The Rise of the Modern Hospital and Early-Life Health: Evidence from the Hill-Burton Act

Jason Fletcher* Owen Thompson† Karin Wu‡

January 30, 2026

Abstract

The U.S. hospital sector expanded rapidly in the 1950s and 1960s, largely due to construction subsidies provided under federal legislation known as the Hill-Burton Act. This paper examines the impact of Hill-Burton grants on maternity care access and infant health. We find that grants for public hospitals significantly reduced out-of-hospital births and infant mortality, particularly among non-white populations. In contrast, grants for private non-profit hospitals had no measurable effects on out-of-hospital births or infant mortality.

*University of Wisconsin-Madison and NBER.
†Williams College and NBER.
‡National Taiwan University

1 Introduction

The capacity of the US hospital sector doubled between 1948 and 1975, when the number of non-federal general hospital beds grew from 451,000 to 932,000.[1] Much of this growth occurred with the assistance of federal subsidies provided under the Hospital Survey and Construction Act of 1946, commonly known as the Hill-Burton Act after its legislative sponsors. This program provided federal matching funds for the construction, expansion and renovation of public or private non-profit hospital facilities, with a focus on bringing hospital access to medically underserved communities. From its inception in 1946 through the early 1970s the Hill-Burton program subsidized the creation of over 300,000 general hospital beds, more than half of the total new beds added over this period.

The rapid expansion of hospital capacity in the latter half of the 20th Century coincided with large reductions in out-of-hospital births and infant mortality. As of 1948 14% of births took place outside of hospitals, but the out-of-hospital delivery rate had fallen to 1% by 1975. The reduction in out-of-hospital births in this period was even more dramatic for non-white (primarily African American) mothers, for whom the share of births occurring outside of hospitals fell from 67% to 3%. Infant mortality was likewise cut in half over this same period, falling from 32 infant deaths per 1,000 live births in 1948 to 16 such deaths in 1975.

These developments constituted a rapid and fundamental transformation of the national health care system, and established hospitals as the dominant point of contact for labor, delivery and newborn health care provision.

This paper uses variation from hospital construction and expansion projects under the Hill-Burton program to evaluate how the advent of the modern hospital sector affected access to maternity care and infant mortality. Combining archival records on Hill-Burton projects with national Vital Statistics data going back to 1946, we estimate difference-in-difference and event study specifications that track county-level changes in out-of-hospital births and infant mortality in the years surrounding a local Hill-Burton grant, primarily using counties that never received a Hill-Burton grant as controls. We separately evaluate the effects of Hill-Burton grants for public (typically county operated) hospitals versus private non-profit facilities, and wherever possible we disaggregate our analyses by race.

[1] All of the statistics cited in this introduction are the authors' calculations using data from Vital Statistics records, American Hospital Association data, and Hill-Burton registrar data described in detail below.

Our work contributes to two related literatures.

First, we provide a novel evaluation of the Hill-Burton program itself. As noted, the Hill-Burton program was large in scale, contributing to the creation of hundreds of hospital facilities and hundreds of thousands of hospital beds. The program was also an early and ambitious expansion of the federal government's role in the provision of health and human services, which occurred approximately two decades prior to the introduction of Medicare, Medicaid, and the various War on Poverty era programs that had significant health components.

Given the magnitude and historical importance of the Hill-Burton program, it has been the topic of surprisingly little research. Early studies by Brinker and Walker [1962] and Lave and Lave [1974] used state level aggregates to study the association between Hill-Burton funding and hospital capacity, and a more modern paper by Chung et al. [2017] also studies the effect of the program on hospital capacity, as well as on ownership structure and the spatial distribution of hospital beds. These studies find that Hill-Burton grants had large and lasting effects on the number of hospital beds per-capita, as well as the structure of the US hospital industry. However, no published studies directly evaluate how Hill-Burton grants impacted actual health outcomes, a significant gap in the literature given that improving health was the ultimate objective of the program.[2]

Second, we contribute to the literature on the determinants and effects of early childhood health. The literature on this topic is quite large and active, with excellent reviews provided by Almond and Currie [2011] and Almond et al. [2018]. The specific strand of this expansive literature that is most closely related to the current study evaluates the effects of access to hospitals (or other forms of formalized health care) at the time of birth, and includes studies by Almond et al. [2010], Barreca et al. [2011], Lazuka [2018], Chyn et al. [2021], Fischer et al. [2021], Anderson et al. [2024] and Hollingsworth et al. [2024]. We discuss how our findings relate to Hollingsworth et al. [2024] and Anderson et al. [2024] in greater detail in the conclusion.

[2]A recent working paper by McCloy [2025] uses an IV strategy to study the effect of Hill-Burton grants on a wide variety of outcomes, and reports some findings for all-age and infant mortality. However, the estimates for these health outcomes under McCloy [2025]'s research design are inconclusive, and their study focuses primarily on how Hill-Burton grants affected hospital capacity and quality, physician labor supply, and innovation.

2 The Hill-Burton Program

The Hill-Burton program was first created under the Hospital Survey and Construction Act of 1946, was significantly expanded in subsequent 1949 legislation, and continued to award project grants through the early 1970s. The primary purpose of the legislation was to increase national hospital capacity, and to do so with a focus on improving access to hospital care in low income, rural, and otherwise medically underserved areas.

The scale of the Hill-Burton program was quantitatively large. Federal expenditures were around $75 million annually (approximately $1.3 billion in 2024 dollars), which was doubled to $150 million annually ($2.6 billion in 2024 dollars) under the 1949 expansion. As described momentarily, these federal expenditures were paired with matching local funds, and the Hill-Burton projects studied in the current paper included approximately $135 billion in total expenditures (in 2024 dollars) that were used to provide over 300,000 beds through the construction or expansion of over 5,500 general hospitals throughout the US.

The Hill-Burton program was an early example of the federal government and the states jointly funding and administering a social welfare program, a policy practice which became much more widespread in the 1960s and 1970s. The federal government allocated aggregate funds to states using a formula specified in the legislation, which provided more per-capita funding to states with lower per-capita income levels, reflecting the legislation's stated goal of improving hospital access in underserved areas.[3] The federal government also set broad parameters on how the funds could be used, most importantly requiring non-profit or public status for recipient facilities, with for-profit hospitals being ineligible. At the state level, agencies conducted comprehensive surveys of existing hospital capacity in each county, and then created "plans" that prioritized the counties within each state for receiving Hill-Burton projects based on their existing hospital capacity relative to need, with some additional preference given to rural counties. These state plans needed to be approved by the Surgeon General, after which the federal funds became available and organizations proposing specific projects would apply to the state for funding. States were to approve projects in an order that broadly complied with the county level prioritization laid out in their federally approved plans, and the federal government gave the final approval for each individual project.

In addition to this joint administrative structure, program costs were divided between the federal government and the states. The original 1946 legislation required that the local

[3]See Lave and Lave [1974] for illustrative examples of the Hill-Burton funding formula.

sponsor provide at least 2/3 of the project cost, with the federal contribution capped at 1/3, while the 1949 revision increased the federal contribution to between 1/3 and 2/3 of total project costs. For the projects studied in the current paper, 28% of total expenditures came from federal sources.

A final federal administrative requirement, critical for current purposes, is that the Hill-Burton Act contained a "separate but equal" provision which stated that the "hospital or addition will be made available to all persons...without discrimination on account of race, creed or color, but an exception will be made in cases where separate hospital facilities are provided for separate population groups if the plan makes equitable provision on the basis of need for facilities and services of like quality for each such group [Largent, 2018]."[4] This separate but equal clause was unsurprisingly not fully enforced, and the Hill-Burton facilities serving Black patient populations were almost certainly provided with fewer resources, on average, than those serving white populations. However, there are historical accounts of at least some meaningful enforcement of the requirement of equal quality facilities (see Smith [2016], Chapter 2) and the Hill-Burton sponsored facilities that were accessible to southern Black patients were generally described as being of good quality, with proponents of full integration sometimes (critically) referring to the Black wards of segregated southern Hill-Burton facilities as "Deluxe Jim Crow" hospitals [Thomas, 2011].

There are also ex-ante reasons to believe that public hospitals could have had different impacts than private non-profits, given their history of accessibility and lower patient costs. The historical record suggests that any distinct role of public hospitals was likely especially pronounced in the non-white patient populations where we find that Hill-Burton's effects were strongly concentrated. For instance Anderson et al. [2024] quote from a 1972 Government Accountability Office report which observed that a "disproportionately large share of minority patients received their health care at government-owned hospitals" while Smith [2016] calculates that although there were approximately 50 hospitals in Chicago as of 1951, over 80% of births to Black mothers occurred at the publicly operated Cook County Hospital. Indeed, many listings of historically "Black hospitals" (for example Jones and Rice [1994]) contain unique Black operated southern non-profits like Meharry Medical Center in

[4]Notably, this is the only major instance of racially segregated facilities being eligible for federal funds in the 20th Century. Subsequent federal programs, most importantly Medicare and the Elementary and Secondary Education Act, actively prohibited funding from going to segregated facilities and were important drivers of integration in hospitals and schools. The separate but equal provision of the Hill-Burton Act was deemed unconstitutional in the 1963 Supreme Court case *Simkins v. Cone.*

Nashville or Saint Agnes Hospital in Wake County North Carolina, but also list public facilities both in and out of the South, such as Kings County Hospital Center (Brooklyn), Charity Hospital of Louisiana (New Orleans), and Memphis Regional Medical Center. Given these considerations, it is reasonable to expect that any differential effects of public and private facilities may have been especially pronounced among Black patients.

3 Data

3.1 Hill-Burton Data

Data on Hill-Burton projects comes from the Hill-Burton Project Registrar, a publication of the US Department of Health, Education and Welfare. These project registrars are available from the Hill-Burton program's first grants in 1947 through 1971, and were transcribed and generously shared by Heidi Williams. The Hill-Burton project registrar data is at the *grant* level. For each Hill-Burton sponsored project, the registrar lists the name of the hospital as well as the county and state where it was located, the year in which the initial project application was approved, the type of capital improvements it funded, the ownership structure of the facility, and the total number of beds and dollars of funding provided for the project.

We focus on general hospitals, since these are the facilities that could reasonably be expected to affect the studied outcomes, and exclude nursing homes, tuberculosis hospitals, mental hospitals and outpatient clinics. We also exclude a small number of Hill-Burton grants that were exclusively for remodels and did not involve new construction, additions or replacements of existing facilities, as well as projects that added zero additional hospital beds, which were typically grants for hospital laboratories, equipment and similar items. After restricting the universe of Hill-Burton grants in these ways, we collapse the registrar data to the county-year level. We then define an absorbing binary treatment measure using the first Hill-Burton grant in each county that meets our criteria, and do so separately for Hill-Burton grants awarded to private non-profit hospitals versus public hospitals.

Since treatment status is ambiguous for counties that received Hill-Burton grants *only* for non-general hospitals or for projects that were strictly remodels or created no new beds, we exclude these counties from our analysis entirely, rather than coding them as untreated and retaining them as controls. Counties that received multiple grants are included and their

treatment date is defined using the first project that meets our sample criteria. Continuous treatment measures that incorporate multiple grants where applicable are discussed below and reported in Appendix 5.

