DOCTORS WITH(OUT) BORDERS: EFFECTS OF EASING OCCUPATIONAL LICENSING FOR FOREIGN-TRAINED PHYSICIANS*

Wenni Yang[a],†

[a]University of California, Davis

October 2026

Click here for the latest version

Abstract

This paper studies how easing licensing requirements for non-U.S.-citizen physicians affects the supply and composition of the physician workforce and hospitals' investment in other inputs. From 1960 through the 1970s, more than thirty states removed citizenship requirements for medical licensure. I use foreign medical graduates (FMGs) as a proxy for non-citizen physicians. Using newly assembled licensure and physician data, I show that easing a licensing requirement changes physician supply through expanding entry, with a suggestive decline among established foreign incumbents, no detectable crowd-out of U.S.-trained physicians, and no short-run deterioration in the average quality of the licensed FMG stock. Hospitals respond by expanding staffed beds by about 4 percent. Effects are concentrated among counties with above-median Medicaid exposure, suggesting physician supply is a binding constraint on hospital input adjustment in anticipation of a growth in demand for health care. Estimated effects on population mortality are imprecise. These effects emerge primarily after the federal government relaxed J-1 visa return requirements in 1970, indicating complementary constraints between state and federal regulation.

Keywords: Occupational licensing; foreign-trained physicians; high-skilled immigration; physician labor supply; hospital capacity.

JEL codes: J44, J61, I11, I18, N32.

*I am grateful to Santiago Perez, Marianne Bitler, and Giovanni Peri for their guidance and feedback throughout my exploration. I also thank Joshua Gottlieb for his invaluable feedback on this project. I thank participants at ASHEcon, the AEA Annual Meeting, the All California Labor Conference, the Bay Area History Workshop, the Summer School on the Economics of Migration, and the UC Davis History Talk for all their helpful comments. I thank Amy Finkelstein for sharing the AHA data. I thank Ethan Po, Henry Htut, Matthew Low, and Thomas Lin for their excellent research assistance. All remaining errors are my own.

†Wenni Yang, Email: wniyang@ucdavis.edu

1 Introduction

A well-functioning healthcare system depends on a key input: human capital. Yet increasing the supply of physicians remains challenging, given the long training pipeline and the strict entry regulations intended to protect public safety. Since training domestic physicians often takes close to a decade, many countries have adopted policies to attract and retain foreign-trained physicians, who now make up about 25 percent of physicians in the United States and roughly 20 percent of physicians across OECD countries on average (OECD, 2023; Educational Commission for Foreign Medical Graduates, 2025). Licensing remains the last bottleneck for foreign-trained physicians to enter the labor market to provide care without supervision (Friedman and Kuznets, 1945; Kleiner and Krueger, 2013). This bottleneck has become an area of active policy experimentation: by 2026, twenty-four states had eased some licensing requirements for foreign-trained physicians while seventeen others were debating similar legislation (Federation of State Medical Boards, 2025).

Ongoing debates suggest it is unclear whether easing licensing requirements could expand the supply of qualified physicians. Advocates argue that easing redundant licensing requirements could expand physician capacity at relatively low cost by activating the capacity of physicians already in the country who are unable to practice (Mansell, 2025), as well as attracting physicians currently practicing abroad (Cicero Institute, 2025). However, skeptics argue that federal visa restrictions may limit the effect of eased licensing requirements (Advisory Commission on Additional Licensing Models, 2025). Even if the flow of entry expands, the stock effect may remain the same if it displaces incumbents (American Osteopathic Association, 2024). Perhaps most importantly, changes in licensing requirement design may change the quality of physicians, either by lowering standards or by changing selection into licensing, which is often difficult to measure ex ante (Federation of State Medical Boards and Intealth and Accreditation Council for Graduate Medical Education, 2024). These competing claims make the effects of easing licensing requirements on physician supply an empirical question. Yet there is little causal evidence on its impact on physician quantity and quality, and even less is known about whether it changes healthcare capacity and downstream public health.

