X clinics specifically designated to increase provision of free, long-acting reversible contraceptives and for training its staff to provide improved counseling and care for this type of contraceptive. Lindo and Packham (2017) show that the program led to a reduction in teen births of about 7% for counties that had a Title X clinic, compared to those without. Kelly et al. (2020) further show that the program had stronger effects on teen births by geographic proximity, and even find some evidence of reductions in births to women in their twenties between six to eight years after the initial implementation of the program. Notably, these researchers demonstrate that the Title X clinics, most of which previously did not provide long-acting reversible contraceptives, led to a clear substitution among their clients from oral contraceptives to long-acting reversible contraceptives.

This latter point is important, because in order for subsidization of contraceptives to lead to a reduction in unintended births, a new method of contraception must supplant a previously less effective one. Buckles and Hungerman (2018) highlight an example where provided free contraceptives may not have fulfilled this condition: a program to distribute condoms to teenagers resulted in increased teen births when the program was not paired with counseling, whereas there were either no effects or negative effects on births when condom distribution was combined with mandated counseling. In effect, condom distribution may have encouraged greater sexual activity among teens who would not otherwise have been sexually active.

Cuts to funding for clinics that provide contraceptive services have also demonstrated their importance for contraceptive usage. When Texas substantially reduced funding for family planning services in 2011, leading to many clinic closures and the implementation of fees at clinics that had previously provided services for free. **?** shows that these cuts led to a decrease in takeup of contraceptives by clients at the clinics among publicly funded clinics, and ultimately to an increase in teen births of 3%, even in counties where the closures were of clinics that did not previously provide abortion, suggesting the effect was attributable to increased financial cost of contraceptives specifically.

In sum, while there is clearly evidence that funding for publicly supported clinics matters for contraceptive takeup and fertility, the evidence overwhelmingly highlights that it affects teen births specifically, rather than for older women. Although teen births presumably are the most costly in terms of disruption to schooling and careers, teen births account for only about one third of all unintended births (**?**).

Moreover, since funding for public clinics often goes hand in hand with support for counseling, it's difficult to distinguish whether the subsequent effects owe primarily to a reduction in the cost of contraceptives or whether the effect is partly driven by improved counseling towards effective methods.

4.2 Subsidized contraceptives via insurance coverage

Evidence on how subsidization of contraceptives for insurance affect contraceptive usage and fertility come from both changes to private and public insurance programs. Overall, this area of research provides promising evidence that changes in financial cost can induce increases in usage of effective contraceptives and may have effects on abortion or fertility for some groups.

4.2.1 Private insurance

Studies of state-level regulations to require private insurance companies to cover prescription contraceptives (although not necessarily without additional co-pays) find that such coverage generally led to increases in claims for prescription contraceptives (**??**), although the data sources do not allow for distinguishing effects on different types of contraceptives. **?** also finds evidence of a 3% decrease in the abortion rate, although no significant effect on the birth rate.

In addition to Medicaid expansions, one component of the Affordable Care Act passed in 2010 was to require all private insurers to cover prescription contraceptives without a copay. **?** and **?**, each using claims data from private (unidentified) insurers, find increases in overall prescription contraceptive claims of about 2 to 3 percentage points, with a particular increase in long-term contraceptives. **?** finds an increase primarily in injectable contraceptives using national claims data and wholesale data on contraceptives.

4.2.2 Public insurance

Whether or not low-income individuals should be more or less affected by expansions of coverage through insurance than individuals with private insurance is ambiguous: on the one hand, with less income, financial constraints may be more pressing. On the other hand, they are likely already eligible for subsidized contraception through publicly supported clinics, whereas those with private insurance (who typically have insurance through an employer) are less likely to have this alternative for subsidized contraception. The evidence on whether subsidization of contraceptives through public health insurance coverage leads to reduced fertility is mixed.

Two studies consider the effects of prior family planning waiver programs on fertility, focusing on the period of expansions between 1997 and 2002 (Kearney and Levine, 2009; Lindrooth and McCullough, 2007). Both estimate overall reductions in fertility of about 1.5-2%, while Kearney and Levine (2009) also estimate larger reductions in fertility for teens[1]. The authors also find suggestive evidence of a reduction in the proportion of women having sexual intercourse without any form of contraception, which they attribute to the family planning waiver programs.

Finally, **?** studies the expansion of both Medicaid coverage and non-Medicaid subsidized coverage for women ages 20-44 between the years 2011 to 2016, finding no effects on the birth rate of Medicaid coverage for this age group. One difference with the present study is that the author does not distinguish states that previously had family planning waiver programs in place.

5 Identification

In this section, I describe the estimands of interest, the estimators I will use, and how the estimators differ in terms of identification assumptions. I choose to employ and compare multiple estimators because they each have pros and cons in terms of how restrictive the assumptions they make are, as well as their efficiency and susceptibility to bias from violation of the central identification assumption for difference-in-differences, which is the assumption of parallel counterfactual trends. In the following section, I describe how the estimators differ in terms of estimation and inference.

I focus on two treatment estimands of interest. I am interested in a policy-relevant estimate to inform whether future expansions of coverage for subsidized contraceptives are likely to have an impact on take-up of highly effective contraceptives or on fertility. Thus, I will estimate the average treatment effect on the treated states, averaged over all periods of treatment. In potential outcomes notation, this estimand can be characterized as:

$$ATT = E[Y(1)_{st} - Y(0)_{st} | D_{st} = 1] \tag{1}$$

Here, $Y(d)_{st}$ gives an outcome of interest Y for potential treatment d in state s at time t, given the states and periods that are actually treated (indicated by $D_{st} = 1$).

Secondly, I am interested in considering whether treatment effects vary dynamically over time. As such, I will also aim to identify average treatment effects on the treated states by time since treatment:

$$ATT_k = E[Y(1)_{st} - Y(0)_{st} | D_{st} = 1, t = g_s + k] \tag{2}$$

Here, the difference with the first equation is that g_s indicates the time period of initial treatment for state s, and k indicates the time periods elapsed since treatment.

The challenge for identification of these estimands is to identify the counterfactual outcome $Y(0)_{st}$.

[1] Appendix B replicates and discusses the methodological choices of this paper to provide insights in differences in results found in their paper and the present one. Specifically, while the authors use population-weighted estimation, I argue that unweighted estimation yields a more relevant causal estimand for extrapolating future effects of policy expansions. With unweighted estimation, there are no significant effects of the family planning waiver program on births.

Under a difference-in-differences model, we will generally make two basic assumptions to achieve identification of this counterfactual:

1. **No anticipation.** This assumption indicates that units that will experience treatment do not respond prior to the first period of treatment.
2. **Parallel counterfactual trends.** This assumption characterizes that, in the absence of treatment, treated and untreated units would have followed parallel trends with respect to the outcome.

Next, I briefly discuss how the identification assumptions differ for the three estimators I will use for identification of the estimands of interest, in the current setting of staggered treatment timing: the two-way fixed effects estimator, the imputation estimator (Gardner, 2021; Borusyak et al., 2021) and the outcome regression estimator from **?**.

5.1 Two-way fixed effects

In general, the two-way fixed effects (TWFE) estimator will only consistently identify the ATT with two additional assumptions:

1. Constant treatment effects across time
2. Constant treatment effects across units

These properties are among the recent findings of work on developing improved difference-in-differences estimators (Goodman-Bacon, 2021; Sun and Abraham, 2020; Borusyak et al., 2021; Callaway and Sant'Anna, 2020; De Chaisemartin and d'Haultfoeuille, 2020; Gardner, 2021). These assumptions are necessary because in the presence of staggered treatment, the parameter estimated when employing TWFE effectively places unequal weights (which are affected arbitrarily by the sample size and fraction of time a unit is treated, among other things) on different treated units, including in some settings negative weights. Thus, given these arbitrary weights, the ATT is only identified when all treated units at all time periods experience the same treatment effect.

Note that the first assumption immediately rules out the possibility of dynamic effects, making the question of estimating ATT_k moot. The second assumption rules out any possible heterogeneity in treatment effects. Except in the perhaps trivial case where there is no true treatment effect, in many setting we might expect that these assumptions could be violated. If they are, then the TWFE estimator does not identify the ATT. Based on this insight, econometricians have worked to develop other estimators that place suitable, non-arbitrary weights on treated units to be able to identify the ATT or other treatment effects of interest, while still allowing for treatment effect heterogeneity over time and across units.

5.2 Imputation estimator

Although in principle the idea of difference-in-differences is predicated only on parallel counterfactual trends – that is, that trends *after treatment* would have been parallel among treated and untreated units, had the treated units not been treated – in practice we often conflate this assumption with the assumption that there would be parallel pre-trends *prior to treatment*. Since parallel counterfactual trends is an inherently untestable assumption, we often instead examine whether there are violations of parallel pre-trends to provide motivation for the untestable counterfactual assumption. The imputation estimator achieves identification by making this assumption explicit: treated and untreated units are assumed to follow parallel trends prior to treatment, as well as assuming parallel counterfactual trends. Concretely, Borusyak et al. (2021) characterize this assumption in terms of potential outcomes as follows:

$$Y_{st} = \alpha_s + \gamma_t + \tau_{st}$$

where Y_{st} represents a potential outcome for unit s in time t, α_s and γ_t are unit and period fixed effects, and τ_{st} is a treatment effect for time t, which is zero (under no anticipation) if the unit is never treated or not yet treated. It's straightforward to see that this assumed structure of outcomes imposes parallel trends among treated and untreated units, since α_s is presumed to be constant before and after treatment, and γ_t is a common time trend across the whole sample. Moreover, as long as α_s and γ_t can be identified (for which there are simple data requirements: at least two observations for all units, including two pre-treatment observations for treated units, and at least one untreated unit in all periods), the treatment effect τ_{st} can also be identified. From the individual treatment effects $\tau_s t$, the ATT can then be constructed by assigning the appropriate weights v_{st}, where $v_{st} = 1/N$ when N represents the total number of treated unit-time observations.

$$ATT = \sum v_{st}\tau_{st}$$

Likewise, the ATT_k parameters can be constructed by assigning equal weights on treated observations' treatment effects within each time period relative to treatment. Importantly, this identification strategy for the ATT places no restrictions on heterogeneity in treatment effects over time or across units.

5.3 Callaway and Sant'Anna's outcome regression estimator

Identification can still be achieved even without relying on the assumption of parallel pre-trends, and this is a key feature of Callaway and Sant'Anna's difference-in-differences estimator. To do so, they instead make a more specific assumption about the evolution of counterfactual trends at the cohort level, where cohorts are defined by the period in which they first were treated (or if they were never treated).

$$E[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = E[Y_t(0) - Y_{t-1}(0)|X, C = 1]$$

In this equation, $Y_t(0)$ indicates the potential outcome under no treatment in time t, G_g indicates units that are first treated in time g, and $C = 1$ indicates never-treated units, while X are pre-treatment covariates. Note that this assumption does not impose anything about pre-treatment periods (although there are variations on this assumption which may do so, even for their estimator), but rather expresses the parallel trends assumption in terms of calendar time.

Whereas the imputation estimator builds upon individual treatment effects, the CS estimator uses what they call the group-time average treatment effect as the core building block. Assuming we do not condition on covariates, from this parallel trends assumption the average group-time treatment effect on the treated can be identified with an intuitively similar structure to the canonical two-period, two-group difference-in-difference setup:

$$ATT_{g,t} = E[Y_t - Y_{g-1}|G_g = 1] - E[Y_t - Y_{g-1}|C = 1]$$

Note that because the parallel trends assumption is expressed in terms of pairs of calendar time periods, when constructing the $ATT_{g,t}$ both the treatment group and the control group are compared to the calendar period prior to the treated group's first period of treatment. As long as the parallel trends assumption as described is fulfilled, what happens prior to one period prior to treatment has no import for identification.

From the $ATT_{g,t}$ Callaway and Sant'Anna propose aggregations that are comparable to the ATT and ATT_k. Specifically, we can construct a weighted average of all $ATT_{g,t}$ for all groups and periods, where the weights are in proportion to group size. The resulting estimand is conceptually similar to the ATT by placing equal weights on treatment effects for each of the treated units. The approach to ATT_k is the same, but for relative time periods.

Although this estimator is less restrictive in terms of not imposing parallel pre-trends, it does restrict heterogeneity in treatment effects, since the lowest "building block" of treatment effects is at the cohort level.

6 Estimation

In the previous section, I described the assumptions necessary for each of the estimators to identify the ATT. In this section, I discuss estimation for each of these estimators of the ATT, as well as the implications for efficiency and bias from the different estimation approaches. In short, the imputation estimators are more efficient, but also more susceptible to bias owing to undetected violation of this assumption because they rely on the pre-treatment parallel trends assumption for identification and estimation, whereas the Callaway and Sant'Anna estimator is less prone to bias but also less efficient precisely because it does not use the pre-treatment observations for identification or estimation. Given these tradeoffs, it is desirable to consider both and compare the resulting estimates.

6.1 Two-way fixed effects

For TWFE, the estimation is straightforward, consisting of a single linear regression:

$$y_{st} = \beta(D_{st}) + \gamma_s + \delta_t + \epsilon_{st} \tag{3}$$

The subscripts s and t denote state and time periods, and D is an indicator for whether a state has been treated in state s and time t, and thus β_1 reflects the coefficient of interest (although it may not yield an appropriate estimate of the estimand of interest). Standard errors are clustered at the state level. While I initially present results without additional control variables, given the potential issues with including time-varying control variables raised by Sant'Anna and Zhao (2020), an additional specification includes controls, which are described in the contraception and birth implementation sections respectively.

When the identification assumptions are in fact fulfilled, TWFE is the most efficient estimator for the ATT.

6.2 Imputation estimator

In practice, I will consider two variations of the imputation estimator: that of Gardner (2021) and that of Borusyak et al. (2021), who I will refer to as BJS. While the identification assumptions are the same, they differ in terms of implementation, particularly with respect to inference.