3.2 Vital Statistics Data

Our primary outcomes of interest are the out-of-hospital birth rate (OHBR) measured in percentage points, and the infant mortality rate (IMR) measured as the number of infant deaths per-1000 live births. Data on these outcomes is drawn from US Vital Statistics records, which are available nationwide at the county-year level beginning in 1946, one year prior to the first Hill-Burton grants being awarded.[5]

We view the OHBR as being of interest for two distinct reasons. First, increased access to hospital based labor and delivery services was an explicit objective of the Hill-Burton program and the most obvious mechanism through which Hill-Burton grants may have impacted infant mortality, making the effect of Hill-Burton projects on the OHBR an important "first stage" relationship. Second, we view the OHBR as a proxy for *overall* access to medical care, and one that was collected administratively in an historical period when few if any reliable measures of access to medical care are available. Any reductions in the OHBR from Hill-Burton grants can therefore be viewed as suggestive evidence that health care access and utilization may have also improved for other types of hospital care.

The IMR is of course itself an important health outcome, and is also widely used as a proxy for general public health levels and health system efficacy. Hill-Burton funds were often used to construct separate labor and delivery wings, allowing separation by disease process such that mothers of newborn infants were not directly exposed to contagious disease, and our entire study period occurs after the introduction of sulfa drugs in the late 1930s, which is when hospitals were first able to provide tangible maternal and child health benefits [Thomasson and Treber, 2008].

Both outcomes are defined with respect to county of *residence* rather than *occurrence*, since traveling to receive hospital care is likely common and could make occurrence based measures

[5] We use the Vital Statistics records that were compiled and audited by Bailey [2016] and generously made available as ICPSR data set 36603. While US Vital Statistics suffered from a substantial under-registration of births through the early 20th Century [Eriksson et al., 2018], by 1950 98.1% of US births were registered [U.S. Department of Health, Education, and Welfare, 1960]. We exclude Alaska and Hawaii because they gained statehood during the study period, and Virginia because large numbers of county boundaries were changed in Virginia during the study period.

misleading. Approximately 2% of county-year observations had zero live births, and these observations are excluded. Approximately 7% of county-year observations had no infant deaths while approximately 20% of county-years had no out-of-hospital births, and these observations are retained and coded as zeros in our preferred specifications.

Wherever possible we report results disaggregated by race, but there are three limitations to the racial disaggregations available in the Vital Statistics data from this period. First, while the OHBR is available by race for the full sample period, racial disaggregations of the IMR are not available for any counties until 1959. Second, racial disaggregations of either outcome are only available in county-years where more than 10% of the population was non-white. As a result, race-specific specifications can only be estimated on a balanced panel within the approximately 1/3 of counties where 10% or more of the population was non-white throughout our sample period, although we note that these 1/3 of counties do contain approximately 50% of the total population and approximately 65% of the total non-white population over our study period. Finally, racial disaggregations are limited to white versus non-white, rather than more detailed race and ethnicity categorizations. For accuracy we use the term "non-white" in describing our results, but we do note that African Americans were the dominant non-white racial group in most counties during this period.

4 Methodology

4.1 Estimation

We estimate two-way fixed effects specifications that regress the OHBR and IMR onto measures of Hill-Burton treatment as well as county and calendar year fixed effects. Our preferred models use binary treatment measures and estimate "intent to treat" effects of the program, employing both a static difference-in-difference specification and a dynamic event study specification:

$$y_{ct} = \beta HillBurton_{ct}^{type} + \gamma_c + \lambda_t + \varepsilon_{ct}. \tag{1}$$

$$y_{ct} = \sum_{et=-4, et\neq 0}^{et=16} \theta_{et} 1[t - HB_c^{type} = et] + \gamma_c + \lambda_t + \varepsilon_{ct}. \tag{2}$$

In both of these equations y_{ct} denotes the OHBR or IMR in county c and calendar year t, while γ_c and λ_t are county and year fixed effects. In the first equation treatment is measured with a binary indicator of there being a Hill-Burton project of a given type (public or private) in county c and year t, which is denoted $HillBurton_{ct}^{type}$. In the second equation, HB_c^{type} denotes the year that county c first received a Hill-Burton grant of a given type, such that treatment is measured with a set of "event-time" dummies indicating whether it has been et years since the first Hill-Burton project was initiated in county c.

OLS estimation of two-way fixed effects models like Equations 1 and 2 in settings with staggered treatment adoption are known to be problematic [Roth et al., 2023, Baker et al., 2025], and our preferred estimation technique is the two stage difference-in-difference method proposed by Gardner et al. [2025], which is very closely related to the estimator developed in Borusyak et al. [2024]. The two-stage approach first estimates the unit and time fixed effects *using only untreated observations*, which in the current application consist of 569 counties that were never awarded any Hill-Burton grant, as well as the pre-treatment years of counties that eventually received a Hill-Burton grant. The outcomes are then residualized of these fixed effects, and these residuals are regressed onto the static and dynamic treatment measures. We use standard errors with unit level clustering derived by Gardner et al. [2025]. This two-stage process is a practical estimation approach for the current application, since it allows us to flexibly and transparently estimate specifications with different treatment measures and sample structures using a consistent method. In Appendix 3 we report and discuss results from various alternative estimators, including OLS and those proposed by Sun and Abraham [2021] and Callaway and Sant'Anna [2021].

Because we distinctly estimate the impacts of public and private Hill-Burton treatments, our estimation sample can be classified into four mutually exclusive groups: Counties that received only public Hill-Burton projects (37% of counties), counties that received only private Hill-Burton projects (29% of counties), never-treated counties that received neither type of project (28% of counties), and counties that received both public and private projects (6% of counties). Since counties receiving both types of projects are rare and their treatment status is ambiguous, we exclude them from our main analyses. The specifications for both public and private projects are then estimated separately with samples that include a common set of never-treated controls, and the identifying assumptions and underlying comparisons for estimating Equations 1 and 2 with this data structure are discussed momentarily.

Our preferred models are weighted by the total number of births in each county-year. These

weighted estimates approximate the treatment effect for the average *birth* that would be obtained if birth-level microdata were available, whereas unweighted estimates approximate the treatment effect for the average *county* [Solon et al., 2015, Baker et al., 2025]. We report and discuss unweighted estimates in Appendix 4. Our preferred models also do not include time varying controls, both because there are very few relevant characteristics measured at the county level on an annual basis going back to 1946, and also because many of the characteristics that would be natural to attempt to control for, such as county level incomes or racial compositions, could very plausibly be affected by the treatment itself and are therefore arguably best thought of as mechanisms rather than true covariates. We therefore find it most informative and transparent to report unconditional reduced form treatment effect estimates.

We restrict our estimation samples to be balanced in *event-time*, so that each event-time coefficient is estimated using a common set of counties. Because Hill-Burton projects began in 1947 and the Vital Statistics data is available beginning in 1946, the main tradeoff we face is between including a sufficient number of pre-program years to evaluate pre-trends and excluding counties that received Hill-Burton projects in the initial years of the program. In our baseline models we opt to include four pre-treatment years and 16 post-treatment years, which allows us to use all counties with Hill-Burton projects occurring in 1950 or later.[6] In Appendix 2 we report models that use alternative event-time ranges and that do not require the sample to be balanced in event-time. For precision and visual clarity we group event-time dummies into two year bins. We omit event-time 0 (the year the grant was approved) as the reference category, which we prefer to the somewhat more common choice of event-time -1 because construction lags make it unlikely that the actual relevant services funded by the grants were available in the year they were approved, making event-time 0 the last untreated year.

4.2 Identifying Assumptions

In the taxonomy of Miller [2023], our data has a "hybrid" structure: There are both treated and never-treated units, and among the treated units there is variable treatment timing. The identifying variation therefore comes both from comparisons of treated observations

[6]More precisely, we estimate all possible event-time coefficients, but restrict the sample to the set of all counties with treatment years such that event-times -4 through +16 are observed and only report event-time coefficients in the balanced range. The results are extremely similar if we fully restrict the sample to observations with event-times in the balanced range.

versus never-treated counties, and from "timing based" comparisons within the set of treated counties.[7] Given this data structure, causal identification requires two related parallel trends assumptions. First, the trends in counties that never received a Hill-Burton project must accurately characterize the counterfactual outcome trends that would have happened in treated counties, absent treatment. Second, the *timing* of projects in treated counties must be unrelated to counterfactual outcome trends, such that outcome trends in not-yet-treated counties accurately characterize the outcome trends that would have happened in treated counties, absent treatment.

Our most important method of evaluating whether these assumptions hold is to inspect the pre-trends that are built into our event study models. These test whether treated and untreated counties were trending similarly in untreated periods, and provide strong suggestive evidence on both of the (inherently untestable) parallel trends assumptions.

Additionally, Table 1 presents "balance tests" that compare the means of various pre-treatment demographic and socioeconomic characteristics, as well as baseline values of our outcomes, in the sets of never-treated counties, publicly-treated counties, and privately-treated counties that are used in our primary models.[8] We report these comparisons separately for our all-race sample (Columns 1-3) and for the subsets of counties used in our main specifications for race specific out-of-hospital birth rates (Columns 4-6) and race-specific infant mortality rates (Columns 7-9).

A clear pattern in Table 1 is that never-treated counties are reasonably similar to counties receiving Hill-Burton grants for *public* hospitals, but that counties receiving a Hill-Burton grant for *private* hospital projects are much more populous and had substantially higher income and education levels than publicly-treated or never-treated counties. Additionally, privately-treated counties had *much* lower baseline OHBRs than the other groups of counties. These differences may be particularly problematic, since they could lead the OHBR to fall relatively slowly in privately-treated counties due to a floor effect around zero, while the OHBR had much more "room to fall" in the never-treated counties that help define the counterfactual OHBR trend. We return to this point below when we discuss our results for

[7] We note that there is a third source of identifying variation, treatment *intensity*, which arises because Hill-Burton projects had highly variable sizes and also because some counties received multiple projects, and we report estimates that measure treatment using the cumulative number of Hill-Burton sponsored hospital beds per-capita in Appendix 5.

[8] Socioeconomic and demographic variables are drawn from the 1950 Census. Note that while the first Hill-Burton grants were issued in 1946, our estimation sample uses only treatments occurring in 1950 or later to maintain a balanced panel, such that 1950 is the latest "pre-treatment" year for the counties we study.

private Hill-Burton treatments.

The second parallel trends assumption, regarding early versus late treated counties, also faces potential violations, especially given the within-state project prioritization process described above. As an initial assessment of potential systematic differences in the timing of projects, in Table A1 of Appendix 1 we regress the year of treatment onto the pre-treatment county characteristics from Table 1 within the sets of publicly and privately treated counties. For publicly treated counties, we find a significant association between treatment timing and baseline total population, which may reflect the noted targeting of medically underserved counties, but year of treatment is otherwise uncorrelated with county characteristics, including baseline values of the OHBR and IMR. For privately treated counties there is again a significant correlation between year of treatment and baseline total population, and also significant correlations between treatment timing and baseline education and income levels. We report these associations for completeness, but emphasize that the timing-based parallel trends assumption does *not* require that treatment timing is as good as randomly assigned, but rather that the *outcomes* in not-yet-treated counties accurately characterize counterfactual *outcomes* in treated counties. We believe the best way to assess this assumption is by observing the pre-trends in our event study specifications.

5 Estimates for Public Hill-Burton Projects

We begin by reporting the estimated effects of Hill-Burton grants for projects at publicly operated hospitals. Static estimates of our baseline models are reported in the upper panel of Table 2, while event study estimates are reported in Figure 1.

Column 1 of Table 2 estimates that the presence of a public Hill-Burton project reduced the all-race OHBR by 3.5 percentage points, while Column 4 estimates a reduction in the all-race IMR of 1.8 infant deaths per 1,000 live births. Both estimates are statistically significant at conventional levels. The corresponding event study diagrams in Figure 1 reassuringly show no statistically significant or visually apparent pre-trends in the 4 years leading up to the awarding of a grant to a public hospital, and then display a clear trend break around the time of treatment. At the end of our window, 16 years after the awarding of a grant for a public hospital, the all-race OHBR has declined by 4 percentage points while the all-race IMR has declined by 2 infant deaths per 1,000 live births.