To shed light on this debate, this paper exploits a major wave of licensing reform in which over thirty states removed the requirement that physicians be U.S. citizens to obtain a full medical license between the early 1960s and the 1970s. In 1938, the AMA House of Delegates passed a resolution endorsing citizenship requirements (Stevens and Vermeulen, 1973), and many states' licensing laws required non-U.S.-citizen physicians to become citizens before they could be licensed to practice independently. These restrictions remained in place until the 1960s, when a clear shift toward repeal emerged and continued into the 1970s

(Butter and Sweet, 1977). This large-scale, clear-cut change in licensing law provides a rare setting to study easing licensing requirements. I examine how removing citizenship requirements affects the flow of newly licensed physicians and the licensed stock, and then trace the downstream consequences for hospital capacity, utilization, and population mortality. Because licensing records and physician directories from this period do not record citizenship, I use foreign medical graduates (FMGs) as a proxy for non-citizen physicians and discuss how closely the two groups overlap in Section 3.2.

Understanding the impact on physician composition requires data that distinguishes already-licensed incumbent from new licensees. To my knowledge, there is no existing public data source available that tracks licensing timing at the individual level (Gittleman et al., 2018).[1] I therefore digitize six waves (1964–1981) of the American Medical Directory (AMD), which provide individual-level records for the near universe of physicians in the U.S., including each physician's medical school, practice location, as well as license status and year of licensure. These records allow me both to separate the entry and incumbent margins of physician supply and to measure the stock of licensed physicians directly—a quantity the literature typically overlooks because it relies on total physician density. Using name and medical school information, I am able to construct a panel of individual physicians to examine the effect on attracting new migrants. Although the AMD from the 1990s onward is widely used to study physician labor markets (Wasserman, 2023; Gottlieb et al., 2025), the historical AMD has not been comprehensively digitized.

My identification compares removal states with not-yet-removal states, before and after reform, in a staggered difference-in-differences framework, using the twenty-three states whose removal I observe after the first AMD wave. The identifying assumption requires that absent the removal, the trend in outcomes among treated states would evolve the same way as in not-yet-treated states. Three pieces of evidence support this assumption. First, event-study estimates show no differential pre-reform trends between treated and not-yet-treated states across license flow and hospital outcomes. Second, removal timing is uncorrelated with rising healthcare demand (Medicare and Medicaid expansion) and other state population and economic characteristics. The one factor that does predict early removal—a state's pre-existing FMG presence—is controlled for across specifications. Third, the results are robust to dropping counties with concurrent supply side shocks that are correlated with outcomes of interests, such as Community Health Centers (CHC) and Hill-Burton grants, and to placebo outcomes, including state population and cumulative Hill-Burton funding. With only 23 state clusters divided among eight adoption cohorts, each effect is estimated from a few treated states against a comparison group that shrinks as later states adopt. I therefore check whether

[1] Studies typically infer licensure from occupational affiliation, and even surveys that ask about it directly, such as the Survey of Income and Program Participation (SIPP) and, since 2015, the Current Population Survey (CPS), do not record when the license was obtained.

wild cluster bootstrap confidence intervals are correctly sized in a setting like mine. Following Black et al. (2022), I simulate outcomes with a known treatment effect while holding the actual timing of the reforms fixed, and re-estimate my main specification on each simulated sample (Appendix A.2). Tests based on bootstrap confidence intervals over-reject the true null, but the estimated effect on FMG licensing flows remains statistically significant when evaluated using a size-adjusted critical value.

I first show that states removing the citizenship requirement experienced an economically meaningful increase in FMG licensing flows relative to not-yet-reforming states. Relative to the pre-reform annual flow of total new licensees (FMGs and U.S. medical graduates, or USMGs, combined), the increase corresponds to approximately 17-28 percent expansion. A 28 percent increase represents roughly 930 additional FMG licenses annually across the sample states—equivalent to 38 percent of their annual cohort of USMGs. The increase is detected in both measures of license flow, one I compiled from state licensing statistics and the other from the license timing of individuals in the digitized AMD. The statistically significant effect is detected in the third post-reform year and lasts persistently.