Gardner's imputation estimator consists of two equations. First, we regress the outcomes on state and time fixed effects, using only observations that are never-treated or not-yet treated:

$$y_{st} = \gamma_s + \delta_t + \psi_{st} \tag{4}$$

Note that to conduct this estimation, each treated unit must have at least two pre-treatment observations to be included (in order to estimate γ_s). Where additional control variables are included in the specification (the same ones as described above) they are added only in this first step.

Next, we calculate the residuals $\tilde{psi}_{st}$ based on the linear predictions of the first regression, including extrapolating to the treated observations that were not in the sample for the first regression.

In the second step, we regress these residuals on our treatment variables of interest:

$$\tilde{psi}_{st} = \beta(D_{st}) + \epsilon_{st} \tag{5}$$

In practice, these two equations are estimated using a two-stage GMM estimator, to correct standard errors for the fact that the outcome variable is predicted (see Gardner (2021) and Butts and Gardner (2021) for details.) These adjusted standard errors are also clustered at the state level. These procedures are implemented in the package `did2s` (Butts, 2021) and used with Stata 17.

To instead estimate an event-study specification using Gardner's approach, D_{st} can be replaced by a series of relative time indicators D^k_{st} in the second equation.

The first step of the imputation estimator from Borusyak, Jaravel, and Hull (BJS) is identical to that of Gardne's, to generate residuals. By considering $\tilde{psi}_{st}$ for treated observations to be an estimate of the individual treatment effect τ_{st}, they suggest making an explicit choice of weights on treated observations for each estimand of interest to form the average (or even cumulative) treatment effect of interest. Concretely, the estimands of interest are estimated as a weighted sum, where Ω indicates the sample:

$$\hat{\beta_w} = \sum_{st \in \Omega} v_{st}\tau_{st} \tag{6}$$

In the case where weights are equal for all treated observations across all periods, the estimator coincides numerically with Gardner's where a simple indicator variable for treatment is used. Likewise, they will be numerically identical if we wish to estimate horizon-specific treatment effects and weights for treated observations observed at a given horizon are equal.

However, while the estimator itself is numerically equivalent between Gardner and BJS under the assumption of equal weights for all treated units, the latter develops a method for inference that is more efficient than Gardner's (which is based on a simple adjustment for the two-stage estimation process via General Method of Moments), which requires the additional condition that the treatment parameter to be estimated is based on "many" treated units rather than just a few. While this condition is often fulfilled for a typical treatment parameter where the ATT is estimated for all periods, it may not be fulfilled when the treatment parameters to be estimated are based on smaller groups, such as when states comprise the observation units and we wish to estimate ATTs by time since treatment. Morever, in terms of implementation, including control variables often makes it impossible for the standard errors to converge under this process. Thus, although the point estimates without control variables coincide, it can still be informative to use both estimators.

Concretely, BJS analytically derive the variance of the estimator with clustering, which is given as:

$$\omega_w^2 = \mathbb{E}[\sum_s (\sum_t v_{st}\psi_s t)^2]$$

In practice, to implement this variance estimator, a choice must be made regarding the construction of a sample analogue for ϵ_{st}, as without specific assumptions restricting treatment heterogeneity, it is not possible to distinguish between τ_{st} and ϵ_{st}. The solution BJS propose involves replacing ϵ_{st} with $\tilde{\psi}_{st} = \hat{\tau}_{st} - \hat{\bar{\tau}}_{st}$ where $\hat{\bar{\tau}}_{st}$ is an average of individual treatment effects for the parameter of interest. It is for this reason that they caution it may be problematic if there are too few treated units included in the sample for estimating a particular parameter. As a result, I only use the BJS estimator for ATTs across all periods, and not for event-study estimates.

6.3 Callaway and Sant'Anna estimator

The last estimator I use is Callaway and Sant'anna's outcome regression estimator.[2]

Estimation essentially proceeds in two steps for this estimator when covariates are not included. First, we estimate ATTs for each group-period cohort (where groups are identified by their timing of treatment, not as individual states), by estimating regressions of the following form for each period and treatment cohort g, compared to the never-treated control group C, using only observations from the period prior to treatment and time t, where G_g is an indicator for whether an observation belongs to group g:

$$y_{gt} = \alpha_1^{gt} + \alpha_2^{gt} * G_g + \alpha_3^{gt} * \mathbb{1}\{T = t\} + \beta^{gt} * (G_g \times \mathbb{1}\{T = t\}) + \epsilon^{gt} \tag{7}$$

Next, the ATT_{gt} parameters, given by β^{gt} in this regression, are aggregated to form either an overall ATT across all periods or event-study-style ATTs by time relative to treatment through specific choices of weights.

To generate an estimate of the ATT across all periods, we construct a weighted average across all groups with weights determined by the relative size of groups.

To generate aggregations that reflect ATTs for a given time relative to treatment, we average the group-time ATTs for groups that have been treated for a given number of time periods, with weights again determined by relative group sizes.

The difference in whether information from pre-treatment periods is incorporated into the estimation of the treatment effects themselves is significant for two reasons. First, it improves efficiency, a fact directly demonstrated in Borusyak, Jaravel, and Spiess. At the same time, Roth and Sant'Anna (2020)

[2] Callaway and Sant'anna also develop a doubly robust estimator which builds simultaneously on both outcome regression and inverse probability weighting. However, without variation in the propensity score for treatment owing to differences in initial covariates, there is no difference between the outcome regression and doubly robust estimators.

has demonstrated that our tests for violations of parallel pre-trends are often underpowered and may fail to detect small but meaningful deviations from parallel pre-trends and can in fact lead to biased estimates of treatment effects. The Callaway and Sant'anna approach is less susceptible to this bias, since neither identification nor estimation require parallel pre-trends, yet it will be less efficient since it incorporates less information.

7 Contraception

For an expansion in coverage of prescription contraceptives to have an impact on births, it must be that contraceptive usage increases. This first part of the analysis makes use of high-quality administrative data on reimbursement of contraceptives to analyze whether subsidization through public insurance increases takeup of prescription contraceptives, particularly of highly effective forms of birth control. The estimates from this analysis can be used to make predictions about the likely effects of these policies on fertility, via their effects on contraceptives.

7.1 Data

High-quality state-level data on usage is not readily available,[3] but we can gain insights by examining comprehensive data on Medicaid reimbursement of prescription contraceptives. Reimbursement of prescriptions provides an upper bound for the potential increase in contraceptive usage: usage cannot increase more than prescriptions themselves. Based on the effects on reimbursements for different types of contraceptives combined with existing evidence on the effectiveness of different types, we can infer what plausible effects on fertility are likely to be.

To study contraceptives, I use the State Drug Utilization Dataset (SDUD), years 1992-2019, which includes all outpatient drugs for which state Medicaid programs request reimbursement through the Medicaid Drug Rebate Program. To identify contraceptive drugs, I follow the listing of contraception drug codes given by Sumarsono et al. (2020),[4] who are physicians, to identify prescription contraceptives in the three categories described in Sec. 2:

In the SDUD, prescriptions are tallied quarterly. To standardize prescriptions across categories, I rescale prescription quantities in terms of months of contraceptive protection. For example, pills are sometimes prescribed in 90-pill packs, which represents three months of coverage. Similarly, a full month of coverage with a patch requires three patches, and therefore one patch represents 0.33 months of coverage. Each injection offers 3 months of coverage. Finally, while implants and IUDs can work for between three and ten years depending on the type, since we cannot know how long individuals use it, I make the conservative assumption that these contraceptives provide 12 months of coverage.

[3] The only source of data on contraceptive usage is the National Survey on Family Growth; however, its sampling strategy is not designed to be representative at the state level. This issue would be further exacerbated when excluding early expansion states and states that changed status with respect to a family planning waiver over the study period

[4] Sumarsono et al. (2020) have a similar aim of assessing how birth control prescriptions have changed following the ACA. There are some differences with my analysis: first, they examine the outcomes in levels and relative to the population of Medicaid enrollees–the latter of which is simultaneously increasing due to Medicaid expansion–whereas my outcome of interest is prescriptions relative to the whole population of women of child-bearing age. Second, they compare only 2014 expander states to never-expanders, while I include later expansion states, and exclude early expansion states as well as states that already had family planning waiver status during the ACA expansion period. Finally, I also assess the effects of earlier family planning waiver program policies on prescription reimbursements.

Next, I adjust these measures to be relative to the population of women of 15-44, namely women of child-bearing age. I use data from the U.S. Census intercensal estimates by state, age and sex as measures of this group's population. I construct the quantity of potential coverage months by multiplying the population of women in each state by three, since for each quarter each woman has a potential contraceptive coverage need of three months. Finally, I divide the number of months of coverage for each category of prescriptions by the number of potential contraceptive coverage months per quarter. This procedure gives the rate of contraceptive coverage per woman of childbearing age owing to prescriptions reimbursed by Medicaid. For easier comparison with fertility rates, which are often expressed per thousand woman of childbearing age, I multiply this rate by 1,000.

Table 2 provides summary statistics on reimbursement of prescription contraceptives. Examination of rates of coverage highlight that, in proportion to the whole population of women in the country, prescription contraceptives reimbursed through Medicaid account on average for a relatively small proportion of coverage for the country as a whole, perhaps for just 1 to 2% of all women who may need coverage.

Table 2: Summary statistics for quarterly Medicaid reimbursement of prescription contraceptives

Variable	Mean (1)	Standard deviation (2)
Short-acting hormonal, units	37240.57	(76271.60)
Depot medroxyprogesterone acetate, units	14153.56	(26181.67)
Long-acting reversible, units	8985.88	(24240.86)
Short-acting hormonal, rate	9.35	(8.87)
Depot medroxyprogesterone acetate, rate	3.98	(5.64)
Long-acting reversible, rate	2.58	(3.80)
Population of women aged 15-44	1223513.79	(1405557.48)
Observations	5722	

Data on prescription reimbursements is from the State Drug Utilization Dataset between 1992 to 2019. Units are aggregated by state, quarter, and type of contraceptive, where contraceptives are identified and categorized by National Drug Code according to Sumarsono et al. (2020). The population of women aged 15-44 is drawn from U.S. Census Bureau intercensal estimates by state and year. Rates indicate the rate of coverage per 1,000 women aged 15-44 for a given type, and are calculated by converting units to monthly coverage equivalents, dividing by the months of potential coverage need in a given quarter (three times the population of women aged 15-44), and multiplying by 1,000.

7.2 Implementation

As described in the prior section, estimation of a difference-in-differences model will employ four estimators: TWFE, Gardner's imputation estimator, Borusyak's imputation estimator, and Callaway and Sant'anna's outcome regression estimator.

When control variables are included, these include controls for the proportion of women in different age groups (15-19; 20-24; 25-29; 30-34; 35-39, with 40-44 excluded) and the fraction of white women and Hispanic women respectively.

Results are estimated separately for the effect of family planning waiver programs between 1992 and 2009, and for the ACA Medicaid expansions between 2010 and 2019.

7.3 Results

On average, both family planning waiver programs and the ACA Medicaid expansion increased reimsbursement of short-acting hormonal contraceptives, while family planning waiver programs appear to have had a larger impact on DMPA prescriptions, and the ACA Medicaid expansion marginally increased long-acting reversible contraceptives. Panel (a) of Fig. 1 presents the overall ATT for the three types of contraceptives across all periods for all treated states through 2009, while Panel (b) presents analogous estimates for the effects of ACA Medicaid expansion between 2010 and 2019.

Confidence intervals for the effects of ACA Medicaid expansion on short-acting hormonal contraceptives are larger than for the family planning waiver programs, ranging between 4 and nearly 13 additional women covered per 1000 women of childbearing age for the former and 2 to 6 for the latter. For DMPA, the point estimates for the effects of family planning waiver programs are approximately 3.5 additional women covered of coverage per 1000 women of childbearing age, while LARC estimates are centered around zero and small in magnitude. Finally, for the ACA Medicaid expansion, all estimates of the effects on DMPA are not significantly different from zero, while for long-acting reversible contraception, the confidence intervals range from barely more than zero to about 6 women covered per 1000 women of childbearing age.

Despite widespread concerns about TWFE and time-varying controls, the results across methods are strikingly similar, with TWFE estimates even appearing somewhat biased towards zero.

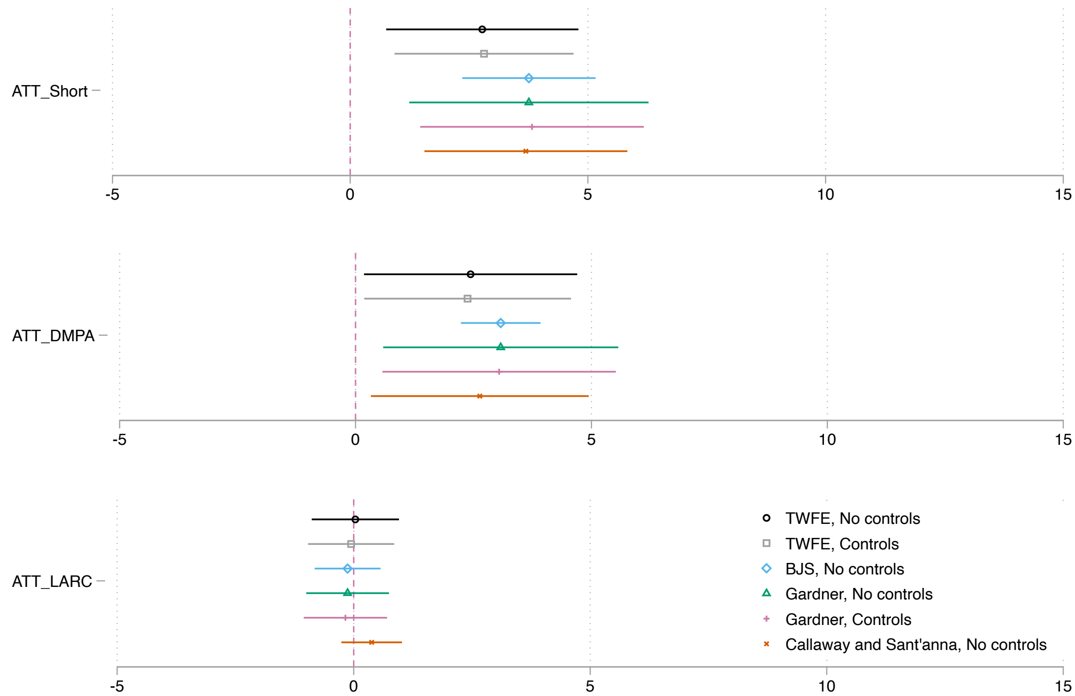
To assess both the likelihood of differential pretrends and whether the treatment effects exhibit dynamic patterns, Fig. 3 presents estimates for family planning waiver programs by contraceptive type, and quarters since treatment using the Gardner imputation method. For family planning waiver programs, point estimates for short-acting hormonal contraceptives increase over time, but so too do confidence intervals, such that it is not possible to distinguish whether treatment effects are arguably constant or increasing over time. For DMPA, pre-treatment coefficients using never- and not-yet treated observations only are very precisely estimated and centered around zero. Notably, although the overall treatment effect is positive and significant, estimates by horizon from four quarters after treatment and onward exhibit large standard errors and the periods between eighteen quarters (four and a half years) and onward exhibit a sudden increase in point estimates, suggesting that the positive overall effect may be driven primarily be few early adopting states. For long-acting reversible contraceptives, point estimates before and after treatment largely remain centered around zero, and the imprecision of post-treatment estimates increases only slightly relative to pre-treatment estimates, suggesting that there really was no average affect on LARC takeup for family planning waiver programs.