The remaining columns in the upper panel of Table 2 and the bottom panels of Figure 1 report

race-specific estimates for public Hill-Burton projects. Because of the discussed limitations in the availability of race-specific data, the race-specific OHBR effects are estimated using a set 278 publicly-treated counties (and 85 never-treated controls), while the race-specific IMR effects are estimated using data from 1959 and later and a set of 57 publicly-treated counties (and the same 85 never-treated controls).[9] The estimates within these samples suggest that the benefits of public Hill-Burton projects were concentrated almost entirely in non-white populations.

Specifically, the event study estimates in the bottom panel of Figure 1 exhibit no statistically significant or visually apparent pre-trends for either outcome or racial group, consistent with the parallel trends assumptions holding in these sample for both racial groups. There are then large, visually clear declines following treatment in both outcomes within non-white populations. By the end of the event-time window, the non-white OHBR has declined by 20 percentage points and the non-white IMR has declined by over 5 infant deaths per 1,000 live births. The static models in Table 2 also show qualitatively large and statistically significant declines in both outcome among non-white mothers and infants. In contrast, the estimated effects for the same set of Hill-Burton projects on the OHBR and IMR in white populations are nominally negative, but are statistically insignificant and practically close to zero.

These estimates imply a substantial closing of racial gaps in the studied outcomes in counties receiving a public Hill-Burton grant. When racial *gaps* are used as the dependent variable in our baseline static specification (not shown), we estimate that public treatments reduced the racial gap in the OHBR by 16 percentage points and reduced the racial gap in the IMR by 4 percentage points. These respectively represent reductions of 30% and 16% from the pre-treatment values of the outcome gaps.

We do caution that these estimates were generated within the subsets of counties and years where racial disaggregations were available, and are not necessarily equal to what would be obtained using national data over the full study period. Nonetheless, the race-specific results in Table 2 and Figure 1 suggest that the pooled results, which included almost all counties in the US and 40 years of outcome data, likely reflect a weighted average of very large treatment effects among non-white (primarily Black) mothers and infants and null effects in white populations.

[9] In Appendix 2 we show that the race-specific IMR results are similar if we do not require our data to be balanced in event-time, which increases the number of publicly-treated counties from 57 to 407, although many of these treatments happened much earlier and therefore only contribute to identifying very late event-time coefficients.

The results for public treatments are robust. In the appendices we report results from estimating our main models via OLS or alternative estimators like Sun and Abraham [2021] and Callaway and Sant'Anna [2021], results that vary the length of the pre-period or use samples that are not balanced in event-time, results that use continuous treatment measures like cumulative hospital beds per-capita, results that use the natural log of the outcomes, and unweighted results. In a large majority of cases our findings for public Hill-Burton treatments are not qualitatively changed under these alternative approaches. In cases where there are meaningful differences, we believe those differences are best understood as treatment effect heterogeneity or as situations where the alternative specifications target a different parameter, and these interpretational points are discussed in detail in the appendices.

6 Estimates for Private Hill-Burton Projects

The estimated effects of Hill-Burton grants for projects at privately operated non-profit hospitals are reported in the bottom panel of Table 2 and in Figure 2.

The covariate imbalance between privately-treated and never-treated counties in Table 1 raised concerns that the parallel trends assumptions may not hold for private treatments, especially with respect to the OHBR. These concerns are borne out in the event study results for the all-race OHBR, which clearly exhibit a strong positive pre-trend. Given the overall environment of declining out-of-hospital births, this indicates that the OHBR was falling faster in never-treated counties than in privately-treated counties *even prior to* the awarding of Hill-Burton grants to private hospitals. These positive trends continue into the post-treatment period, and there are large positive point estimates for the post-treatment event-times in Figure 2 and an estimated static treatment effect of positive 7.4 in Table 2.[10]

We find it implausible that constructing or improving private non-profit hospitals would lead to large increases in out-of-hospital births, and attribute these unexpected signs to parallel trends violations. More specifically, the relatively high baseline OHBR of 20.5 in never-treated counties (shown in Table 1) means that there is mechanically greater scope for OHBR reductions in those controls than in privately treated counties, where the baseline OHBR was just 10.2. As a result, the OHBR trends in never-treated counties likely overstate the

[10]The downward spike at the reference period, event-time=0, resembles that in Roth [2024] when using the estimator from Borusyak et al. [2024] in a simulated data set with common trends violations. Appendix 3 discusses why this appears as a discontinuous jump at event-time = 0 rather than a smooth continuous pre-trend, and also discusses more generally the validity of visually evaluating pre-trends with our estimator.

counterfactual OHBR reductions for privately-treated counties, leading to positive treatment effect estimates. Given these issues, our preferred interpretation of the all-race OHBR results for private treatments is simply that the requirements of our research design are not met and that we are unable to draw any conclusions about this effect.[11]

The remainder of Table 2 and Figure 2 report the estimated effects of private Hill-Burton treatments on the all-race IMR and on the race-specific OHBR and IMR. In contrast to the all-race OHBR, the imbalance in baseline levels of these outcomes are more modest and the event studies in Figure 2 do not display visually clear pre-trends, such that our research design may be able to provide credible estimates. The estimated treatment effects, however, are small and statistically insignificant. A partial exception is that there are non-trivial negative point estimates for the non-white IMR, but these estimates are imprecise and do not approach conventional levels of statistical significance. We interpret these findings as suggestive evidence that, in contrast to grants going to public facilities, Hill-Burton grants for private hospitals did not have large effects on the IMR in any population or on the race-specific OHBR.

7 Conclusion

We close with two observations about how our findings may inform important open research questions.

First, the recent study by Hollingsworth et al. [2024] evaluated a philanthropic hospital construction program sponsored by the Duke Endowment, which operated primarily in North Carolina beginning in the 1920s. The Duke Endowment program served as an influential state-level model for the Hill-Burton Act, and contrasting our findings with those of Hollingsworth et al. [2024] provides an important case study of how effective model programs can scale up, a question of perennial interest across a wide variety of policy domains. Hollingsworth et al. [2024] estimated that county level exposure to a Duke Endowment hospital grant reduced the all-race IMR by 7.5% and the nonw-white IMR by 13.6%. When we convert our own estimates into percentages using the IMR in the last pre-treatment year as a baseline, the results from Table 2 indicate that exposure to public Hill-Burton projects reduced the overall IMR by 5.5% and the Black IMR by 13.1%. Taken at face value, this

[11] We have considered pursuing matching strategies to construct a more suitable set of never-treated controls, but there is very little common support in the observable characteristics of privately-treated and never-treated counties.

provides suggestive evidence that the benefits of the Duke Endowment program did indeed scale nationally.

Of course, there were many differences between the Duke and Hill-Burton programs that make these estimates not directly comparable, including different baseline public health levels and technology environments stemming from the time periods and geographic scopes of the two programs. And perhaps the most important difference is that we are referring to our estimates for public hospitals, while the Hollingsworth et al. [2024] estimates come from jointly evaluating public and private non-profit hospitals. However, this itself may provide an important historical lesson about the role of coordination between federal and local governments in scaling up model programs. Indeed the overall administrative structure of the Hill-Burton program, with both the federal government and the states contributing funding and local governments having jurisdiction over many aspects of the actual service provision, foreshadowed the structure of many subsequent policy initiatives including Medicaid, Head Start, and Title I education financing, among many others. It may be that this type of administrative structure is particularly effective for scale up.

Second, recent work by Anderson et al. [2024] found that southern hospital desegregation, driven by the 1966 introduction of Medicare, did not materially reduce Black infant mortality. This finding contrasts with influential earlier work on the same topic by Almond et al. [2006]. While we do not directly study southern hospital desegregation, our findings are relevant to this debate since they indicate that hospital access, in general, significantly decreased Black infant mortality in the mid 20th Century US, including in the pre-desegregation period. One possible implication of our own results is that the null findings of Anderson et al. [2024] may be because publicly operated Hill-Burton hospitals were already providing meaningful health care access to Black mothers, such that the marginal health impact of formal desegregation was relatively small.

References

D. Almond and J. Currie. Killing me softly: The fetal origins hypothesis. *Journal of economic perspectives*, 25(3):153–172, 2011.

D. Almond, K. Y. Chay, and M. Greenstone. Civil rights, the war on poverty, and black-white convergence in infant mortality in the rural south and mississippi. 2006.

D. Almond, J. J. Doyle Jr, A. E. Kowalski, and H. Williams. Estimating marginal returns to medical care: Evidence from at-risk newborns. *The quarterly journal of economics*, 125 (2):591–634, 2010.

D. Almond, J. Currie, and V. Duque. Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature*, 56(4):1360–1446, 2018.

D. M. Anderson, K. K. Charles, and D. I. Rees. Imposing policy on reluctant actors: The hospital desegregation campaign and black postneonatal mortality in the deep south. *Review of Economics and Statistics*, pages 1–46, 2024.

M. Bailey. *US county-level natality and mortality data, 1915-2007*. 2016.

A. Baker, B. Callaway, S. Cunningham, A. Goodman-Bacon, and P. H. Sant'Anna. Difference-in-differences designs: A practitioner's guide. *arXiv preprint arXiv:2503.13323*, 2025.

A. I. Barreca, M. Guldi, J. M. Lindo, and G. R. Waddell. Saving babies? revisiting the effect of very low birth weight classification. *The quarterly journal of economics*, 126(4): 2117–2123, 2011.

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285, 2024.

P. A. Brinker and B. Walker. The hill-burton act: 1948-1954. *The Review of Economics and Statistics*, pages 208–212, 1962.

B. Callaway and P. H. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230, 2021.

A. P. Chung, M. Gaynor, and S. Richards-Shubik. Subsidies and structure: the lasting impact of the hill-burton program on the hospital industry. *Review of Economics and Statistics*, 99(5):926–943, 2017.

E. Chyn, S. Gold, and J. Hastings. The returns to early-life interventions for very low birth weight children. *Journal of Health Economics*, 75:102400, 2021.

K. Eriksson, G. T. Niemesh, and M. Thomasson. Revising infant mortality rates for the early twentieth century united states. *Demography*, 55:2001–2024, 2018.

M. Fischer, M. Karlsson, and N. Prodromidis. Long-term effects of hospital deliveries. Technical report, CINCH Series, 2021.

J. Gardner, N. Thakral, L. T. Tô, and L. Yap. Two-stage differences in differences. 2025.

A. Hollingsworth, K. Karbownik, M. A. Thomasson, and A. Wray. The gift of a lifetime: The hospital, modern medicine, and mortality. *American Economic Review*, 114(7):2201–2238, 2024.

W. Jones and M. Rice. *Public policy and the black hospital: From slavery to segregation to integration*. Bloomsbury Publishing USA, 1994.

E. A. Largent. Public health, racism, and the lasting impact of hospital segregation. *Public Health Reports*, 133(6):715–720, 2018.

J. R. Lave and L. B. Lave. The hospital construction act: An evaluation of the hill-burton program. 1974.

V. Lazuka. The long-term health benefits of receiving treatment from qualified midwives at birth. *Journal of Development Economics*, 133:415–433, 2018.

S. McCloy. Hospital expansion and health outcomes: Evidence from the hill-burton program. *Available at SSRN 4637931*, 2025.