An increase in the licensing flow raises the stock only if new licensees remain in the state and are not offset by exit among those already licensed. In the short run, the reform increased the licensed FMG stock by an amount equal to about 2 percent of the total stock of licensed physicians. Because treated states show differential pre-reform trends in the licensed FMG stock, I assess this estimate using robust confidence intervals (Rambachan and Roth, 2023): it remains statistically significant when post-reform deviations from parallel trends are allowed to be as large as a quarter of the largest pre-reform deviation. In the medium run, the effect is noisy and statistically indistinguishable from zero. To identify which group accounts for this offset, I further decompose the licensed stock by physician group and licensing timing. The decomposition suggests that the offset is concentrated among established FMGs (those licensed before the reform), whose density declines, whereas I find no consistent evidence of crowd-out among USMGs in licensing flows, the aggregate licensed stock, or specialty-specific stocks.

Is the expanded licensed stock explained by attracting more FMGs (migration channel) or activating the license status among those who would otherwise migrate but unlicensed (reclassification channel)? The evidence supports a reclassification channel. I find that the reform raised the share of the physician stock that is licensed. The estimates suggest a statistically significant increase in the share of licensed physicians among all physicians by approximately 3 percentage points. Among FMGs, the license rate increased by 12 to 13 percentage points. I rule out the explanation that the relative increase in license stock outweighs the total stock by showing that no statistically significant increase is detected in the total stock of FMGs (including unlicensed physicians). Migration channel is directly tested by estimating the effect on newly

arrived FMGs and FMGs in training, and the result suggests no statistically significant increase in either measures.

Beyond its effect on quantity, removing the citizenship requirement may also change the quality of the physician workforce. Two channels could be operating: a shift in who selects into licensing when its time cost substantially falls, and the displacement of lower-quality incumbents. Using license exam pass rate at medical school level as a quality measure for FMGs, I show that the licensed FMG stock does not experience a statistically significant or economically meaningful deterioration in the short run; medium-run quality estimates are too sensitive to trend assumptions to be conclusive. Estimate on the quality among new licensees is positive but noisy – the confidence interval is too wide to reject either a large increase or a large decrease in quality.

To understand firms' responses to the physician licensing regulation, I use American Hospital Association Annual Survey to examine the responses in the hospital sector. Evidence suggests that hospitals respond to the removal of citizenship requirement by expanding bed capacity. Short-term general hospitals increased bed capacity by about 4 percent, an effect that is stable in magnitude across control specifications. Utilization effects are less robust: admissions in short-term general hospitals rise in the medium run but are sensitive to the choice of comparison group, and long-term hospital utilization is not robust to controlling for Medicaid.

Because this period of study experienced large shifts in both the demand for and the supply of physicians, I examine whether the effects differ across counties facing different conditions determined at baseline year. I compare counties by the size of the demand shock they faced from the introduction of Medicaid, by amenities that make a county attractive to physicians (income and urban share), and by their pre-existing supply of physicians (physician density), testing whether the effects differ significantly between groups. Hospitals in counties with high Medicaid-eligible population (proxied by the baseline share of women receiving AFDC) expand non-physician employment and beds significantly more than those in low-eligibility counties, whereas no such difference appears for splits by income, urban share, or physician density. This suggests the citizenship requirement relatively constrained hospital expansion where Medicaid increased demand for care, and with it the demand for physicians, rather than relatively constrained hospital expansion where the amenities are less attractive to physicians or lower physician density.

When is removing citizenship requirement an effective policy to expand license flow? Decomposing average treatment effect by calendar time, I find that the effects on FMG license flow and hospital beds become substantially larger and more persistent after 1971, when federal reforms relaxed J-1 return and waiver constraints. This timing is consistent with federal immigration policy complementing state licensing

deregulation and with local payer mix shaping downstream hospital responses.