For the ACA Medicaid expansion, short-acting hormonal contraceptives exhibit a much clearer jump in reimbursements beginning two periods after expansion and remaining relatively constant in subsequent periods, suggesting constant treatment effects. In contrast, pre-treatment and post-treatment estimates for DMPA are largely overlapping, although the post-treatment coefficients are more imprecise. Likewise, although the overall treatment effect for LARCs is positive and significant, when comparing estimates by horizon there is again substantial overlap between post-treatment estimates and pre-treatment estimates, suggesting that the effect may be modes perhaps not robust to mild violations of the parallel trends assumption.

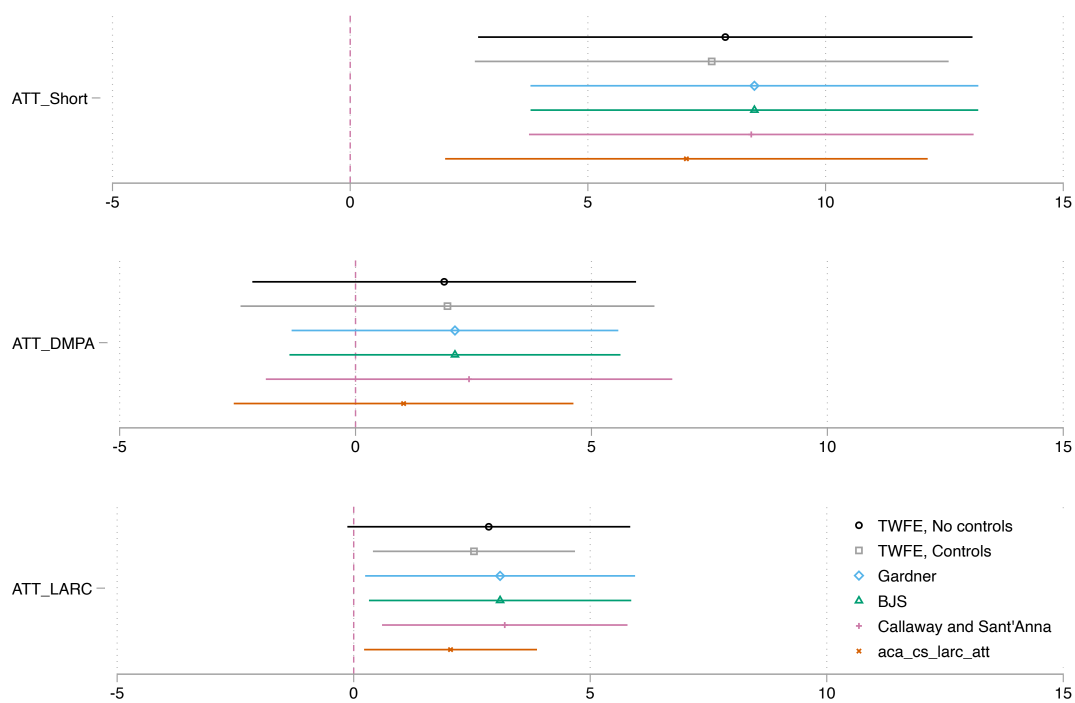
Altogether, these results provide evidence for an increase in reimbursement of contraceptives owing to both policies, particularly for short-acting hormonal contraceptives, and no indications to suggest that there are obvious violations of the parallel trends or no anticipation assumptions.

Figure 1: Overall ATTs for Reimbursement of Prescription Contraceptives

(a) Family Planning Waiver Programs, 1992-2009



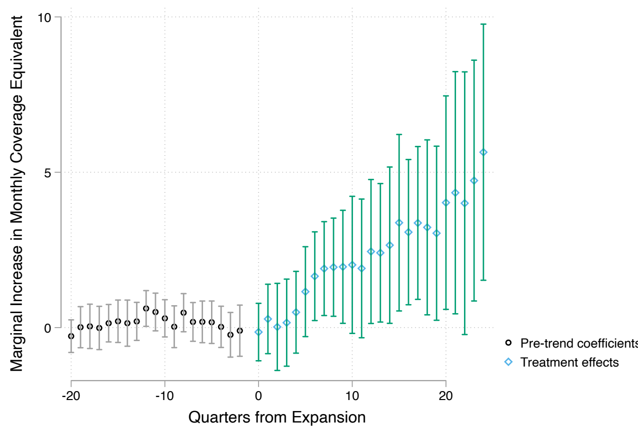
(b) ACA Medicaid Expansion, 2010-2019



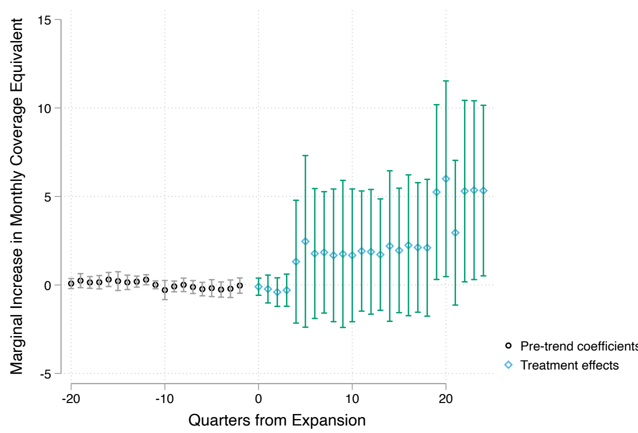
This figure presents overall average treatment effects on treated states using data from the State Drug Utilization Program, 1992-2019, for Family Planning Waiver Programs (Panel (a)) and ACA Medicaid expansions (Panel (b)). Coefficients are in terms of contraceptive coverage rates per 1,000 women aged 15-44, by type of contraceptive (short-acting hormonal; DMPA; or long-acting reversible). Legends indicate which estimator and control variables were used, following the description in Sec. 7.2.

Figure 2: Dynamic effects of family planning waiver expansions on Medicaid-reimbursed contraceptive prescriptions (1992-2009)

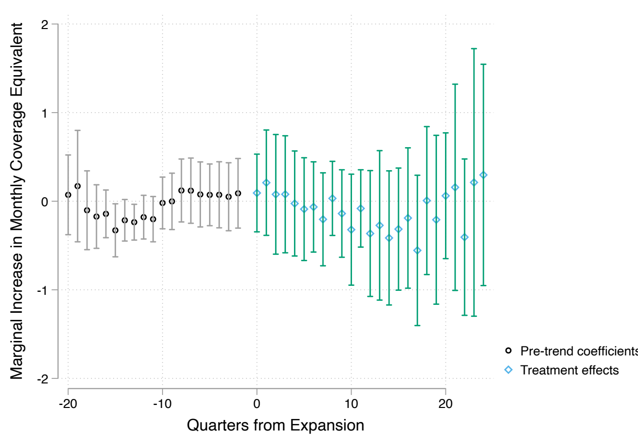
(a) Short-acting hormonal contraceptives



(b) DMPA Injections



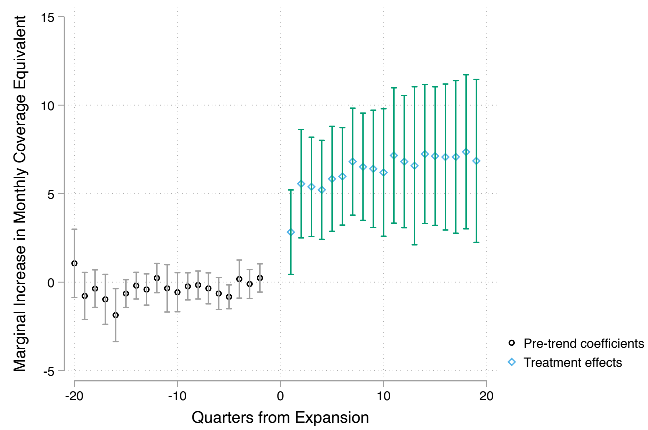
(c) Long-acting reversible contraceptives



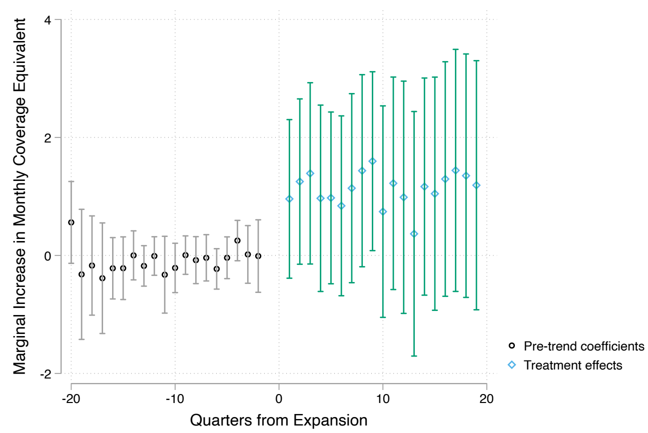
This figure presents average treatment effects on treated states by quarters since policy implementation using data from the State Drug Utilization Program, 1992-2009, for Family Planning Waiver Programs, by contraceptive type (short-acting hormonal; DMPA; or long-acting reversible). Coefficients are in terms of contraceptive coverage rates per 1,000 women aged 15-44. Treatment effects are estimated using the imputation estimator from Gardner (2021) without covariates.

Figure 3: Dynamic effects of ACA Medicaid expansion on Medicaid-reimbursed contraceptive prescriptions (2010-2020)

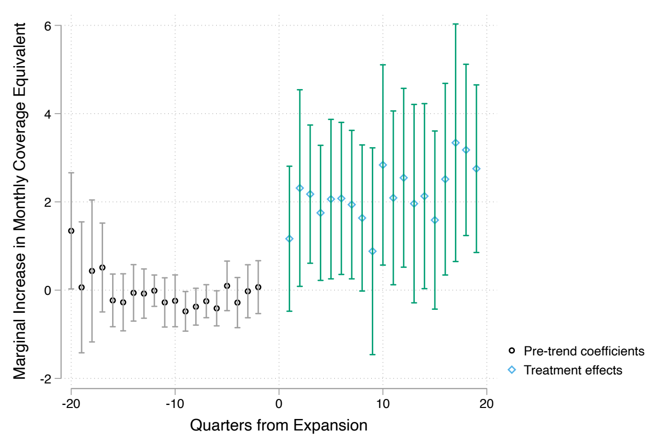
(a) Short-acting hormonal contraceptives



(b) DMPA Injections



(c) Long-acting reversible contraceptives



This figure presents overall average treatment effects on treated states by quarters since policy implementation using data from the State Drug Utilization Program, 2010-2020, for ACA Medicaid expansions, by contraceptive type (short-acting hormonal; DMPA; or long-acting reversible). Coefficients are in terms of contraceptive coverage rates per 1,000 women aged 15-44. Treatment effects are estimated using the imputation estimator from Gardner (2021) without covariates.

7.4 Discussion

The results for reimbursement of prescription contraceptives provide clear evidence that both types of programs increased takeup of contraceptives through Medicaid. They also provide a basis for predicting the plausible ranges of effects that these policies may in turn have on unintended fertility.

As previously discussed, the treatment effects on reimbursement of prescription contraceptives provide an upper bound for the effects on *usage* of prescription contraceptives. In turn, we can infer the predicted upper bound on the treatment effect for fertility by assuming that the treatment effects on reimbursements are equal to the treatment effect on usage, or in other words, that the women receiving prescriptions as a result of these programs did not previously use any contraceptives and were sexually active.[5] We can also consider a range of other plausible predictions for effects on fertility by varying our assumptions about the prior contraceptive usage rates of women who took up prescription contraceptives through Medicaid as a result of the two key policies.

Table 3: Estimated contraceptive failure rates

Method	Failure Rate
No contraceptive method	85%
Fertility-based awareness methods	24%
Withdrawal	22%
Male condom	18%
Pill, patch, ring	9%
Injectable (DMPA)	6%
LARCs (IUD/Implant)	0.8%

Estimated contraceptive failure rates from tyical first-year usage, according to Trussell (2004).

For these assumptions, I consider five scenarios about prior contraceptive usage of the women represented by the treatment effects on reimbursement. In all scenarios, I assume women are sexually active and use typical usage failure rates by method from Trussell (2004), which are rather more conservative in the estimation of the likelihood of failure than clinical studies (e.g. Mansour et al. (2010)), given that they are based on survey data for first-year usage of a given method. Table 3 summarizes the estimates of typical usage failure rates by contraceptive type.