D. L. Miller. An introductory guide to event study models. *Journal of Economic Perspectives*, 37(2):203–230, 2023.

J. Roth. Interpreting event-studies from recent difference-in-differences methods. *arXiv preprint arXiv:2401.12309*, 2024.

J. Roth, P. H. Sant'Anna, A. Bilinski, and J. Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023.

D. B. Smith. *The power to heal: civil rights, medicare, and the struggle to transform America's health care system*. Vanderbilt University Press, 2016.

G. Solon, S. J. Haider, and J. M. Wooldridge. What are we weighting for? *Journal of Human resources*, 50(2):301–316, 2015.

L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, 225(2):175–199, 2021.

K. K. Thomas. *Deluxe Jim Crow: civil rights and American health policy, 1935-1954*. University of Georgia Press, 2011.

M. A. Thomasson and J. Treber. From home to hospital: The evolution of childbirth in the united states, 1928–1940. *Explorations in Economic history*, 45(1):76–99, 2008.

U.S. Department of Health, Education, and Welfare. *Vital Statistics of the United States, Volume I*. U.S. Government Printing Office, Washington, DC, 1960.

Table 1: Baseline Means by Treatment Status

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All-Race Sample			*Race-Specific OHBR Sample*			*Race-Specific IMR Sample*		
	Never Treated	Publicly Treated	Privately Treated	Never Treated	Publicly Treated	Privately Treated	Never Treated	Publicly Treated	Privately Treated
Total Population	14,267	23,149	65,862	33,445	31,078	86,349	33,444	28,712	52,004
Median Education	8.7	8.5	9.2	7.6	7.5	8.2	7.6	7.4	7.5
Median Income	2,038	2,026	2,698	1,535	1,571	2,069	1,535	1,526	1,742
Percent Black	6.6	14	5.1	35.9	33.7	28.8	35.9	35.0	35.3
Infant Mortality Rate	30.5	32.7	29.8	40.4	36.4	38.5	40.4	37.7	41.1
Out of Hospital Birth Rate	20.5	25	10.2	41.9	41.2	29.6	41.9	40.7	35.9
Observations (Counties)	569	747	590	85	278	67	85	57	15

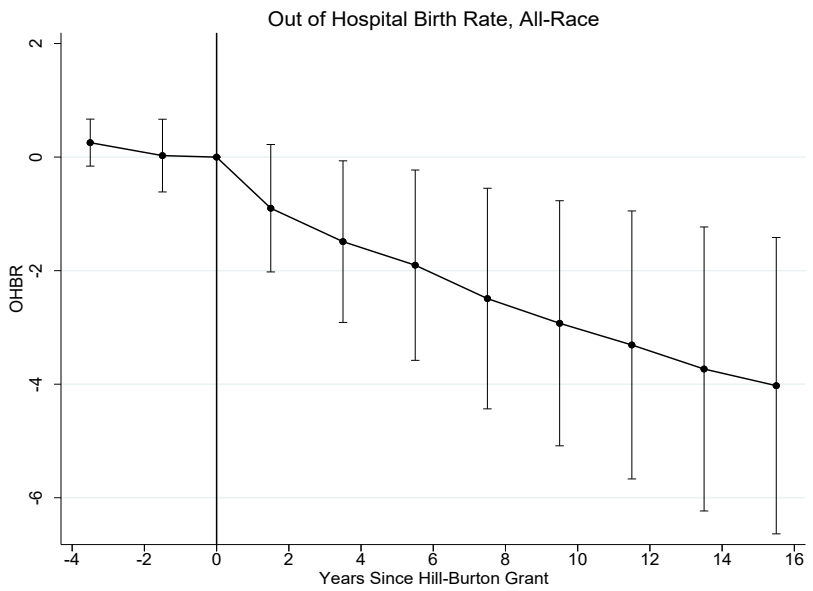
Notes: Reported means are all for calendar year 1950. Total population, percent Black, median income and median education are drawn from the City and County Data Books of the 1950 Census, contained in ICPSR Data Set 07736. The Infant Mortality Rate and Out of Hospital Birth Rate are from the Vital Statistics data described in the text. All of the reported means pool racial groups, and the "race-specific" headings in Columns 4-9 indicate that the samples are restricted to the subset of counties for which race-specific values of that outcome are available in the data.

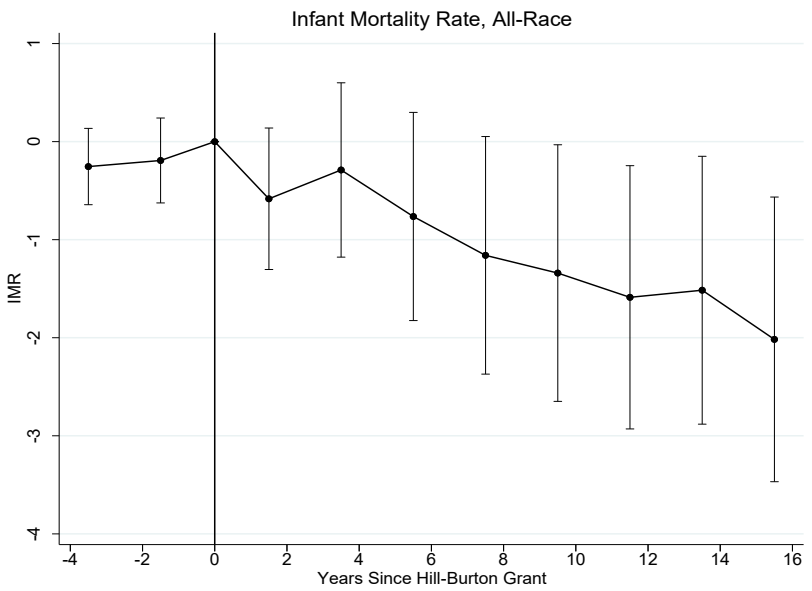
Table 2: Static Difference-in-Difference Estimates

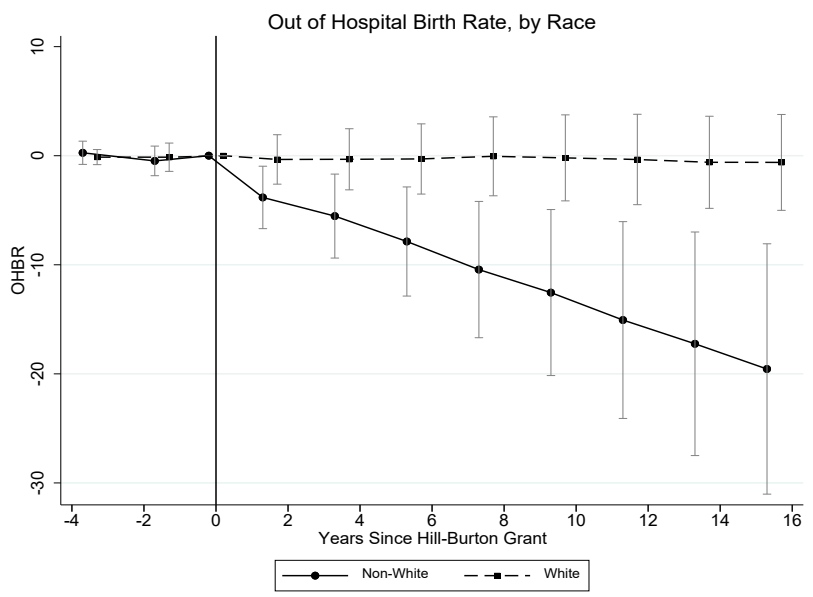
	(1) OHBR, All-Race	(2) OHBR, Non-White	(3) OHBR, White	(4) IMR, All-Race	(5) IMR, Non-white	(6) IMR, White
			Panel A: Public Hill-Burton Projects			
Treatment Indicator	-3.5**	-16.2***	-0.1	-1.8**	-6.0***	-0.8
	(1.4)	(6.0)	(2.5)	(0.9)	(1.9)	(1.0)
Observations (county-years)	56,588	15,609	15,609	56,588	4,260	4,260
Unique Treated Counties	747	278	278	747	57	57
Unique Never-Treated Counties	569	85	85	569	85	85
			Panel B: Private Hill Burton Projects			
Treatment Indicator	7.4***	5.6	4.4	0.8	-3.3	-0.3
	(1.0)	(9.9)	(3.0)	(0.7)	(4.0)	(1.4)
Observations (county-years)	49,837	6,536	6,536	49,837	3,000	3,000
Unique Treated Counties	590	67	67	590	15	15
Unique Never-Treated Counties	569	85	85	569	85	85

Notes: Data is at the county-year level. Column titles indicate the dependent variable used in each specification. The out-of-hospital birth rate is measured from 0-100, while the infant mortality rate is measured as the number of infant deaths per 1,000 live births. All models are estimated using the technique from Gardner et al. (2025), which in this application residualizes outcomes of county and calendar year fixed effects that were estimated using only untreated observations. There are no additional covariates. All models are weighted by the number of births, with race-specific values of these weights used where applicable. Standard errors, reported in parentheses, are clustered at the county level and calculated using the formulas from Gardner et al. (2025). *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Figure 1: Event Study Estimates, Public Hospitals