With respect to their prior contraceptive usage, I assume:

- **Scenario 1**: Women were not using any form of contraception previously, which typically results in pregnancy for about 87% of women over a year.
- **Scenario 2**: Women were using only fertility-based awareness methods or withdrawal, resulting

[5] There is an implicit assumption made here that the effect on fertility will be weakly negative. There is an argument to be made that access to contraceptives can potentially increase pregnancies and fertility, as Myers (2017) found in the early stages of access to the pill in the United States. However this primarily matters only in a setting where access to contraception causes women who were previously sexually abstinent to become sexually active, since sexual abstinence fully prevents pregnancy and all contraceptive methods have some level of failure. In the setting of this paper, where contraception was already widely available and used and most adult women were sexually active, it is more plausible to assume these policies caused women to use a method of the same or greater effectiveness for preventing pregnancy. Under such a scenario, subsidization of contraception should either decrease fertility or have no effect, such as if all women taking up contraception under Medicaid already used that very method.

in pregnancy for about 22-24% of women over the course of a year.

- **Scenario 3**: Women whose prior contraceptive usage reflects the average distribution of methods for uninsured women surveyed in the National Survey of Family Growth, as given in Frost and Darroch (2008), and assuming typical usage failure rates for each method. The average expectation of pregnancy over a year is 16%.

- **Scenario 4**: Women were previously using the next-most effective method of contraceptives: women who took up short-acting hormonal contraceptives were previously using condoms; those taking up DMPA or LARCs were previously using short-acting hormonal contraceptives.

- **Scenario 5**: Women were previously using the exact same contraceptive method, only now paid for by Medicaid. Naturally, this would result in zero expected change in pregnancies.

The predicted reductions in fertility are estimated by calculating:

$$\widehat{F}_i = \sum_{j=1}^{3} \alpha_i (\mu^i_{j,New} - \mu_{j,Prior})$$

for $i = 1, 2, 3, 4, 5$ and $j = 1, 2, 3$, where i indicates the scenario for prior contraceptive assumptions, j indicates the contraception type (short-acting hormonal, DMPA, and long-acting reversible, respectively), α_j represents the treatment effect parameter associated with contraceptive type i, $\mu^i_{j,Prior}$ represents the prior estimated probability of pregnancy for a given scenario, and $\mu_{j,New}$ represents the probability of pregnancy for the new method. Finally, $\widehat{F}$ represents the predicted total treatment effect on the pregnancy rate for a given scenario and policy.

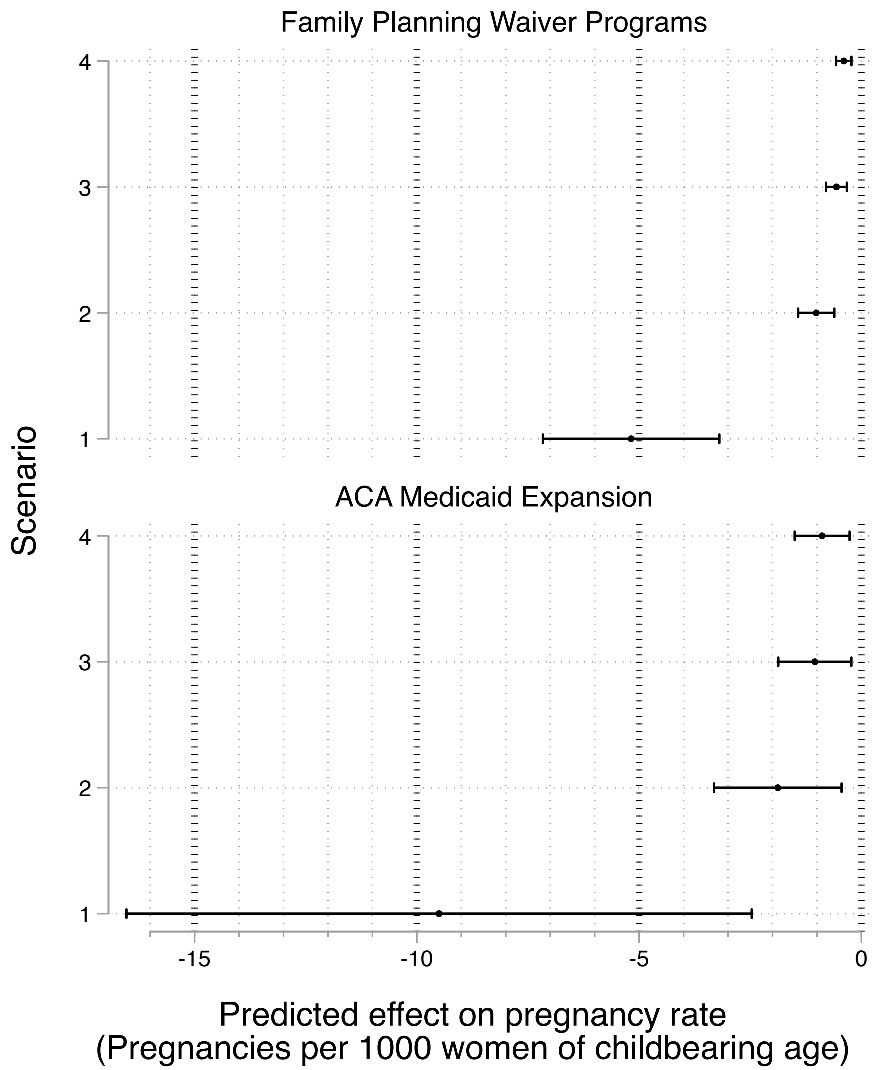
Fig. 4 displays the predicted reductions in fertility, based on estimates using the BJS estimator and summing across the predicted effects for different birth control types. For Scenario 1, which is arguably implausible but provides a useful upper bound, the range of predictions given the estimated effects on contraceptive reimbursements is wide: for family planning waiver programs, we would predict reductions between 3.2 to 7.2 pregnancies per thousand women aged 15-44. For the ACA Medicaid expansion, the implied prediction range is even broader, between 2.5 and 16.5 fewer pregnancies. To put this in perspective, the Guttmacher Institute estimates that in 2011 there were 98 pregnancies per 1000 women aged 15-44, of which slightly less than half were unintended. Moreover, Buckles et al. (2022) estimate that over the past twenty years, the rates of unintended birth range between about 20 to 25 per thousand women aged 15-44. The upper bounds of these predictions would thus represents enormous proportional cuts to the rates of unintended pregnancies or births.

The other three scenarios likely provide more plausible predictions, based on the reasonable assumption that most women that take up contraceptives under the Medicaid programs were making at least some efforts to avoid pregnancy previously. For these, the predicted reductions in pregnancies are much more modest. Under Scenario 2, the predicted reduction for the family planning waiver programs ranges from 0.6 to 1.4 fewer pregnancies per thousand women, while for the ACA Medicaid expansion it ranges from 0.4 to 3.3 fewer pregnancies. However, this scenario still assumes that no women previously used medical or barrier methods of contraceptives, which based on what we know of survey data on contraceptive usage is unlikely, even for uninsured women. Scenarios 3 and 4 thus present the most plausible predictions, based on the assumptions that most women did use medical or barrier methods, though perhaps less effective ones. For these scenarios, the predicted effects on pregnancies are quite small, with upper bounds of less than one pregnancy per thousand women for the family planning waiver programs under both scenarios, and less than two pregnancies per thousand women for the ACA Medicaid expansions. Scenario 5 is not graphed, since it of course predicts exactly zero change in pregnancies, but it is important nevertheless to keep in mind the possibility that subsidization functions only as a monetary transfer rather than an incentive to change behavior.

There are two final points to keep in mind when interpreting these predictions. First, the reduction in births is likely to be smaller in magnitude than the reduction in pregnancies. For unintended pregnancies, a substantial proportion (perhaps almost as many as half, according to the Guttmacher's estimates) are terminated. Thus, reductions in pregnancies owing to the use of more effective contraception is likely to only partially translate to reductions in births, since had they occurred, some of those pregnancies would in any case have been terminated.

Second, these predictions represent anticipated average treatment effects on the treated states at the population level, but we already know that in treated states, only between one fourth to one third of women will in fact be eligible for the programs. Within the subpopulations of eligible individuals, the predicted effects on pregnancies may thus be three to four times larger. Given that the predicted treatment effects at the population level seem likely to be quite small (1 birth per thousand women aged 15-44 represents approximately 1-2% of the fertility rate in a given year, with fertility rates ranging from 60-70 births per thousand women over the last two decades), this latter point suggests that focusing on subsamples of eligible individuals in the analysis of birth may be useful for improving power and precision in estimating effects.

Figure 4: Predicted effects on fertility based on estimated changes in contraceptive reimbursement



Predictions of plausible effects on fertility owing to increases in contraceptive reimbursement are calculated by combining estimated treatment effects from Fig. 1 for the BJS estimator, with estimated first-year typical usage failure rates by contraceptive method from Trussell (2004), as well as explicit assumptions about prior contraceptive usage.

8 Births

Given the preceeding analysis of the effects of Medicaid policies on contraceptive reimbursements, it is possible that there is little or no effect of the policies on fertility, but there are also plausible scenarios where there is an effect. Since only a minority of the population is treated, at the overall population level, the plausible average treatment effects on the treated states will most likely be modest, in the range of 1-3% reductions in births overall. However, if this type of average effect reflects changes in births within only the subset of newly eligible individuals, then the effect sizes for these subgroups may be three to four times larger. As a result, a key aim of the empirical strategy to study the effects of these policies on births will be focused on maximizing power in the analysis. To do so, I supplement aggregate natality data with individual data from the the American Community Survey, which allows for restricting the analysis to individuals that are directly exposed to treatment.

8.1 Data

The ideal dataset for analyzing the effect of policy changes on births would be to observe the same set of individuals and their birth outcomes over the years of the study period as well as their income to determine their eligibility status. Existing panel surveys are too small to be useful for this purpose; tax data are not comprehensive (and in particular do not include many low-income individuals who would be eligible for Medicaid but may not be required to file federal income taxes); and the U.S. Natality Detail files are expressly forbidden from being linked to other data sources and do not include information on income.

The standard approach in studies of fertility in the United States is to define fertility rates using aggregate births from U.S. Vital Statistics for detailed subgroups, combined with population estimates from the U.S. Census Bureau for the analogous subgroups. This approach has the advantage that births reported are comprehensive, being drawn from the official records of all births in the United States. However, a disadvantage is that certain variables that may be of interest in the analysis to identify the population exposed to treatment, such as income, are not available. Moreover, to analyze the effects on fertility rates, population counts used as denominators for the group of interest are always necessary. The Census provides intercensal estimates of population subgroups, such as by state, age, sex, and race or ethnicity, but they limit how fine-grained these groupings can be since estimates become increasingly noisy as they get smaller. For example, they publish state- and year-level population estimates by sex and single year of age, but to obtain population estimates for race or ethnicity they instead group them into multi-year age categories.

Given these limitations, I carry out complementary analyses using both the standard approach with Vital Statistics natality data as well as the American Community Survey, which is a large representative cross-sectional sample of the U.S. population that includes information on births in the past year, as well as detailed information on income and demographics. Using the ACS allows for studying more specifically whether there was an effect on individuals that became newly eligible for Medicaid coverage.

Table 4: Summary statistics for births: American Community Survey

(a) Family planning waiver programs sample (2001-2009)

	Full Sample			Eligible		
	All (1)	Non-Parents (2)	Parents (3)	All (4)	Non-Parents (5)	Parents (6)
Age	30.34 (8.92)	26.15 (8.65)	35.67 (5.91)	28.62 (8.62)	24.49 (8.10)	33.37 (6.49)
High school grad.	0.55 (0.50)	0.48 (0.50)	0.63 (0.48)	0.64 (0.48)	0.56 (0.50)	0.72 (0.45)
College grad.	0.28 (0.45)	0.27 (0.44)	0.28 (0.45)	0.10 (0.30)	0.12 (0.33)	0.08 (0.27)
Married	0.48 (0.50)	0.27 (0.44)	0.74 (0.44)	0.28 (0.45)	0.13 (0.34)	0.45 (0.50)
Total family income, thousands	69.43 (69.40)	67.95 (69.10)	71.31 (69.73)	18.75 (15.76)	16.86 (16.34)	20.92 (14.78)
Worked last year	0.76 (0.43)	0.74 (0.44)	0.78 (0.42)	0.63 (0.48)	0.60 (0.49)	0.67 (0.47)
Birth, last 12 mos.	0.06 (0.24)	0.10 (0.30)	0.02 (0.14)	0.09 (0.28)	0.14 (0.34)	0.03 (0.17)
No. of children			2.01 (0.98)			2.20 (1.15)
Eldest child			10.66 (5.51)			10.65 (5.43)
Youngest child			7.19 (4.95)			6.60 (4.76)
Observations	2810874	1572677	1238197	782906	418541	364365

(b) ACA Medicaid expansions sample (2010-2019)

	Full Sample			Eligible		
	All	Non-Parents	Parents	All	Non-Parents	Parents
	(1)	(2)	(3)	(4)	(5)	(6)
Age	29.96	26.18	35.81	28.43	24.64	33.61
	(8.68)	(8.15)	(5.72)	(8.42)	(7.75)	(6.28)
High school grad.	0.51	0.47	0.57	0.64	0.58	0.72
	(0.50)	(0.50)	(0.50)	(0.48)	(0.49)	(0.45)
College grad.	0.34	0.32	0.36	0.13	0.15	0.10
	(0.47)	(0.47)	(0.48)	(0.34)	(0.36)	(0.30)
Married	0.42	0.23	0.72	0.22	0.10	0.38
	(0.49)	(0.42)	(0.45)	(0.41)	(0.30)	(0.49)
Total family income, thousands	82.27	80.15	85.53	16.73	15.16	18.87
	(87.83)	(86.63)	(89.56)	(18.86)	(19.45)	(17.80)
Worked last year	0.73	0.71	0.76	0.56	0.53	0.58
	(0.44)	(0.45)	(0.43)	(0.50)	(0.50)	(0.49)
Birth, last 12 mos.	0.06	0.09	0.02	0.08	0.12	0.03
	(0.24)	(0.28)	(0.14)	(0.27)	(0.32)	(0.17)
No. of children			2.05			2.25
			(1.03)			(1.20)
Eldest child			10.44			10.68
			(5.58)			(5.55)
Youngest child			6.79			6.30
			(4.79)			(4.65)
Observations	1194200	724443	469757	272617	157421	115196

This table presents summary statistics by sample and study period for the American Community Survey, waves 2001-2019. All samples include women between the ages of 18 to 44. Samples are further defined by prior parental status (a child older than one) and eligibility with respect to the program of interest: under 185% of the Federal Poverty Line for the family planning waiver programs, and under 138% of the Federal Poverty Line for the ACA Medicaid expansions. Averages are reported for each characteristic, with the standard deviation in parentheses.

Given that existing evidence highlights potentially meaningful differences in effects of subsidization on teens and non-teens, I consider differential effects by these age groupings. For the Vital Statistics natality data, I construct fertility rates by year for teens and non-teens. Throughout the analysis, non-teens indicates women between the ages of 20-45. However, teens differ: For the family planning waiver program analysis, teens includes those ages 15-19, since in most states teens could access care through these programs. For the ACA Medicaid analysis, those under the age of 18 were unaffected, so I characterize teens as those who were 18 or 19. Note that while ages for mothers in Vital Statistics data is as of the time of birth, age for women in the ACS is at the time of interview, and the birth variable regards birth in the prior 12 months. Thus, I assume that the age of (potential) birth for women is one year less than their reported birth, e.g. to estimate the effects on 18 and 19-year-olds I select women who report being 19 or 20 at the time of the interview.

One limitation of using the ACS is that we only observe births in the past year after the year 2000. Thus, while Vital Statistics natality data are available for the full period of 1990-2019, the ACS cannot offer a full comparison of the 1990-2009 period of family waiver planning program expansions.

8.2 Implementation

Analysis of the aggregate Vital Statistics natality follows the same approach as for the contraceptive analysis, using Eqs. 3- 7, with the exception that the dependent variable is the log of the fertility rate for women in a given state, year, and age group (teen or non-teen). This approach is also comparable to the main results in Kearney and Levine (2009). In specifications with controls, for teens I control for the fraction of married individuals in the population, the fraction that are non-white, and the fraction that are Hispanic. For non-teens, I additionally control for the state's age distribution by controlling for the fractions that are between ages 20-24, 25-29, 30-34, and 35-39.[6]

With the ACS, we observe individual-level birth outcomes, and can therefore carry out individual-level estimation. Moreover, in contrast to the state-level analysis for contraception and for the aggregate natality data, I can make use of several different samples in an attempt to improve power by narrowing in on the subpopulations that are more likely to be directly "treated" by becoming eligible for coverage. In this respect, in addition to doing an initial comparison by age groups, I will also split the sample along two dimensions: whether or not an individual had household income relative to the federal poverty line that put them under the eligibility threshold for a given program (I consider the relevant threshold for family planning waiver programs to be 185% of the federal poverty line and 138% for ACA Medicaid expansions), and whether or not individuals were already parents. However, for these analyses I will focus on the older age group of women age 20-43 at the time of birth (or 21 to 44 at the time of interview), both since very few teens are already parents, and since income may not be measured accurately for teens since many live with other adults including their parents.

To be consistent with the aggregate analysis, where treated states are weighted equally, the individual-level analysis to yield comparable estimates must correct for the fact that larger states will have more observations. To do so, I apply weights for each individual in a regression equal to $\frac{1}{N_s}$ where N_s is the total number of observations for a given state in the sample of interest.

[6]Controls for the period 1990-2009 are taken from the Current Population Survey, since the ACS is not available annually prior to 2000 and this is the same source of controls used in Kearney and Levine (2009). For 2010 onwards, the controls are constructed from the ACS.

To provide a baseline comparison with the contraception and aggregate natality estimates, the first sample will simply include the full population without regard to parenthood or eligibility. Next, I will consider the full populations of non-parents and parents respectively. Finally, I will consider the same three groupings restricted only to individuals with levels of income under the relevant eligibility threshold.

Using a difference-in-difference approach for subgroup analysis in this way is valid for identification of treatment effects under the assumption that there are not regional spillovers of the expansion policy (even if there are within-state spillovers) and the "potentially exposed" subgroups under the threshold follow parallel counterfactual trends (Huber and Steinmayr, 2019). Being able to restrict focus to eligible individuals may yield improved power, since it will exclude individuals who should be unaffected.

The dependent variable in each analysis is simply whether or not an individual gave birth within the last 12 months. Eligibility is defined whether an individual's household income falls under the eligibility threshold relative to the federal poverty line, given the number of individuals in their household (this variable is calculated by the Census Bureau and provided in the ACS). In specifications described as without controls, I include only state and year fixed effects (excepting the Callaway and Sant'Anna estimator, which does not incorporate unit fixed effects). In specifications with controls, I include fixed effects for single year of age, parental status, four levels of educational attainment, and Hispanic ethnicity.

8.3 Results

In this section, we consider the estimated average treatment effects on the treated for both family planning waiver programs and the ACA Medicaid expansions in terms of births. While the estimates within this section are largely displayed in figures, corresponding tables which indicate exact coefficient values as well as number of observations are included in Appendix A.

In general, I find little evidence of negative effects on fertility, and even considering the full span of confidence intervals, most estimates generally rule out effect sizes greater than 5% and in many cases even 3% reductions in births. This pattern largely holds even when considering effects only for the population of individuals that were eligible for a given policy. There are two exceptions where I find significant and negative effects on fertility: first, for teens 18 to 19 years old at the time of birth, I find reductions of approximately 6% under the ACA Medicaid expansions considering Natality data. This finding is corroborated although less precisely with ACS data.

Second, when examining subgroups among non-teen adults using Natality data, I estimate reductions in fertility among eligible parents for family planning waiver program expansions, for which several of the estimates imply negative effects on births equivalent to around 5% reductions.

Beginning with estimates by age using Natality data, Fig. 5 displays the estimated *ATTs* using different estimators. As was the case with the contraception results, there is little difference across specifications in most of the point estimates, which is suggestive of limited heterogeneity across time or units. The notable result here is with respect to teens under the ACA, who exhibit significant decreases in fertility across all specifications.

Pre-trends are graphed along side ATT_k estimates in Fig. 6 based on the Gardner estimator with controls. Considering pretrends by inspection, there do not appear to be obvious differences in trends between treated and untreated units. For all the estimates for teens under the ACA, post-treatment coefficients are largely centered close to zero. For the estimates for teens under the ACA, several although not all of the post-treatment coefficients are negative and significant at the 5% level, and there is no clear dynamic trend, with confidence intervals largely overlapping.

We can compare the estimates from the aggregate data by age using the ACS instead. Since this analysis is estimated with individual data as a linear probability model with the outcome being whether or not an individual gave birth in the last year, coefficients are expressed in terms of the absolute effect of a treatment on the likelihood of birth. Fig. 7 shows the overall estimates across all treatment periods. Note that the estimates for the family planning waiver period are based only on 2001-2009, rather than 1990-2009 as for the Natality data.

Similar to the results from the Natality data, I find no significant effects except for teens under the ACA, although here the are somewhat more imprecisely estimated: the imputation estimators are all negative and significant at the 10% level. The point estimates imply somewhat larger effects, equivalent to about 10% reductions relative to baseline levels of births, but but but given substantial overlap they provide support for the estimates based on Natality data.

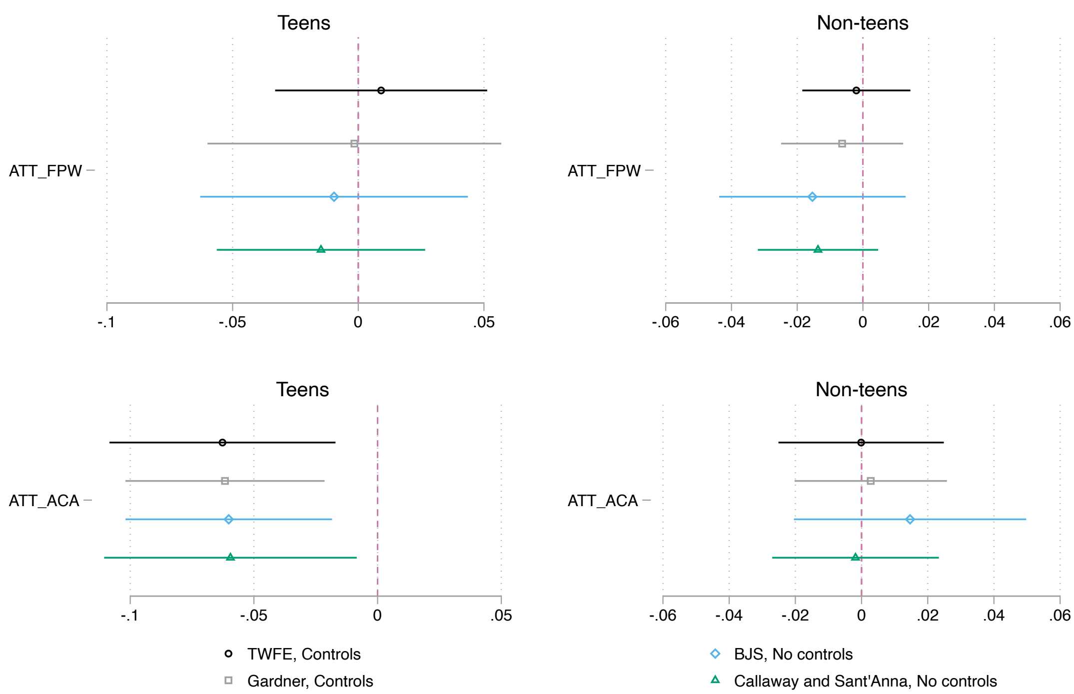
Considering the event-study estimates in Fig. 8, aside from the first period pre-treatment coefficients (which may be significantly different from zero because they are based on fewer treated units, and hence noiser), there do not appear to be different pre-trends for the treated and untreated. Curiously, there are some and significant coefficients for family planning waiver post-treatment estimates (for 4 and 6 years after treatment), but these confidence intervals still overlap substantially with pre-treatment coefficients, suggesting an interpretation as a meaningful treatment effect should be taken with a grain of salt. For teens under the ACA, point estimates are necessarily less precise than even for the overall treatment effect, but the point estimates are nevertheless negative and ranging from -0.002 to -0.006. For non-teens, there is one significant post-treatment coefficient only for the family planning waiver programs, six years after treatment. Again, this may owe more to the fact that the latest and earliest periods are sometimes noiser due to being based on fewer treated units. Even if we interpret it as representative of a true treatment effect, this estimate would correspond to approximately a 6% reduction in fertility. With respect to the non-teen results for the ACA, there are no significant coefficients either for the pre-treatmetn period or the post-treatment period.

These initial results by age provide evidence that the ACA at least led to a reduction in teen births.

Given we find no effect for non-teens, one question is whether we do not find an effect simply because the treated individuals represent only a minority of the overall population included for estimation in the aggregate fertility rate. To consider this possibility, we turn to the ACS, which allows us to explore deeper by looking at individauls who specifically became eligible for one or the other policy under expansion, or by parental status, since non-parents truly had no other source of public health insurance coverage to pay for contraceptives.

Fig. 9 illustrates the average effects across all periods, for the different samples, estimators, and specifications. While the treatment effects are expressed as the marginal effect of the policy on the probability of a birth, red and blue dotted lines above and below the estimate provide indications of effect sizes that would be 3% and 5% relative to the baseline mean of births for the given sample (in the first year of the sample, which is 2001 for the family planning waiver program analysis and 2010 for the

Figure 5: Overall Average Treatment Effects on the Treated for Births (Natality data)

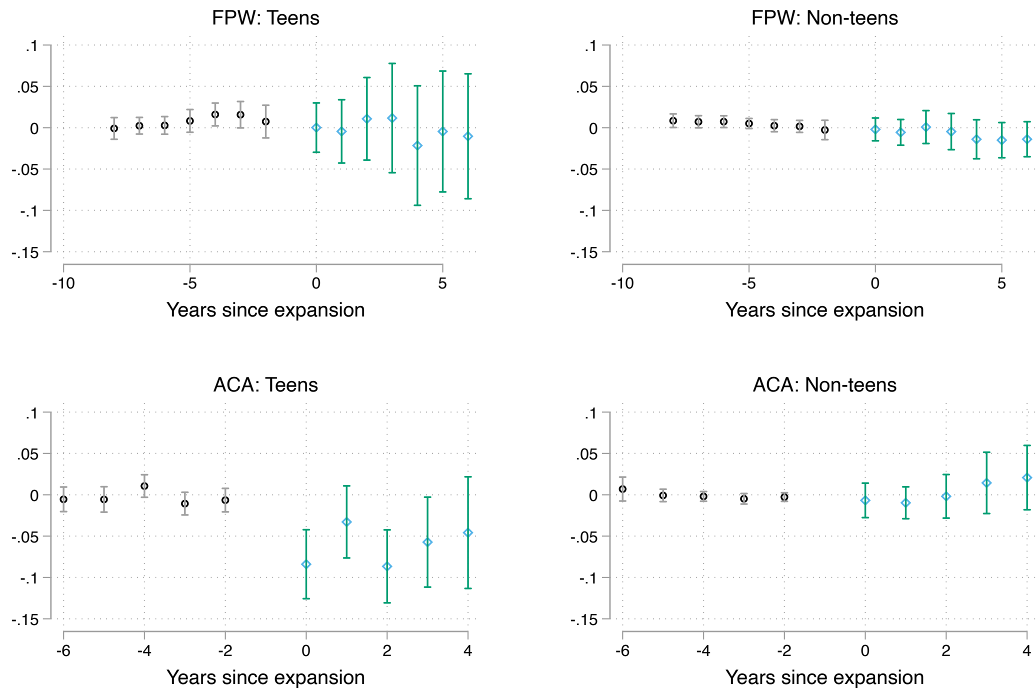


This figure displays estimated ATTs for the log fertility rate, based on fertility rates constructed from U.S Vital Statistics Natality data combined with population estimates from the U.S. Census Bureau between 1990 to 2019, where family planning waiver programs are the treatment of interest for the period 1990-2009, and ACA Medicaid expansions for 2010-2019. Non-teens include women ages 20-44 at the time of potential birth. Teens include 15-19 year olds for family planning waiver program analysis, and 18 and 19 year olds for the ACA Medicaid expansions. Parameters are estimated according to the methods described in Sec. 6.