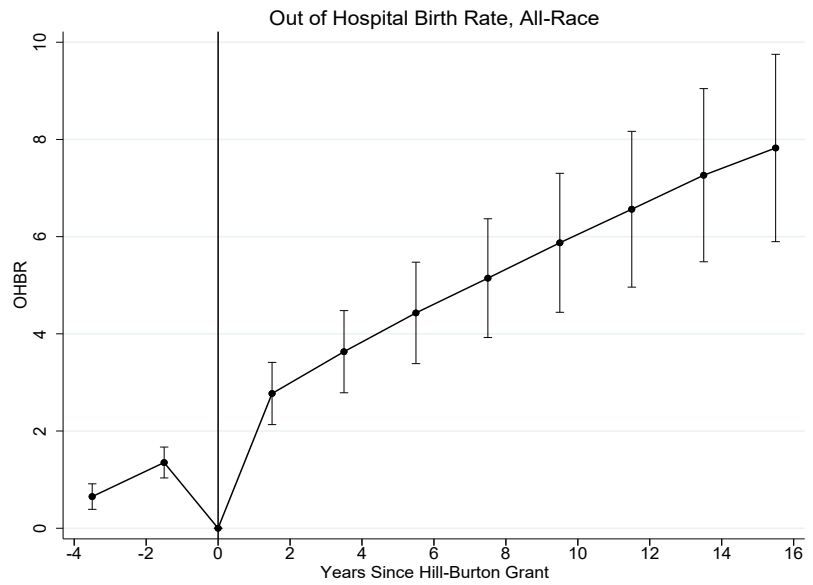


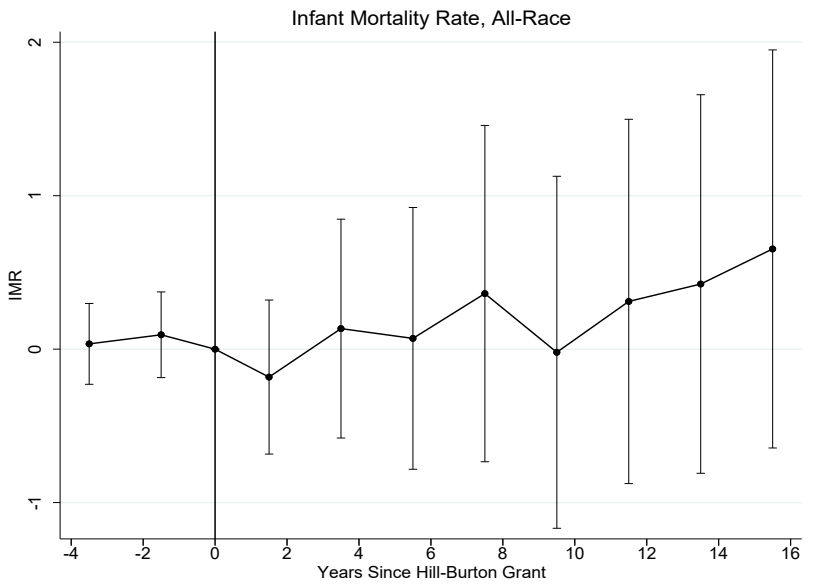


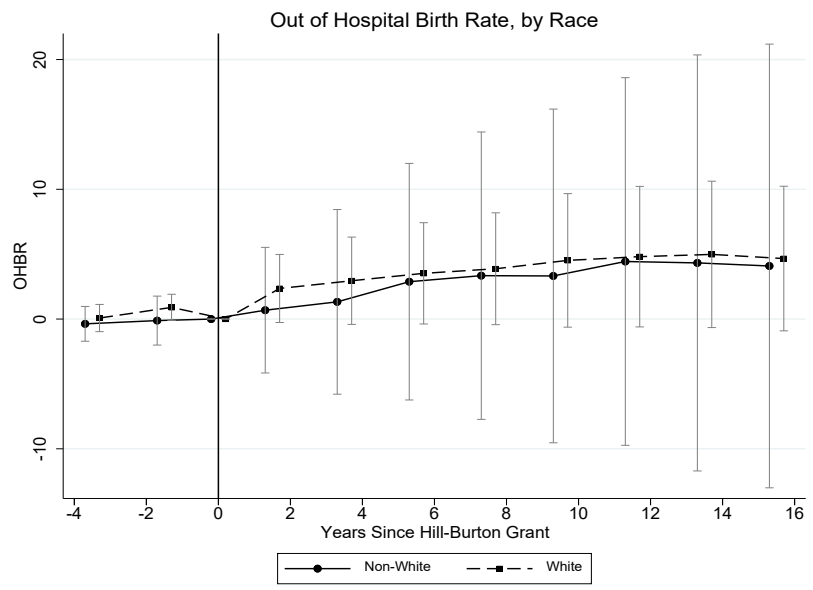


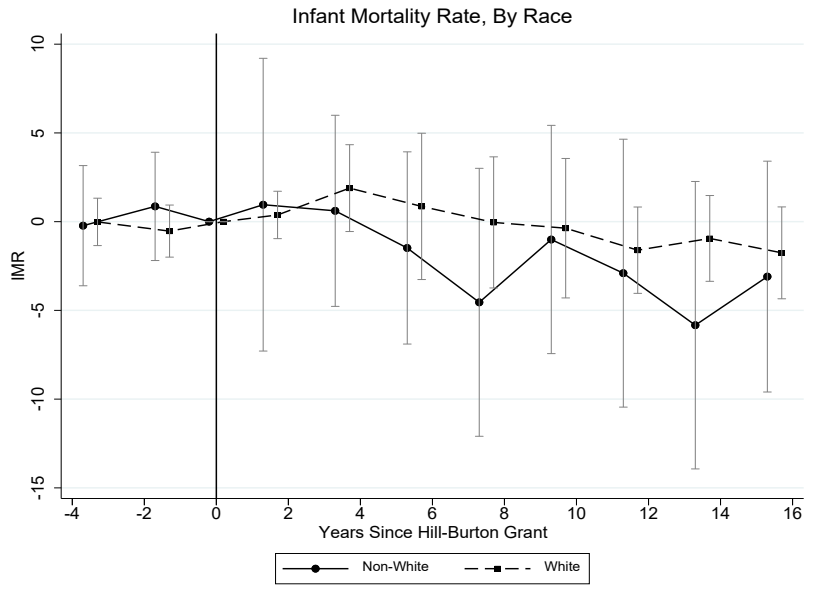
Notes: Data is at the county-year level. Samples are balanced in event-time, such that each reported event-time coefficient is estimated using a common set of counties. Given the years for which outcomes are available, this balancing requires restricting to treatments occurring in 1950 or later for the all-race outcomes and the race-specific OHBR, and to treatments occurring in 1963 or later for the race-specific IMR. The number of treated and never-treated counties for each outcome and the total sample sizes are reported in the corresponding static models from Table 2. The out-of-hospital birth rate is measured from 0-100, while the infant mortality rate is measured as the number of infant deaths per 1,000 live births. All models are estimated using the technique from Gardner et al. (2025), which in this application residualizes outcomes of county and calendar year fixed effects that were estimated using only untreated observations. There are no additional covariates. All models are weighted by the number of births, with race-specific values of these weights used where applicable. 90% confidence intervals are calculated using standard errors from the formulas in Gardner et al. (2025) with county-level clustering.

Figure 2: Event Study Estimates, Private Hospitals









Notes: Data is at the county-year level. Samples are balanced in event-time, such that each reported event-time coefficient is estimated using a common set of counties. Given the years for which outcomes are available, this balancing requires restricting to treatments occurring in 1950 or later for the all-race outcomes and the race-specific OHBR, and to treatments occurring in 1963 or later for the race-specific IMR. The number of treated and never-treated counties for each outcome and the total sample sizes are reported in the corresponding static models from Table 2. The out-of-hospital birth rate is measured from 0-100, while the infant mortality rate is measured as the number of infant deaths per 1,000 live births. All models are estimated using the technique from Gardner et al. (2025), which in this application residualizes outcomes of county and calendar year fixed effects that were estimated using only untreated observations. There are no additional covariates. All models are weighted by the number of births, with race-specific values of these weights used where applicable. 90% confidence intervals are calculated using standard errors from the formulas in Gardner et al. (2025) with county-level clustering.

Online Appendix, The Rise of the Modern Hospital and Early-Life Health: Evidence from the Hill-Burton Act

Jason Fletcher, Owen Thompson, Karin Wu

January 30, 2026

This appendix reports various supplementary estimates, extensions and robustness checks. In most cases we focus on our findings for public Hill-Burton treatments, both all-race and race-specific. Private Hill-Burton treatments are not our focus here because there are not positive findings for which to establish robustness, and because in many cases the analyses of private treatments likely suffer from parallel trends violations, as discussed in the main text.

1 Correlates of Treatment Timing

As discussed in detail in the main paper, some of our identifying variation comes from comparisons of earlier-treated counties with not-yet-treated counties, with a corresponding parallel trends assumption that outcome trends in not-yet-treated counties accurately characterize the outcome trends that would have occurred in treated counties in the post-treatment periods. This assumption would hold if treatment timing were as good as randomly assigned, although random timing is not required and in fact if treatment timing were random preferable alternative estimators exist (see Roth et al. [2023] Section 6).

Appendix Table A1 uses county-level data for the sample of treated counties, and regresses the year of treatment for each county onto pre-treatment county characteristics. We do so for both publicly and privately treated counties, and report both bivariate models with each characteristic as well as a model that includes all pre-treatment county characteristics simultaneously. As described in the main paper, there are significant associations in Table A1 between treatment timing and baseline total population, possibly reflecting that the program intentionally prioritized medically underserved counties. Year of treatment is also associated with county level education and income in bivariate specifications within privately treated counties, but not in the full model. We report these associations for transparency but emphasize that our parallel trends assumption requires that *outcomes* in not-yet-treated counties accurately characterize counterfactual *outcomes* in treated counties, and that the paucity of pre-trends in our event-study specifications for public treatments support this assumption.

2 Alternative Event-Time Windows and Imbalanced Panels

Because the first Hill-Burton grants were awarded in 1947 and the Vital Statistics data is available beginning in 1946 (or later for race-specific IMR) we face tradeoffs between including many pre-treatment event-times, the number of treated counties included in the working sample, and whether each event-time coefficient is estimated on a common set of counties ("balance"). Our baseline models restricted the estimation samples to be balanced from event time -4 through event time +16, such that each event-time in this range was observed for each county in the estimation sample, which required excluding grants awarded prior to 1950 (or prior to 1963 in the race-specific IMR models). This represents our preferred balancing of these competing objectives, but in Figures A1 and A2 we present results that make different choices about these tradeoffs.

In Figure A1 we report results that continue to require a balanced panel but that extend the pre-period back eight years, rather than four. For all-race OHBR and IMR and for race-specific OHBR this requires excluding grants awarded prior to 1954 (rather than 1950 in the baseline models), while for the race-specific IMR models this requires excluding grants awarded prior to 1967 (rather than 1963 in the baseline models). Because many counties were treated relatively early in the study period the reductions in sample size to obtain this longer pre-period are substantial: The number of treated units is reduced from 747 to 557 in the all-race models, from 278 to 196 in the race-specific OHBR models, and from 57 to 23 in the race-specific IMR models. The key feature of the results in Figure A1 is that all of the pre-trends over this longer pre-period are visually consistent with parallel trends in the pre-period holding, which is suggestive evidence that the (untestable) parallel trends assumption holds as well. The treatment effect magnitudes are also similar, although the results for the race-specific IMR are somewhat erratic, likely due to the small sample of treated units.

Figure A2 reports the results of models that do not require a balanced panel, and instead use all of the counties for which each included event-time is observed to estimate the coefficients. We report a wide range of event times for these models, -8 through +20. Note, however, that with balance not imposed the coefficients for narrower ranges are numerically identical to those for the wider range reported in Figure A2. For example results using our baseline -4 through +16 event time range without balance imposed are simply the coefficients in Figure A2 over that range. The increases in the number of observations contributing to the estimates

when balance is not imposed are large: The number of treated units increases from 747 to 1063 in the all-race models, from 278 to 407 in the race-specific OHBR models, and from 57 to 407 in the race-specific IMR models. These sample sizes are somewhat misleading though, because the nature of an imbalanced panel is such that the number of counties contributing to each specific event-time coefficient is different, and relatively distant event times in particular are estimated on small (and possibly non-random) subsets of counties that were treated very early or very late. This is especially true for the race-specific IMR models, where outcome data is not available until 1959, but the models in Figure A2 use events from well before 1959 to identify these distant positive event times. Nonetheless, in practice the estimates using the imbalanced samples in Figure A2 are generally quite similar to the baseline estimates.

3 Alternative Estimators and Evaluating Pre-Trends

Our baselines results came from the two-stage DiD estimator of Gardner et al. [2025]. In Figure A3 we reproduce the baseline results with this method for reference, but also report estimates from OLS, from the method proposed by Sun and Abraham [2021], and from the method proposed by Callaway and Sant'Anna [2021].[1] The four panels of Figure A3 report results from each estimator for (a) the all-race OHBR, (b) the all-race IMR, (c) the non-white OHBR, and (d) the non-white IMR. In each panel, all four estimators show clear declines in the years following a Hill-Burton project and we characterize our estimates as being robust to using these alterative estimators.

That being said, the OLS and Sun and Abraham [2021] estimates for the OHBR outcomes exhibit patterns in the coefficients on leads of treatment status that could be characterized as pre-trends. We emphasize that we prefer the Gardner et al. [2025] estimator for substantive reasons. The issues with OLS estimates in staggered adoption designs are well known, and are shown here only for transparency. The Sun and Abraham [2021] approach relies on aggregating treatment-cohort specific event study coefficients that use the never-treated units as controls, and therefore do not incorporate timing-based comparisons, leading to a substantial loss of identifying variation in our application. We therefore feel that the Gardner et al. [2025] approach is, for current purposes, the most transparent and flexible of the "new" difference-in-difference methods, and the authors shows that it performs well along various

[1] We do not report estimates from the Borusyak et al. [2024] method because the point estimates are numerically identical to the baseline Gardner et al. [2025] estimates.

dimensions in simulations. We also emphasize that any potential concerns apply only to the OHBR, and that there are no visually apparent pre-trends for the IMR in any population and using any estimator.