ACA Medicaid expansions analysis). Panel (a) displays results from the analysis of family planning waiver programs between 2001 and 2009. Point estimates are generally close to zero, and no estimates are significantly different from zero at even the 10% level except for the sample of eligible parents. Moreover, for the non-parents sample, TWFE estimates are very similar to the imputation estimator results, while CS estimates are slightly more positive (note that if the true effect of the policy is zero, and hence constant across groups and time, TWFE would be unbiased and the different estimators would be expected to yield the same results).

In contrast, there is some evidence of negative effects on births for eligible parents. While for the sample of eligible parents, the TWFE estimates are extremely close to zero, while both the estimates using the imputation estimators and CS estimator are more negative, with point estimates of approximately -0.002 exceeding 5% reductions in births relative to the baseline. Here we see also the benefits of the more efficient inference procedure for the BJS estimator: while the point estimates are identical for Gardner and BJS, the smaller confidence interval for the BJS estimates implies that these treatment effects are significantly different from zero at the 10% level. Finally, the CS estimate is in fact even more negative at -0.0044, and significant at the 1% level. Considering these specifications together, they suggest reductions in births for eligible parents of between 2 to 4 births per thousand women of

Figure 6: Event-Study Estimates for Births (Natality data)



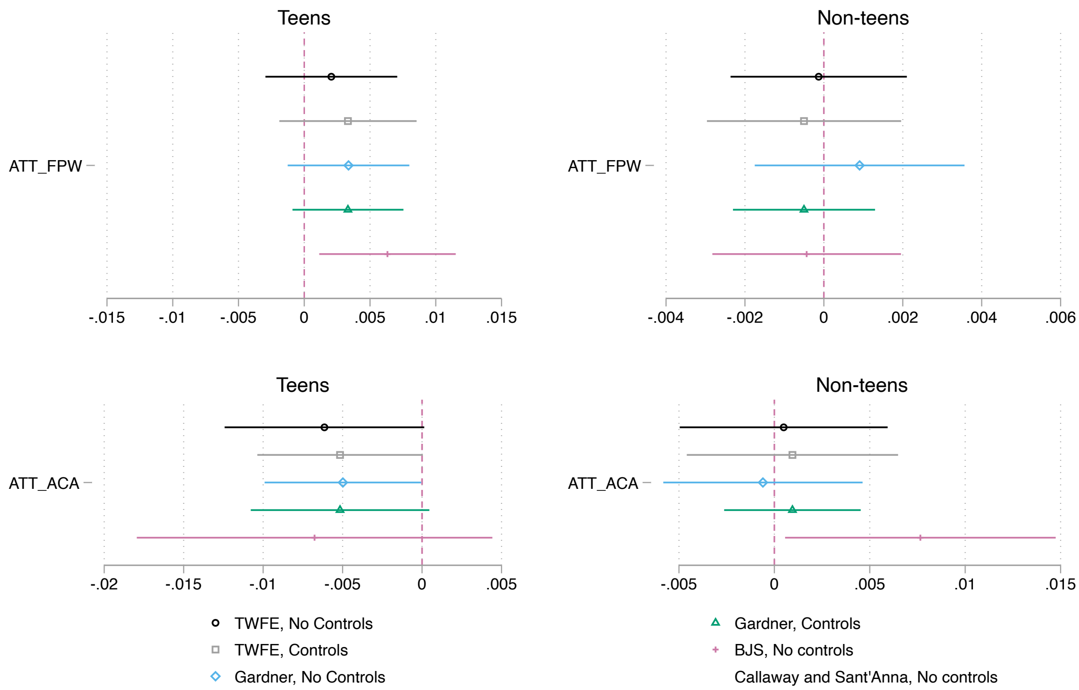
childbearing age.

Results are largely similar for the estimated effects of the ACA Medicaid expansion. When the sample is not restricted by eligibility, point estimates are close to zero across all samples and confidence intervals are generally within an effect size bound of 3% or 5% relative to the baseline level of births, although for the parental sample confidence intervals extend below this effect size bound (despite being more precise in absolute terms compared to the non-parental sample) given that the baseline level of births for parents is substantially lower.

For the samples restricted by eligibility status, point estimates are more negative, though still not significantly different than zero, with results from TWFE and imputation estimators ranging from -0.002 to -0.003 and centered around the level of a 3% reduction in births, with lower bounds for confidence intervals exceeding the 5% effect size. These more negative estimates appear to be driven by non-parents, where the TWFE and imputation estimator results range from -0.0018 to -0.003, whereas for parents the point estimates for the analogous estimates range from -0.0015 to -0.0019. At the same time, the lower bound of the confidence intervals, while higher in absolute levels than for non-parents, extend to the equivalent of an approximately 10% reduction relative to the baseline for births.

One notable feature of the results for the ACA Medicaid expansion is that the Callaway and Sant'anna estimator is both notably more positive and more imprecise across nearly all samples and specifications, although confidence intervals substantially overlap with the other estimators in all cases. The difference in precision is one of the factors that has been previously discussed for this estimator, but the fact that the confidence intervals nevertheless largely overlap implies there is little

Figure 7: Overall Average Treatment Effects on the Treated for Births (ACS)



This figure displays estimated ATTs on births using data from the ACS between 2001 to 2019, where family planning waiver programs are the treatment of interest for the period 2001-2009, and ACA Medicaid expansions for 2010-2019. Non-teens include women ages 20-43 at the time of potential birth. Teens include 15-19 year olds for family planning waiver program analysis, and 18 and 19 year olds for the ACA Medicaid expansions. Parameters are estimated according to the methods described in Sec. 6.

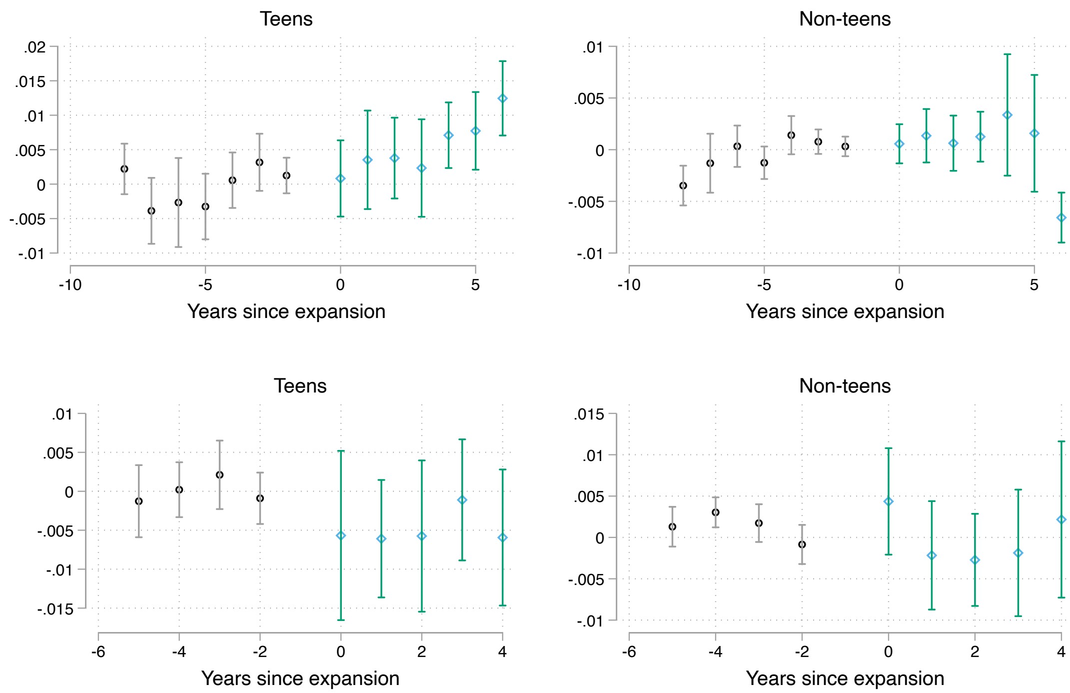
reason for these more positive estimates to affect the overall interpretation of the results.

To examine pre-trends and whether there are dynamic treatment effects present within these subgroups, Fig. 8 also presents event-study estimates using the Gardner imputation estimator with demographic controls. For samples unrestricted by eligibility, there is little evidence of dynamic treatment effects, with both pre- and post-treatment coefficients and confidence intervals largely overlapping.

For the samples of individuals under the eligibility threshold, none of the pre-treatment coefficients are significantly different from zero. However, the post-treatment coefficients are declining between year zero and three. When broken down by parental status, we see a further corrobation of the ATTs averaged across all periods, where for non-parents all coefficients are not significantly different from zero, while for parents all periods from three years after treatment are significantly different from zero at the 5% level and the point estimates appear to be trending downward, suggesting an effect that is increasing in magnitude over time.

For the ACA Medicaid expansion, with event-study results depicted in Panel (b), there is likewise little evidence of differences in trends in the four years prior prior to treatment compared to never-treated units, as none of the coefficients on pre-periods is significantly different from zero at the 5% level. For the post-treatment periods, several samples show significant estimated decreases in births for two years after treatment, but no other periods. This includes for the whole sample, with a point

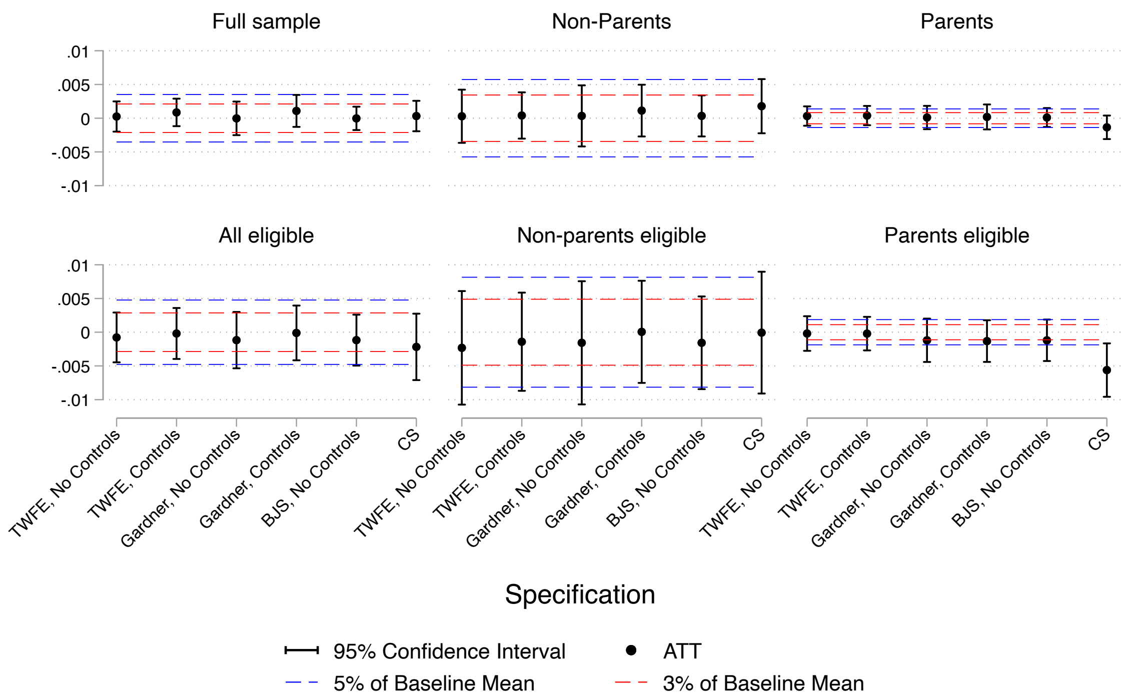
Figure 8: Event-Study Estimates for Births (ACS)



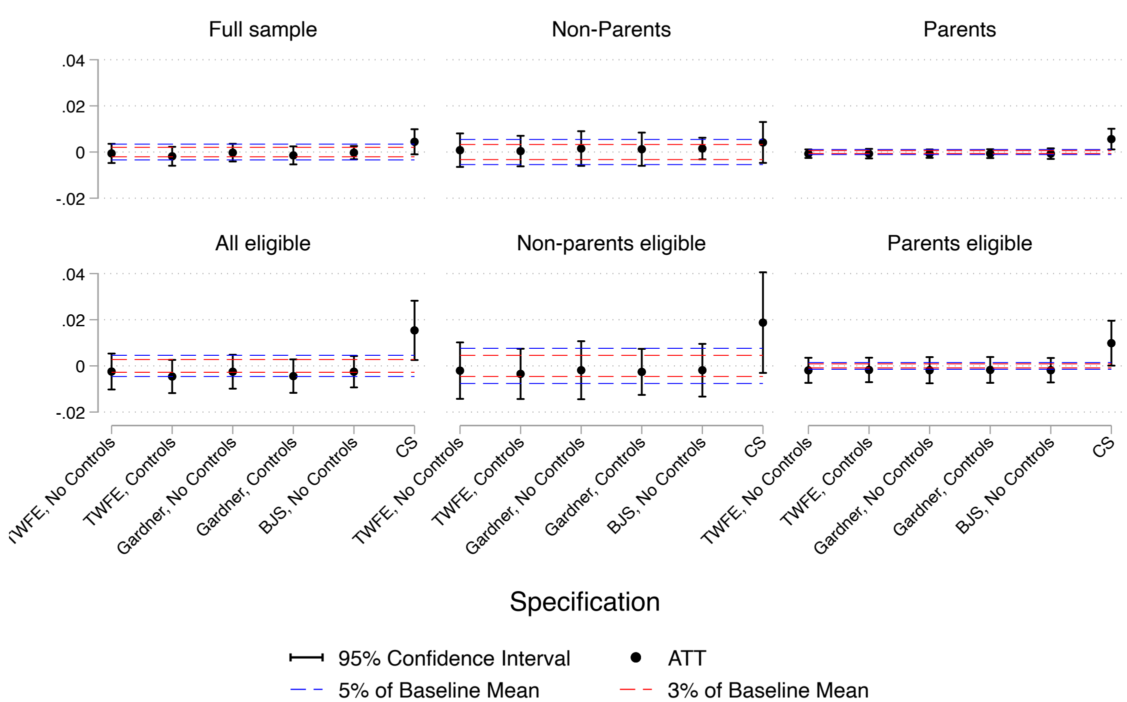
This figure displays estimated ATT_k's using data from the ACS between 2001 to 2019, where family planning waiver programs are the treatment of interest for the period 2001-2009, and ACA Medicaid expansions for 2010-2019. Non-teens include women ages 20-43 at the time of potential birth. Teens include 15-19 year olds for family planning waiver program analysis, and 18 and 19 year olds for the ACA Medicaid expansions. Parameters are estimated using the Gardner imputation method.

Figure 9: Overall Average Treatment Effects on the Treated for Births

(a) Family Planning Waiver Programs, 2001-2009



(b) ACA Medicaid Expansion, 2010-2019



This figure presents results estimating the effects on births of two types of Medicaid expansion policies (indicated in the panel title) for women ages 18-44 using data from the American Community Survey between 2001 to 2019. The x-axis labels indicate the estimator and specification used, according to the description in Sec. 8.2. The subplot titles indicate the sample restriction, if any. Eligible individuals include those under a household income threshold of 185% or 138% of the federal poverty line for family planning waiver programs and ACA Medicaid expansions respectively. Red dotted lines denote the level that would be consistent with an effect size of a 3% change in births for the given sample, and blue lines denote 5% effect sizes.

estimate of -0.0025, and for the eligible individuals only, with a very large point estimate of -0.01, equivalent to about a 10% reduction in births for this group. However, in all samples the coefficients of subsequent periods are closer to zero and not significant.

8.4 Discussion

The preceding results regarding births provide limited evidence of a negative effect of either family planning waiver programs or the ACA Medicaid expansion on births. Specifically, there is some indication that there may have been a negative effect on teens under the ACA and on subsequent births for individuals who already were parents under family planning waiver programs.

One possibility why there is a negative effect for teens under the ACA and not under family planning waiver programs could be that doctors only began recommending adolescents and nulliparous receive long-acting reversible contraceptives as of 2009. This conclusion is supported by the fact that the ACA but not the family planning waiver programs led to increases in long-acting reversible prescriptions.

Beyond the teens under the ACA and the parents under the family planning waiver programs, most confidence intervals were contained within a range consistent with about a 5% increase or decrease in births–but one might wish to be able to say something more precise. After all, there are many interventions that change outcomes like fertility by only a few percent that are nevertheless considered to be of interest.

Are the extremes of the confidence intervals plausible? Naturally, given the assumptions of our estimation, we expect there is greater likelihood that the true effects are towards the center of the confidence interval, so in a statistical sense the central point estimates are more plausible, which with the exception of parents under family planning waiver programs are estimates that are largely negative but small in magnitude, in the range of 0 to -0.003 (corresponding to zero to 3 births per thousand women of childbearing age).

Moreover, without strong evidence to the contrary, we should be skeptical that these policies could have led to increases in fertility, which should lead us to largely rule out the plausibility of the positive end of the confidence intervals.

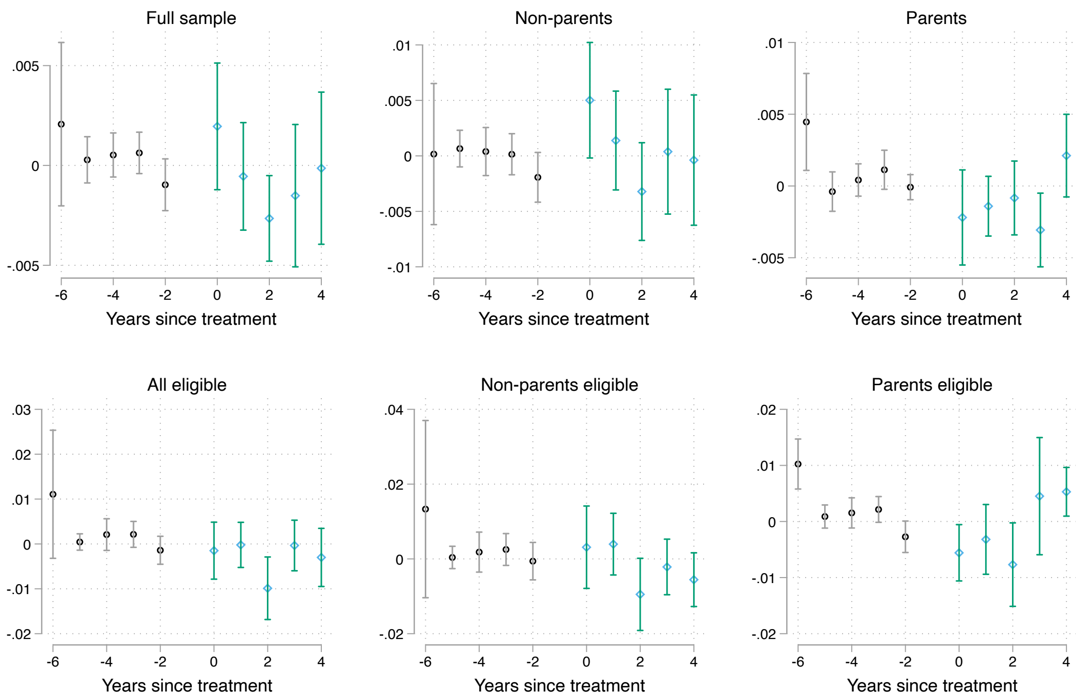
Finally, the results and predictions regarding contraceptive reimbursements can also help guide us in assessing the estimated effects for births. Under the most plausible scenarios, the likely prediction for an overall ATT on births is more in the range of a reduction of about 1 to 2 births per thousand women of childbearing age.

If these values are then scaled up by approximately three or four times to reflect that these decreases should come only from the population eligible for each program, we arrive at plausible possible decreases of about 3 to 4.5 births per thousand eligible women of childbearing age. As a result, the lower bounds of the confidence intervals even for many of the non-significant results are not necessarily implausible even when taking into account the contraceptive results, although the true effect is likely closer to the center of the confidence intervals. Overall, these would still represent modest reductions in the likelihood of birth for gaining eligibility for insurance.

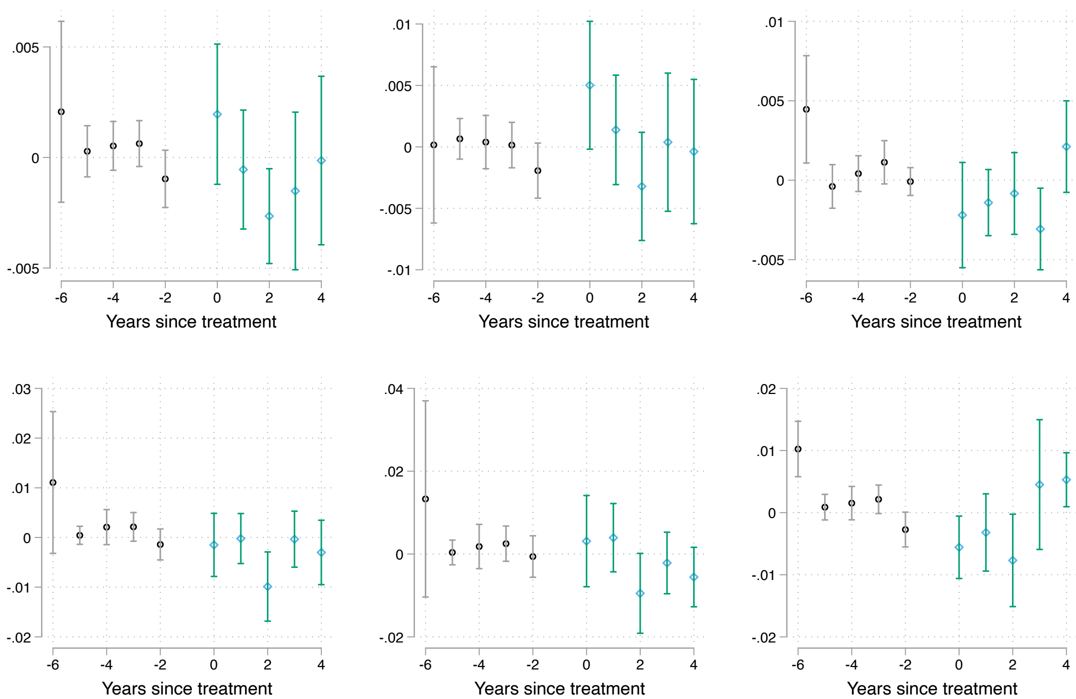
My results also suggest substantively different conclusions in particular about the effects of family planning waiver programs compared to the results in Kearney and Levine (2009) and Lindrooth and McCullough (2007), who find significant reductions in fertility for both non-teens and teens. To

Figure 10: Births: Average Treatment Effects on the Treated by Horizon

(a) Family Planning Waiver Programs



(b) ACA Medicaid Expansion



better understand this difference, in Appendix B I replicate the results from Kearney and Levine (2009) and show the difference depends entirely on the fact that they employ population weights in the regression, whereas I do not. While using population weights is not wrong, per se, it does not estimate the ATT, since it in fact weights treatment effects for individual treatment states by the size of their population, rather than equally. While this might be desirable if we are estimating a treatment effect for historical, retrospective purposes–to try to understand the average effect a particular type of intervention has had on the country as a whole–it is less useful if we wish for our causal parameter to be relevant for thinking prospectively about future implementations of the same policy. To give some intuition, there is no reason (*a priori*) that we should expect future potential expanding states to be more like California than Wyoming, but a population-weighted average treatment effect implicitly suggests just that. Thus, for the purpose of the question of interest in this paper, I would argue that omitting population weights is preferable. Doing so yields results that family planning waiver programs had no effect on fertility, excepting perhaps small effects on eligible parents.

9 Conclusion

Substantial efforts and resources have gone towards reducing the financial cost of contraception via public health insurance programs. The findings in this paper can add to the existing body of evidence that subsidies may decrease births, in particular for teens. However, they appear to produce modest reductions in fertility, if any at all. The findings of this paper suggest that to the extent we wish to further reduce unintended births, we must look to other methods to do so.

If reducing financial costs doesn't do it, what will? Returning to the question of whether improved counseling has a role, studying provider behavior might be an alternative route to yield new insights on this question. Notably, Buckles et al. (2022) point to a change by the American College of Obstetricians and Gynecologists in 2009 towards recommending LARCs for all women including adolescents, not only post-partum women, as a likely major factor in explaining the decline in unintended births between 2007 and 2016. Presumably, this is a change that should have substantially affected provider decisions and recommendations to patients.

Beyond providers, the limited effects of individual coverage also suggests the importance of further exploration of the factors giving rise to unintended birth, including perhaps assessing its welfare implications – if individuals become pregnant in part because they may be ambivalent about pregnancy, how concerning is such an outcome even if they classify the pregnancy as unintended? On the other hand, ambivalence about pregnancy may in part arise because certain individuals' alternative opportunities are limited, making the opportunity cost of unintended childbirth correspondingly low. This framing highlights the possibility that a focus on the costs of contraception might simply be looking at incentives in the wrong part of the equation: a lowering of the costs of contraception may have little effect if the perceived costs of unintended birth are too low. Such a reality would still imply that unintended births may present important challenges to address.

References

Bailey, M. J. (2012). Reexamining the Impact of Family Planning Programs on US Fertility: Evidence from the War on Poverty and the Early Years of Title X. *American Economic Journal: Applied Economics 4*(2), 62–97.

Bailey, M. J., V. W. Lang, I. Vrioni, L. Bart, D. Eisenberg, P. Fomby, J. Barber, and V. K. Dalton (2021). How subsidies affect contraceptive use among low-income women in the us: A randomized control trial. In *UCLA Working Paper*.

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Buckles, K., M. Guldi, and L. Schmidt (2022). The great recession's baby-less recovery: The role of unintended births. *Journal of Human Resources*, 1220–11395R3.

Buckles, K. S. and D. M. Hungerman (2018). The incidental fertility effects of school condom distribution programs. *Journal of Policy Analysis and Management 37*(3), 464–492.

Butts, K. (2021). *did2s: Two-Stage Difference-in-Differences Following Gardner (2021)*.

Butts, K. and J. Gardner (2021). Did2s: Two-stage difference-in-differences. *arXiv preprint arXiv:2109.05913*.

Callaway, B. and P. H. Sant'Anna (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–96.

England, P., M. L. Caudillo, K. Littlejohn, B. C. Bass, and J. Reed (2016). Why do young, unmarried women who do not want to get pregnant contracept inconsistently? mixed-method evidence for the role of efficacy. *Socius 2*, 2378023116629464.

Frost, J. J. and J. E. Darroch (2008). Factors Associated with Contraceptive Choice and Inconsistent Method Use, United States, 2004. *Perspectives on Sexual and Reproductive Health 40*(2), 94–104.

Gardner, J. (2021). Two-stage differences in differences. Technical report, Working paper.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Guzzo, K. B. and S. Hayford (2011). Fertility Following an Unintended First Birth. *Demography 48*(4), 1493–1516.

Huber, M. and A. Steinmayr (2019). A framework for separating individual-level treatment effects from spillover effects. *Journal of Business & Economic Statistics*, 1–15.

Kearney, M. S. and P. B. Levine (2009). Subsidized contraception, fertility, and sexual behavior. *The review of Economics and Statistics 91*(1), 137–151.

Kelly, A., J. M. Lindo, and A. Packham (2020). The power of the iud: Effects of expanding access to contraception through title x clinics. *Journal of Public Economics 192*, 104288.

Lindo, J. M., C. K. Myers, A. Schlosser, and S. Cunningham (2020). How far is too far? new evidence on abortion clinic closures, access, and abortions. *Journal of Human resources 55*(4), 1137–1160.

Lindo, J. M. and A. Packham (2017). How much can expanding access to long-acting reversible contraceptives reduce teen birth rates? *American Economic Journal: Economic Policy 9*(3), 348–76.