Beyond evaluating the specific patterns shown in Figure A3, the applied econometric literature on evaluating pre-trends from more recently developed DiD estimators is active and not fully resolved. A recent working paper by Roth [2024] reports results of different estimators in a simulated data set that has uniform treatment timing, null-treatment effects, and parallel trends violations. Roth [2024] finds that several of the new DiD estimators still produce kinks or discrete jumps at the time of treatment and often fail to clearly indicate pre-trends violations based on visual inspection.

Roth [2024] shows that these surprising patterns arise because the new DiD techniques all estimate the coefficients on pre-treatment event times differently than the post-treatment event time coefficients. In the specific case of Gardner et al. [2025], this asymmetry arises because for the **pre-period** the two-step residualization procedure described in the main paper effectively runs a traditional TWFE model on the untreated units and uses the earliest time period as the reference group for estimating the time period effects. The post-treatment coefficients, in contrast, are normalized to the average outcome in the pre-treatment period. This change in the reference period at the time of treatment is what can lead to discrete jumps at the time of treatment in the [Gardner et al., 2025] and [Borusyak et al., 2024] estimators. Critically, however, the coefficients on leads of treatment status under these methods do still in fact represent valid tests of parallel trends in the pre-treatment period. Roth [2024] is quite explicit on this point, writing that "the pre-treatment coefficients in the BJS event study [which are numerically identical to the Gardner et al. [2025] estimates] are still valid tests of pre-treatment parallel trends..." (p.11).[2] What is different is that the coefficients on pre-treatment event time indicators using Gardner et al. [2025] cannot be interpreted as mean deviations from trends, as they are in OLS. Instead, the coefficients on leads of treatment status test whether average residuals from the first stage of the Gardner et al. [2025] procedure are close to zero in pre-treatment periods and whether they vary with the number of periods until treatment. The key conclusion is that while there are some differences in interpreting pre-trends from traditional TWFE event studies versus our method, the pre-trends from our baseline estimates do in fact represent valid tests of parallel trends in the pre-treatment period.

[2] Gardner et al. [2025] also discuss the issue in detail in their Section 2.6.

Finally, we note the closely related point that the Roth [2024] simulation results for the Borusyak et al. [2024] method showed an increasing trend in the pre-periods, then a discrete drop at the time of treatment, and finally a resumption of positive trends. This simulation result, which contained parallel trends violations by design, bears a striking resemblance to the all-race OHBR results for privately treated counties in Figure 2 of the main paper, where we strongly suspected parallel trends violations.

4 Weighting

This section assesses the impact of not weighting our baseline models by birth counts. Because this involves estimating a large numbers of variants of our baseline models, we report these alternative models using static estimates rather than event studies, but full event studies do not change any of the conclusions. Weighted versus unweighted variants of our main models are reported in Table A2. Panels A-D of Table A2 respectively report separate results for the all-race OHBR, the all-race IMR, the nonwhite OHBR, and the nonwhite IMR.

Beginning with the all-race outcomes, Columns 1 and 2 of Panel A show that the unweighted treatment effect estimate for the all-race OHBR is virtually identical to the weighted version, but Columns 1 and 2 Panel B shows that the point estimate for the all-race IMR falls to -.42 without weighting and is no longer statistically significant at conventional levels.

As noted in the main paper, the weighted baseline estimates target the treatment effect for the average *birth*, which would be obtained from birth-level data if such data were available, whereas the unweighted estimates target the treatment effect for the average *county*, regardless of its size. The weighted and unweighted estimates are therefore simply targeting different parameters, and both target parameters are valid and of potential interest, although we believe the effect on the average birth is somewhat more intuitive and policy relevant.

On a mechanical level, the smaller coefficient in the unweighted IMR model means that treatment effect estimates are greater in larger county-years (as measured by number of births).[3] Precisely why there are smaller all-race IMR estimates in counties with fewer births is not obvious, and perhaps even somewhat puzzling given that hospital access is the primary hypothesized mechanism for the IMR effects and that unweighted OHBR estimates remain large. One possible explanation is measurement error in smaller counties. The IMR

[3]We have verified that this is in fact the case by estimating the base model across quartiles of the total births distribution.

is a *rate*, with total births in the denominator. Because infant deaths are relatively rare events, mismeasurement of a small number of them in a county with very few total births could generate large errors in the measured rates. And even absent such measurement error, idiosyncratic year-to-year fluctuations in the IMR will be higher in smaller counties, possibly making estimates in smaller counties imprecise or unstable. Weighting by total births would mitigate this type of measurement error, since it places less weight on the counties with fewer births where these issues are most likely and severe.[4]

To evaluate this possibility directly, Columns 3 and 4 of Table A2 report weighted and unweighted estimates of our baseline specification, but exclude county-years with zero infant deaths, which are much more common in county-years with fewer overall births and therefore down-weighted in our baseline models (for completeness we also report results excluding observations zero out-of-hospital births, even though these estimates were unaffected by weighting.). The IMR estimates with zeros excluded are somewhat larger than the weighted baseline, and more importantly are virtually identical in the weighted versus unweighted models. This suggests that much of the effect of weighting in our baseline models actually comes specifically from putting very little weight on counties with *zero* infant deaths, which is much more likely in less populous counties with fewer overall births. We interpret this as strong suggestive evidence that the smaller IMR estimates in the unweighted models is due to measurement error related to the relative rarity of infant-deaths in less populous counties, although other explanations for this heterogeneity exist. In any case, as noted, we interpret any disparities between weighted and unweighted estimates as being due to targeting a different parameter for estimation, rather than a failure of robustness per-se. We also note that observations with zero infant deaths would be excluded if we were to use the log of the IMR as the outcome, and that estimates using log IMR (not shown) are indeed insensitive to weighting, with a weighed coefficient of -.12 (p=.007) and an unweighted coefficient of -.13 (p=.000).

Panels C and D of Table A2 report the same sets of weighted and unweighted results with the non-white OHBR and non-white IMR as outcomes. In general similar patterns hold, with small and statistically insignificant IMR estimates in unweighted models, but with weighting largely inconsequential when excluding county-years with zero infant deaths.

We do note that the unweighted non-white OHBR estimates are substantially smaller than

[4]The same reasoning applies to the OHBR, but to a lesser extent since out-of-hospital births are generally more common than infant-deaths.

the weighted versions, whether or not zeros are included, although the estimates are still negative and of non-trivial magnitudes. There is no immediately obvious explanation for these differences, but we do note that the sample for the race-specific OHBR model already excludes counties where less than 10% of the population was non-white. The estimation sample for these models is therefore somewhat mechanically restricted in terms of containing observations that receive zero weight or very little weight. And more generally the distribution of total non-white births within counties where 10% or more of the population was non-white throughout the study period more restricted and idiosyncratic in nature than overall total birth counts.

5 Continuous Treatments

The main paper uses binary treatment measures, or in the event-study models leads and lags of those binary measures. But there is also usable variation in intensity of treatment, since some Hill-Burton projects were larger than others, and also because some counties received multiple distinct Hill-Burton grants in different years (sometimes for the same facility and sometimes for different facilities). In Table A3 we present the results of models that are structurally the same as those from Table 2 of the main paper, but replace the binary indicator of having a Hill-Burton project with the *cumulative* number of hospital beds per thousand residents constructed with Hill-Burton funds in each county-year. The results have the same signs and same approximate magnitudes as the baseline results (the magnitudes are not directly comparable given differences in the units), and the effect of public treatments on white IMR is now marginally statistically significant. Similar models that use cumulative Hill-Burton *funds* per-capita rather than beds per-capita (not shown) are extremely similar, which is unsurprising given that the size and cost of projects are very strongly correlated.

References

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285, 2024.

B. Callaway and P. H. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230, 2021.

J. Gardner, N. Thakral, L. T. Tô, and L. Yap. Two-stage differences in differences. 2025.

J. Roth. Interpreting event-studies from recent difference-in-differences methods. *arXiv preprint arXiv:2401.12309*, 2024.

J. Roth, P. H. Sant'Anna, A. Bilinski, and J. Poe. What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023.

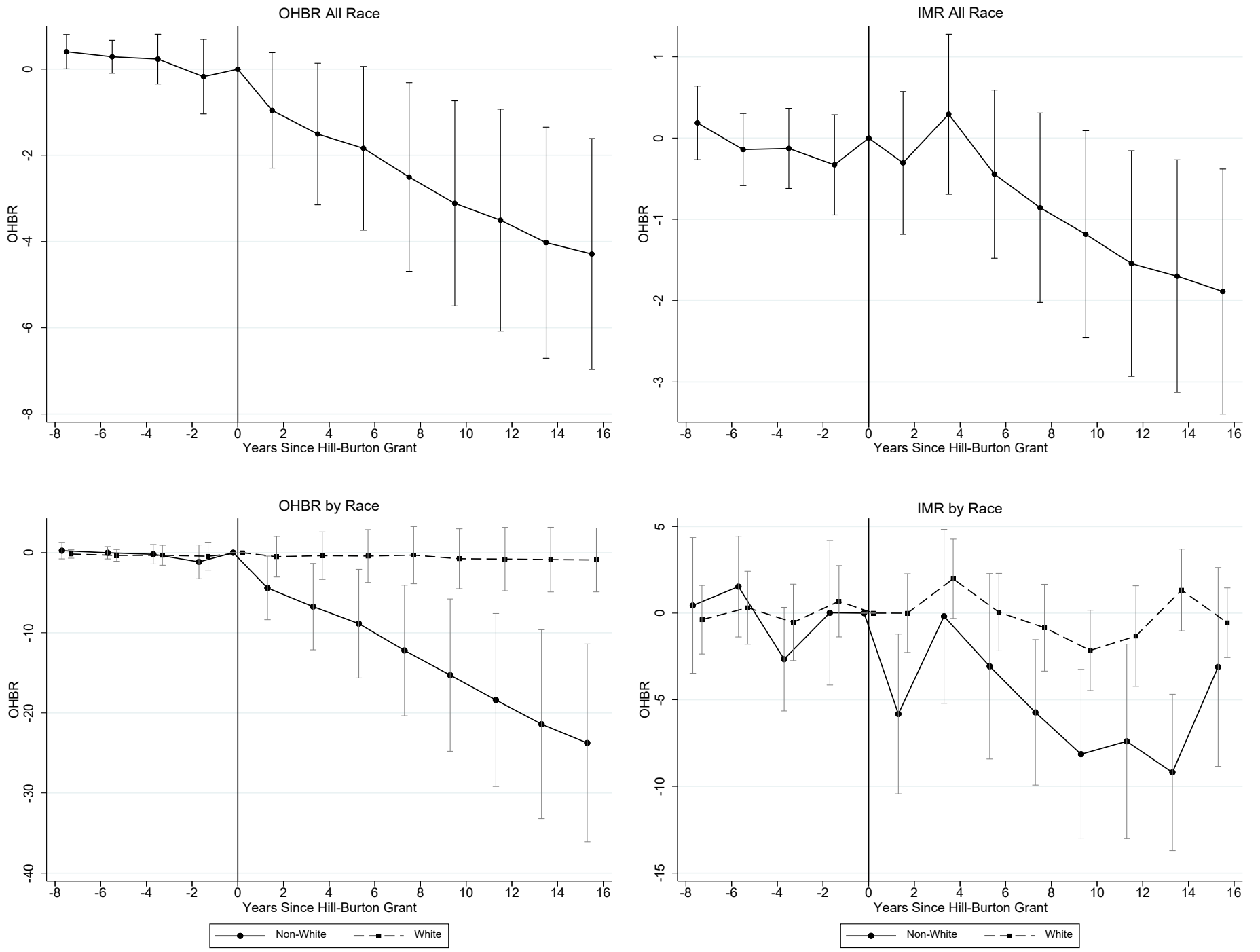
L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of econometrics*, 225(2):175–199, 2021.

Table A1: Timing Correlates

	(1) Total Population	(2) Median Education	(3) Median Income	(4) Percent Black	(5) Infant Mortality Rate	(6) Out of Hospital Birth Rate	(8) All
			Panel A: Public Treatments				
Total Population (1000's)	-0.020**						-0.018**
	(0.008)						(0.008)
Median Education		0.183					0.205
		(0.158)					(0.246)
Median Income ($100s, 1950 dollars)			-0.001				-0.053
			(0.027)				(0.044)
Percent Black				-0.016			-0.001
				(0.011)			(0.016)
Infant Mortality Rate					-0.010		0.001
					(0.016)		(0.018)
Out of Hospital Birth Rate						-0.011	-0.015
Observations (counties)	747	747	747	747	747	747	747
			Panel B: Private Treatments				
Total Population (1000's)	-0.006***						-0.005***
	(0.002)						(0.002)
Median Education		-0.453**					-0.207
		(0.202)					(0.273)
Median Income ($100s, 1950 dollars)			-0.098***				-0.065
			(0.035)				(0.056)
Percent Black				0.013			0.005
				(0.025)			(0.030)
Infant Mortality Rate					0.011		-0.006
					(0.020)		(0.023)
Out of Hospital Birth Rate						0.019	-0.017
						(0.018)	(0.029)
Observations (counties)	590	590	590	590	590	590	590

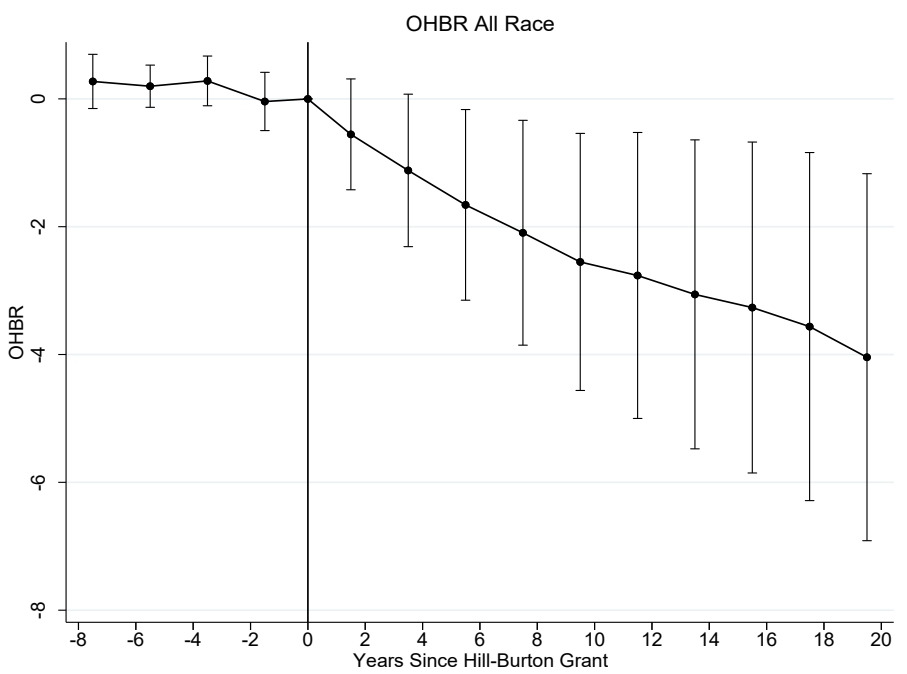
Notes: Reported means are all for calendar year 1950. Total population, percent Black, median income and median education are drawn from the City and County Data Books of the 1950 Census, contained in ICPSR Data Set 07736. The Infant Mortality Rate and Out of Hospital Birth Rate are from the Vital Statistics data described in the text.

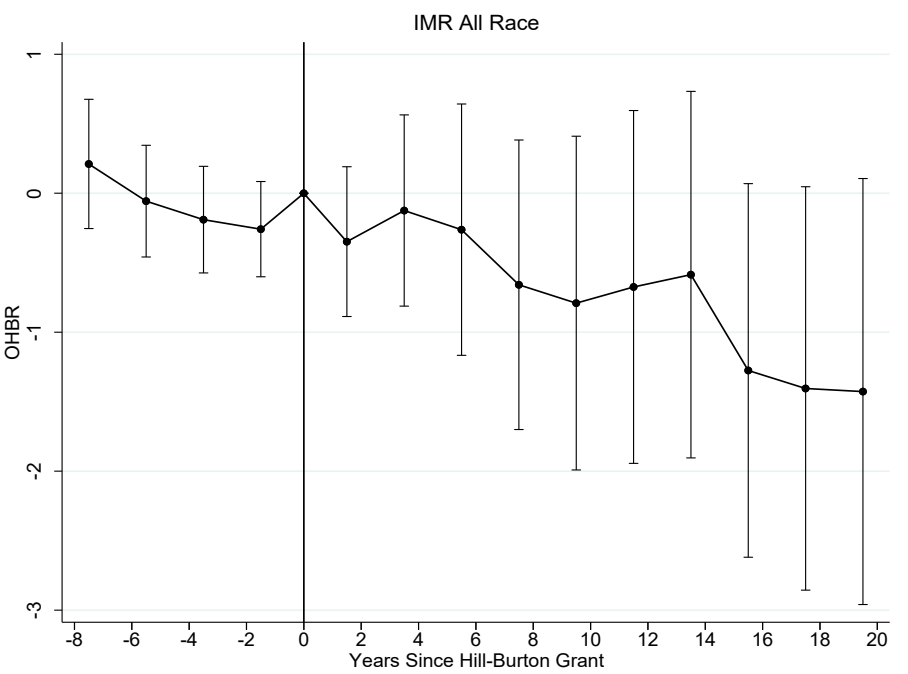
Figure A1: Extended Pre-Treatment Window, Balanced Panel

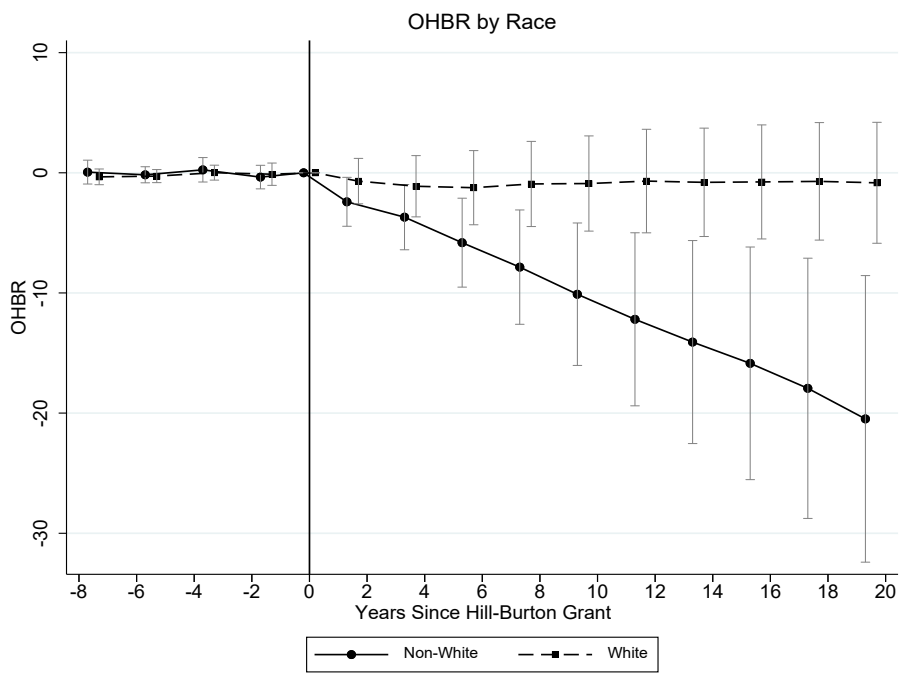


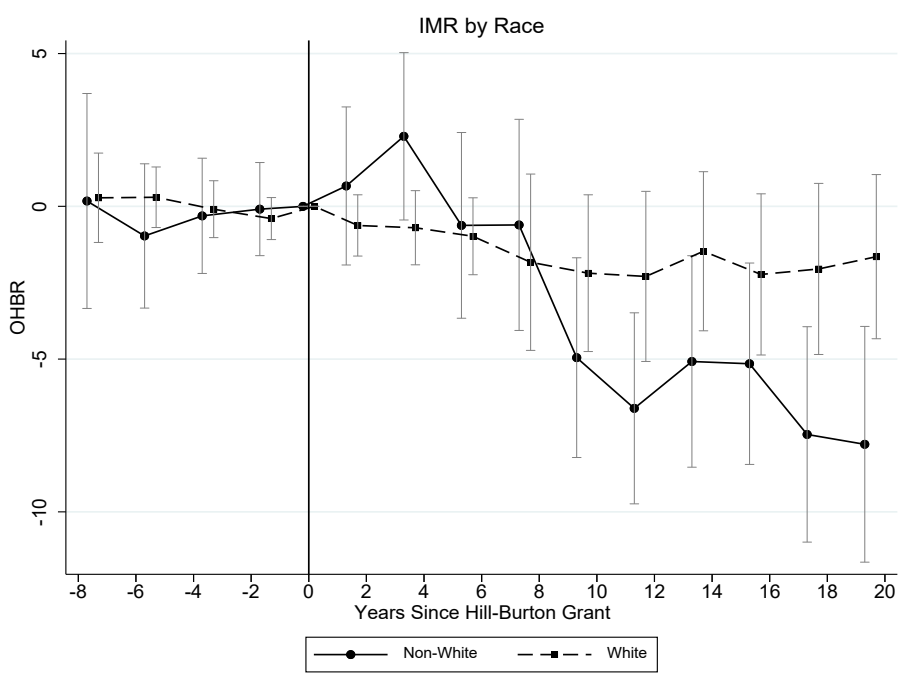
Notes: Samples are balanced in event-time, such that each reported event-time coefficient is estimated using a common set of counties. To obtain 8 pre-treatment coefficients, rather than the 4 pre-treatment coefficients in the main paper, this balancing requires restricting to treatments occurring in 1954 or later for the all-race outcomes and the race-specific OHBR, and to treatments occurring in 1967 or later for the race-specific IMR. As a result, the number of treated units is reduced from 747 to 557 in the all-race models, from 278 to 196 in the race-specific OHBR models, and from 57 to 23 in the race-specific IMR models. Data is at the county-year level. The out-of-hospital birth rate is measured from 0-100, while the infant mortality rate is measured as the number of infant deaths per 1,000 live births. All models are estimated using the technique from Gardner et al. (2025), which in this application residualizes outcomes of county and calendar year fixed effects that were estimated using only untreated observations. There are no additional covariates. All models are weighted by the number of births, with race-specific values of these weights used where applicable. 90% confidence intervals are calculated using standard errors from the formulas in Gardner et al. (2025) with county-level clustering.