Lindrooth, R. C. and J. S. McCullough (2007). The effect of medicaid family planning expansions on unplanned births. *Women's Health Issues 17*(2), 66–74.

Mansour, D., P. Inki, and K. Gemzell-Danielsson (2010). Efficacy of contraceptive methods: a review of the literature. *The European Journal of Contraception & Reproductive Health Care 15*(1), 4–16.

Miller, S., N. Johnson, and L. R. Wherry (2021). Medicaid and mortality: new evidence from linked survey and administrative data. *The Quarterly Journal of Economics 136*(3), 1783–1829.

Myers, C. K. (2017). The power of abortion policy: Reexamining the effects of young women's access to reproductive control. *Journal of Political Economy 125*(6), 2178–2224.

Roth, J. and P. H. Sant'Anna (2020). When is parallel trends sensitive to functional form? *arXiv preprint arXiv:2010.04814*.

Sant'Anna, P. H. and J. Zhao (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics 219*(1), 101–122.

Sumarsono, A., M. W. Segar, L. X. PharmD, F. Atem, S. E. Messiah, J. K. Francis, and N. Keshvani (2020). Medicaid Expansion and Provision of Prescription Contraception to Medicaid Beneficiaries. *Contraception 103*(3), 199–202.

Sun, L. and S. Abraham (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.

Trussell, J. (2004). Contraceptive failure in the united states. *Contraception 70*(2), 89–96.

Yoo, S. H., K. B. Guzzo, and S. R. Hayford (2014). Understanding the complexity of ambivalence toward pregnancy: Does it predict inconsistent use of contraception? *Biodemography and social biology 60*(1), 49–66.

A Tables corresponding to figures

Since many estimates in the text are presented in figures for ease of interpretation, the following tables provide precise numeric estimates, standard errors, and the number of observations for each of the estimates in these figures.

Table A1: Contraceptives: Overall average treatment effects on the treated (corresponding to Fig. Figure 1)

(a) Family planning waiver programs

(1) Contraceptive Type	(2) Specification	(3) Estimate	(4) Standard Error	(5) Observations
Short-Acting Hormonal	TWFE	2.779	1.006	3479
	TWFE, Controls	2.816	0.937	
	BJS	3.759	0.716	
	Gardner	3.759	1.284	
	Gardner, Controls	3.825	1.200	
	CS	3.696	1.089	
DMPA	TWFE	2.439	1.124	3477
	TWFE, Controls	2.374	1.091	
	BJS	3.077	0.430	
	Gardner	3.077	1.271	
	Gardner, Controls	3.043	1.262	
	CS	2.634	1.178	
Long-acting reversible	TWFE	0.036	0.459	3475
	TWFE, Controls	-0.054	0.453	
	BJS	-0.128	0.355	
	Gardner	-0.128	0.446	
	Gardner, Controls	-0.174	0.449	
	CS	0.380	0.327	

(b) Affordable Care Act Medicaid Expansions

(1) Contraceptive Type	(2) Specification	(3) Estimate	(4) Standard Error	(5) Observations
Short-Acting Hormonal	TWFE	7.891	2.453	675
	TWFE, Controls	7.605	2.351	
	BJS	8.503	2.404	
	Gardner	8.503	2.402	
	Gardner, Controls	8.438	2.386	
	CS	7.073	2.589	
DMPA	TWFE	1.879	1.918	659
	TWFE, Controls	1.949	2.069	
	BJS	2.107	1.767	
	Gardner	2.107	1.790	
	Gardner, Controls	2.406	2.198	
	CS	1.018	1.836	
Long-acting reversible	TWFE	2.856	1.410	675
	TWFE, Controls	2.543	1.008	
	BJS	3.096	1.456	
	Gardner	3.096	1.414	
	Gardner, Controls	3.194	1.324	
	CS	2.048	0.932	

This table presents overall average treatment effects on treated states using data from the State Drug Utilization Program, 1992-2019, for Family Planning Waiver Programs (Panel (a)) and ACA Medicaid expansions (Panel (b)). Coefficients are in terms of contraceptive coverage rates per 1,000 women aged 15-44, by type of contraceptive (short-acting hormonal; DMPA; or long-acting reversible). Col 1. indicates the type of contraceptive as the outcome of interest, and Col. 2 indicates which estimator and control variables were used, following the description in Sec. 7.2.

Table A2: Fertility: overall average treatment effects on the treated using Vital Statistics (corresponding to Fig. 5)

(a) Family planning waiver programs, 1990-2009

(1) Sample	(2) Specification	(3) Estimate	(4) Standard Error	(5) Observations
Teens	TWFE	-0.063	0.022	180
	BJS	-0.060	0.021	
	Gardner	-0.062	0.021	
	CS	-0.059	0.026	
Non-Teens	TWFE	0.000	0.012	180
	TWFE, Controls	0.000	0.012	
	BJS	0.015	0.018	
	Gardner	0.003	0.012	
	CS	-0.002	0.013	

(b) Affordable Care Act Medicaid expansions

(1) Sample	(2) Specification	(3) Estimate	(4) Standard Error	(5) Observations
Teens	TWFE	-0.063	0.022	180
	BJS	-0.060	0.021	
	Gardner	-0.062	0.021	
	CS	-0.059	0.026	
Non-Teens	TWFE	0.000	0.012	180
	TWFE, Controls	0.000	0.012	
	BJS	0.015	0.018	
	Gardner	0.003	0.012	
	CS	-0.002	0.013	

Table A3: Fertility: Overall average treatment effects on the treated for family planning waiver programs (corresponding to Fig. 9a)

(1) Sample	(2) Specification	(3) Estimate	(4) Standard Error	(5) Observations
Full Sample	TWFE	0.000	0.001	2508544
	TWFE, Controls	0.001	0.001	
	Gardner	0.000	0.001	
	Gardner, Controls	0.001	0.001	
	CS	0.000	0.001	
Non-Parents	TWFE	0.000	0.002	1292316
	TWFE, Controls	0.000	0.002	
	Gardner	0.000	0.002	
	Gardner, Controls	0.001	0.002	
	CS	0.002	0.002	
Parents	TWFE	0.000	0.001	1216228
	TWFE, Controls	0.000	0.001	
	Gardner	0.000	0.001	
	Gardner, Controls	0.000	0.001	
	CS	-0.001	0.001	
All Eligible	TWFE	-0.001	0.002	687797
	TWFE, Controls	0.000	0.002	
	Gardner	-0.001	0.002	
	Gardner, Controls	0.000	0.002	
	CS	-0.002	0.003	
Eligible Non-Parents	TWFE	-0.002	0.004	330967
	TWFE, Controls	-0.001	0.004	
	Gardner	-0.002	0.005	
	Gardner, Controls	0.000	0.004	
	CS	0.000	0.005	
Eligible Parents	TWFE	0.000	0.001	356830
	TWFE, Controls	0.000	0.001	
	Gardner	-0.001	0.002	
	Gardner, Controls	-0.001	0.002	
	CS	-0.006	0.002	

This table presents results estimating the effects on births for family planning waiver policies implemented between 2001-2009, using data from the American Community Survey on 18-44 year old women for these years. Cols. 1 and 2 indicate the sample restriction and specification (see Sec. 8.2 for details). Eligible individuals include those under a household income threshold of 185% of the federal poverty line.

Table A4: Fertility: Overall average treatment effects on the treated for ACA Medicaid expansions (corresponding to Fig. 9b)

(1) Sample	(2) Specification	(3) Estimate	(4) Standard Error	(5) Observations
Full Sample	TWFE	-0.001	0.002	993904
	TWFE, Controls	-0.002	0.002	
	Gardner	0.000	0.002	
	Gardner, Controls	-0.001	0.002	
	CS	0.004	0.003	
Non-Parents	TWFE	0.001	0.003	571919
	TWFE, Controls	0.000	0.003	
	Gardner	0.002	0.004	
	Gardner, Controls	0.001	0.004	
	CS	0.004	0.005	
Parents	TWFE	-0.001	0.001	421985
	TWFE, Controls	-0.001	0.001	
	Gardner	-0.001	0.001	
	Gardner, Controls	-0.001	0.001	
	CS	0.006	0.002	
All Eligible	TWFE	-0.002	0.004	227747
	TWFE, Controls	-0.005	0.003	
	Gardner	-0.003	0.004	
	Gardner, Controls	-0.004	0.004	
	CS	0.015	0.007	
Eligible Non-Parents	TWFE	-0.002	0.006	121858
	TWFE, Controls	-0.003	0.005	
	Gardner	-0.002	0.006	
	Gardner, Controls	-0.003	0.005	
	CS	0.019	0.011	
Eligible Parents	TWFE	-0.002	0.003	105889
	TWFE, Controls	-0.002	0.003	
	Gardner	-0.002	0.003	
	Gardner, Controls	-0.002	0.003	
	CS	0.010	0.005	

This table presents results estimating the effects on births for ACA Medicaid expansions implemented between 2010-2019, using data from the American Community Survey on 18-44 year old women for these years. Cols. 1 and 2 indicate the sample restriction and specification (see Sec. 8.2 for details). Eligible individuals include those under a household income threshold of 138% of the federal poverty line.

B Replication of Kearney and Levine (2009)

The main analysis of fertility carried out by Kearney and Levine (2009) considers the effects of family planning waiver programs on births occurring between years 1990 to 2002. They use aggregated birth counts for a given year and state drawn from U.S. Vitality Statistics Natality data, divided by state-level population estimates from the U.S. Census Bureau for women, focusing on two age groups in turn: 15-19 year olds (representing teen births) and 20-44 year olds. As is typical in an aggregate fertility analysis, they take the log of the fertility rate as their dependent variable of interest.

They estimate a difference-in-differences model using two-way fixed effects:

$$Y_{st} = \delta(Treated_{st}) + \alpha_s + \gamma_t + X'_{st}\beta + \epsilon_{st}$$

where Y_{st} gives the log fertility rate for women in a state s in year t, β is the coefficient on whether or not women in said state are exposed to an income-based waiver expansion, α_s and γ_t are state and year fixed effects respectively, and ϵ_{st} is the error term. They cluster standard errors at the state level, and the equation is estimated separately for teens and non-teens. The vector X_{st} includes controls for

One additional important detail is that the authors weight all estimation by the state population of women, a choice which I will discuss in detail after presentation of the initial results.

Given developments regarding estimators for DID, I estimate the same results using the Gardner estimator (with their time-varying controls), the BJS estimator (without their time-varying controls) and the CS estimator (without controls) according to identical equations as given in Eqs. 3-7. For each estimator, I show results both with and without population weights included.

Table A5 presents the results for this exercise, with Panel (a) displaying results for the 15-19 year olds and Panel (b) for the 20-44 year olds. Col. 1 of Panel (a) therefore corresponds to the upper panel of Col. 1 of Table 3 i Kearney and Levine, while Col. 1 of Panel (b) corresponds to the upper panel of Col. 4 of the same table.

This point is worth discussing in some detail, as population weights turn out to be critical for the specific results estimated in Kearney and Levine (2009), and they are at the heart of the question about how to develop appropriate estimators for difference-in-differences models.

Weighting by population was once thought to be innocuous and perfectly normal in a setting analyzing aggregate state-level rates such as fertility. However, more recent consideration highlighting that the decision to incorporate population weights in this way needs to be carefully considered, and there are only certain appropriate reasons to motivate the choices (**?**). Moreover, weighting can actually have meaningful impacts on both point estimates, inference, and the interpretation of the causal estimate.

In estimating causal effects, **?** highlight three possible motivations for including weights: (1) to improve precision owing to heteroscedaticity, (2) to correct for endogenous sampling, and (3) to identify average population partial equilibrium effects in the presence of heterogeneity. The first issue cannot be immediately ruled out, but seems implausible: it would be more likely to arise if our fertility rate estimates were based on survey data, where rates calculated from smaller states (based small samples) would have substantially greater variance than those calculated based on samples from larger states. However, we are using administrative data for births, and population estimates

Table A5: Replication of Kearney and Levine (2009))

(a) 15-19 year olds

	TWFE		Gardner		BJS		CS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimate	-0.074	-0.021	-0.135	-0.043	-0.131	-0.028	-0.064	-0.026
	(0.031)	0.034	0.041	0.059	0.027	0.049	0.021	0.027
Pop. Weights	YES	NO	YES	NO	YES	NO	YES	NO
Observations	637	637	637	637	637	637	637	637

(b) 20-44 year olds

	TWFE		Gardner		BJS		CS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimate	-0.037	-0.012	-0.071	-0.030	-0.080	-0.019	-0.045	-0.017
	(0.011)	0.011	0.021	0.027	0.015	0.022	0.012	0.011
Pop. Weights	YES	NO	YES	NO	YES	NO	YES	NO
Observations	663	663	663	663	663	663	663	663

aggregated at quite a high level. If we were attempting to use much more fine-grained groupings, such as county-level single-year age groups, the population estimates might indeed become quite noisy for some smaller states. We can ignore the second motivation directly, since there is no sampling.

The third issue links directly to questions that have been discussed in this paper, namely the choice of weights for the causal estimand of interest. **?** highlight that in some settings, researchers might be interested in the *population average partial equilibrium treatment effect* in the context of heterogenous treatment effects. For an analysis at the state level, in practice this would mean we would find it desirable to put more weight on a larger state like California, because we want to characterize the overall average effect of a treatment on the country. In contrast, the intuition of the average treatment effect on the treated is to characterize the average effect on a treated state, in which we would want to put the same weights on the effects for California and New Mexico as treated states, even if they have heterogenous effects. One way of thinking about the motivation for doing one or the other is about whether we wish to think retrospectively or prospectively: suppose that we want to characterize historically how much a particular type of policy affected the country as a whole; this thinking might motivate weighting by population. Alternatively, suppose that we want to instead consider whether a particular type of policy should be implemented in other states in the future. In this case, it makes less sense to put large weights on California: once California has been treated, there is no other California in the country. While there is no guarantee of course that putting equal weights on the treated states provides a representative average for states that may later implement treatment (this is of course a core point of the Lucas critique), it is certainly less arbitrary than putting a weight on California's effect equal to precisely the proportion of California's population relative to the treated states for the particular time period studied.