Figure A2: Unbalanced Panels



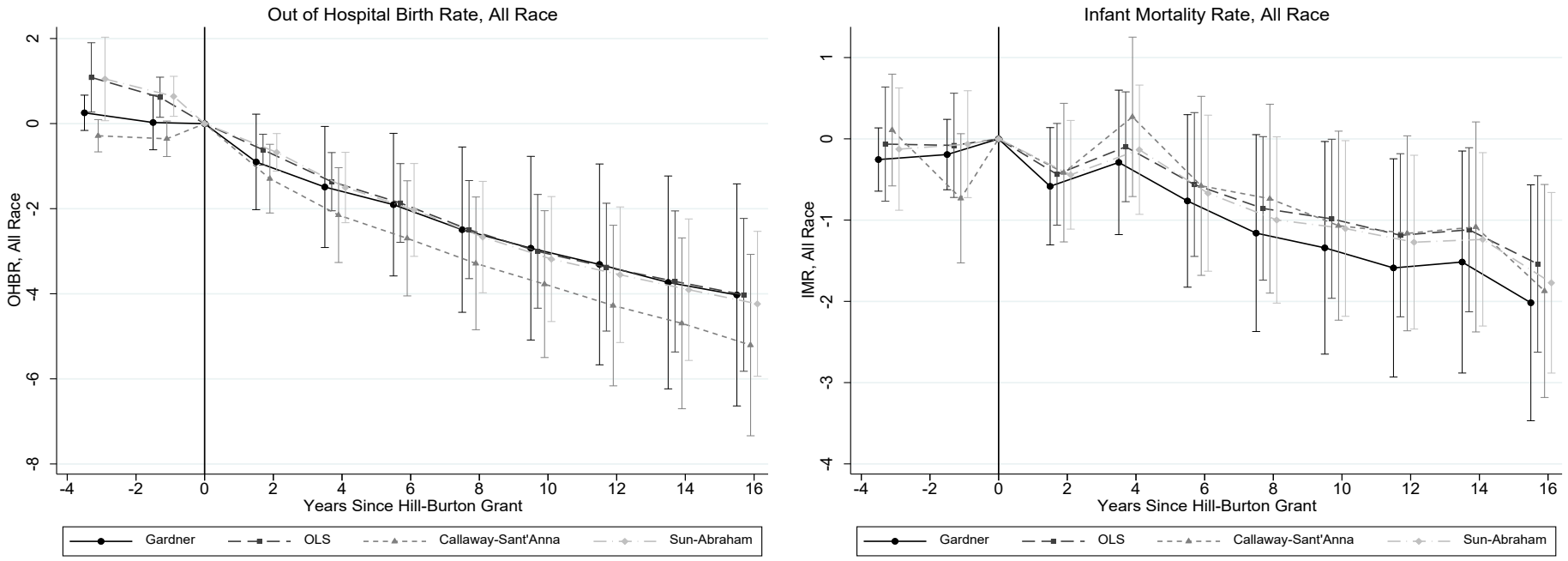


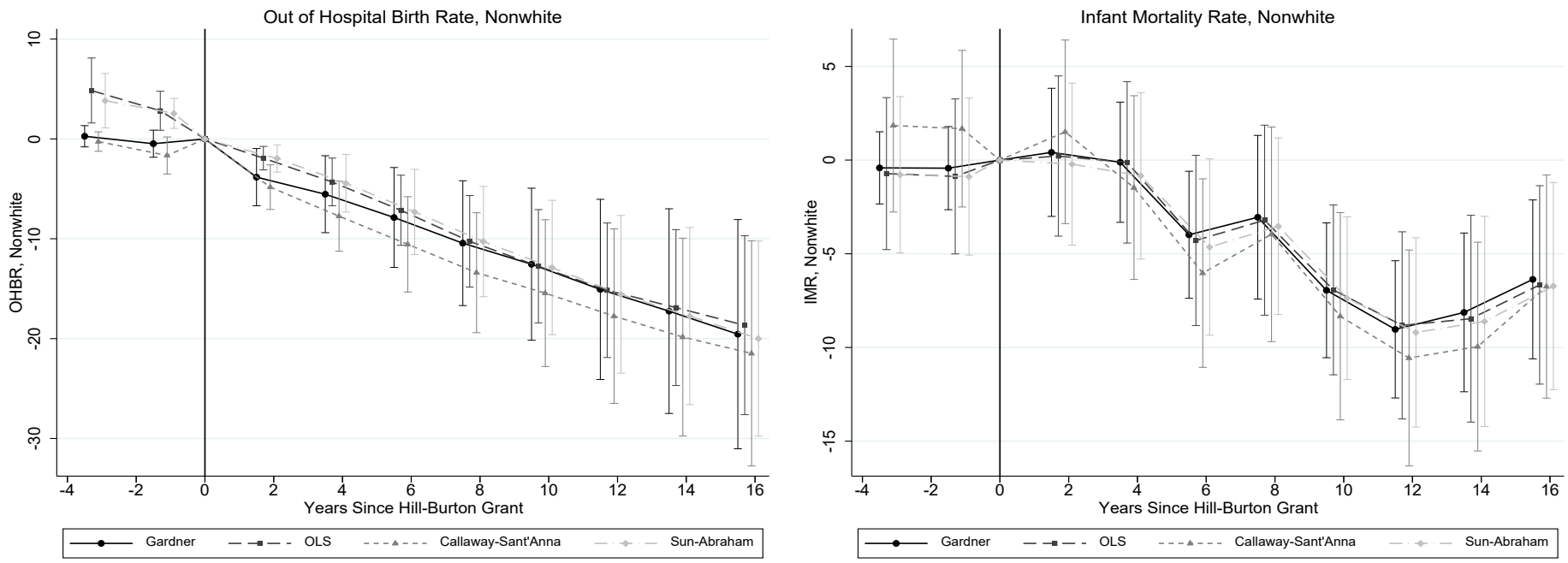




Notes: Samples are **not** balanced in event-time, such that each reported event-time coefficient is estimated using different sets of counties. The **total** number of treated units used to identify **any** event-time increases from 747 to 1063 in the all-race models, from 278 to 407 in the race-specific OHBR models, and from 57 to 407 in the race-specific IMR models. Data is at the county-year level. The out-of-hospital birth rate is measured from 0-100, while the infant mortality rate is measured as the number of infant deaths per 1,000 live births. All models are estimated using the technique from Gardner et al. (2025), which in this application residualizes outcomes of county and calendar year fixed effects that were estimated using only untreated observations. There are no additional covariates. All models are weighted by the number of births, with race-specific values of these weights used where applicable. 90% confidence intervals are calculated using standard errors from the formulas in Gardner et al. (2025) with county-level clustering.

Figure A3: Alternative Estimators





Notes: The Gardner estimates reproduce the results from the main paper, while the remaining estimates are from OLS estimation and the methods outlined in Sun & Abraham (2021) and Callaway & Sant'Anna (2021), as indicated. Data is at the county-year level. Samples are balanced in event-time, such that each reported event-time coefficient is estimated using a common set of counties. Given the years for which outcomes are available, this balancing requires restricting to treatments occurring in 1950 or later for the all-race outcomes and the race-specific OHBR, and to treatments occurring in 1963 or later for the race-specific IMR. The number of treated and never-treated counties for each outcome and the total sample sizes are reported in the corresponding static models from Table 2 of the main paper. The out-of-hospital birth rate is measured from 0-100, while the infant mortality rate is measured as the number of infant deaths per 1,000 live births. There are no additional covariates. All models are weighted by the number of births, with race-specific values of these weights used where applicable. 90% confidence intervals are calculated using standard errors from the indicated estimator with county-level clustering.

Table A2: Weighting

	(1) Base	(2) Unweighted	(3) No Zeros, Weighted	(4) No Zeros, Unweighted
		Panel A: Out of Hospital Birth Rate, All Race		
Treatment Indicator	-3.50**	-3.46***	-3.80**	-3.66***
	(1.42)	(0.82)	(1.61)	(0.89)
Observations (county-years)	56,588	56,588	40,123	40,123
Unique Treated Counties	747	747	747	747
Unique Never-Treated Counties	569	569	569	569
		Panel B: Infant Mortality Rate, All Race		
Treatment Indicator	-1.79**	-0.42	-2.41***	-2.54***
	(0.85)	(0.47)	(0.82)	(0.48)
Observations (county-years)	56,588	56,588	50,521	50,521
Unique Treated Counties	747	747	747	747
Unique Never-Treated Counties	569	569	569	569
		Panel C: Out of Hospital Birth Rate, Non-White		
Treatment Indicator	-16.24***	-3.13	-17.65***	-3.88**
	(5.95)	(2.20)	(5.27)	(1.97)
Observations (county-years)	15,609	15,609	12,259	12,259
Unique Treated Counties	278	278	278	278
Unique Never-Treated Counties	85	85	85	85
		Panel D: Infant Mortality Rate, Non-White		
Treatment Indicator	-6.03***	0.98	-6.64***	-4.19*
	(1.89)	(2.03)	(1.41)	(2.16)
Observations (county-years)	4,260	4,260	3,520	3,520
Unique Treated Counties	57	57	57	57
Unique Never-Treated Counties	85	85	85	85

Notes: Columns 1 and 3 weight by the total number of births taking place in each county-year, using race-specific values of total births in Panel C and D, while Columns 2 and 4 do not apply any weights. Columns 3 and 4 exclude observations with zero out-of-hospital births and zero infant deaths, respectively. Note that the dependent variable for the observations excluded in Columns 3 and 4 is equal to zero. The out-of-hospital birth rate is measured from 0-100, while the infant mortality rate is measured as the number of infant deaths per 1,000 live births. All models are estimated using the technique from Gardner et al. (2025), which in this application residualizes outcomes of county and calendar year fixed effects that were estimated using only untreated observations. There are no additional covariates. Standard errors, reported in parentheses, are clustered at the county level and calculated using the formulas from Gardner et al. (2025). *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table A3: Static Difference-in-Difference Estimates with Continuous Treatments

	(1) OHBR, All-Race	(2) OHBR, Non-White	(3) OHBR, White	(4) IMR, All-Race	(5) IMR, Non-white	(6) IMR, White
			Panel A: Public Hill-Burton Projects			
Cumulative Hill-Burton Beds per 1,000 Residents	-1.0***	-4.1***	-0.4	-0.3**	-1.0**	-0.4*
	(0.3)	(1.1)	(0.5)	(0.1)	(0.5)	(0.2)
Observations (county-years)	56,588	15,609	15,609	56,588	4,260	4,260
Unique Treated Counties	747	278	278	747	57	57
Unique Never-Treated Counties	569	85	85	569	85	85
			Panel B: Private Hill Burton Projects			
Cumulative Hill-Burton Beds per 1,000 Residents	1.4***	-0.2	0.7	0.1	0.3	0.1
	(0.2)	(2.8)	(1.0)	(0.2)	(0.7)	(0.3)
Observations (county-years)	49,764	6,523	6,523	49,764	3,000	3,000
Unique Treated Counties	590	67	67	590	15	15
Unique Never-Treated Counties	569	85	85	569	85	85

Notes: This table is identical to Table 2 of the main paper except the binary treatment indicator is replaced with a continuous measure of the cumulative number of Hill-Burton sponsored beds per 1,000 residents in each county-year. Data is at the county-year level. Column titles indicate the dependent variable used in each specification. The out-of-hospital birth rate is measured from 0-100, while the infant mortality rate is measured as the number of infant deaths per 1,000 live births. All models are estimated using the technique from Gardner et al. (2025), which in this application residualizes outcomes of county and calendar year fixed effects that were estimated using only untreated observations. There are no additional covariates. All models are weighted by the number of births, with race-specific values of these weights used where applicable. Standard errors, reported in parentheses, are clustered at the county level and calculated using the formulas from Gardner et al. (2025). *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.