Finally, I find no evidence of improvement or harm in population mortality. The mortality results are either not statistically significant or sensitive to controlling for Medicaid. Elderly mortality estimates are precise enough to reject any mortality decrease equivalent to the effect of CHC, but not precise enough to reject any smaller but plausible changes driven by the removal of citizenship. Cause of death mortality rates also show no detectable change in mortality.

This paper connects the health economics literature with the labor economics literature on immigration and occupational licensing. Within health economics, it contributes most directly to work studying how physician supply responds to policy-induced incentives and institutional constraints. Prior work shows that physician labor supply responds to financial incentives (Gottlieb et al., 2025), location-based policies (Kulka and McWeeny, 2019), and upstream constraints on residency training capacity (McNamara and Pineda-Torres, 2024). I extend this literature by studying occupational licensing as a downstream regulatory margin that shapes both the supply of licensed physicians and competition within local physician markets.[2]

More broadly, this paper also contributes to work showing that physicians are central inputs in healthcare production. Physicians affect healthcare utilization and spending through practice styles and treatment decisions (Badinski et al., 2023), and patient outcomes depend on physician skill, selection, medical training, and physician–patient concordance (Schnell and Currie, 2018; Alsan et al., 2019; Chan et al., 2022; Clay et al., 2025). I add to this literature by showing that changes in physician supply can also affect hospital capacity. The results therefore speak not only to the labor-market consequences of licensing reform, but also to the role of physicians as a scarce input in the broader healthcare production function.

This paper also contributes to the literature on occupational licensing. A central tension in this literature is that licensing may protect consumers by screening provider quality, but it may also restrict entry and raise incumbent rents. Empirically, a large body of work finds that licensing raises wages or prices and restricts labor supply, while the evidence on quality improvements is more mixed (Kleiner and Kudrle, 2000; Kleiner and Krueger, 2013). Recent work formalizes this trade-off by showing that licensing can simultaneously affect labor supply and labor demand through selection and perceived quality (Kleiner and Soltas, 2023), and physician licensing is a particularly important setting because both provider quality and access to care are first-order welfare concerns (Atal et al., 2025). This paper deepens this discussion by

[2]The two most closely related studies are Noether (1986a) and Peterson et al. (2014). Noether (1986a) studies the same reform—the relaxation of citizenship and other FMG licensing restrictions—but in national time series that aggregate away the cross-state variation exploited here. Peterson et al. (2014) provide the closest research design, a state panel of FMG licensing stringency over 1973–2010, but study a different and later margin (residency requirements) and a single outcome (FMG inflows). This paper revisits that reform with sharper variation than either: staggered state-level removals, and outcomes extending beyond entry to physician competition, hospital capacity, and mortality.

isolating a licensing margin that is not itself a clinical quality screen: citizenship eligibility for licensing. By holding the broader medical training and examination regime fixed, the analysis asks whether a non-quality-related licensing requirement can nonetheless constrain physician supply and healthcare production. In doing so, the paper provides direct evidence that entry barriers embedded within quality-control regimes can have real effects on the supply of physicians and the organization of healthcare delivery.

Lastly, this paper contributes to the immigration literature on immigrant-native substitution and complementarity. Existing evidence is mixed: some work emphasizes displacement among closely competing workers, while other work highlights complementarities, specialization, and demand-side adjustment (Borjas, 2005; Peri et al., 2015; Peri, 2016). I study this question in a high-skill regulated labor market where immigrant labor supply is shaped by both federal immigration policy and state licensing institutions. This complements work showing that immigration policy and visa costs affect the supply of foreign-trained physicians and teaching hospitals' demand for international medical graduates (Lo Sasso, 2021; Braga et al., 2024). Using physician license timing and specialty, I examine competition within county-specialty labor-market cells and show that substitution and complementarity coexist, with the dominant force depending on local market structure. The paper also extends more recent evidence that immigrant healthcare workers relax staffing constraints in nursing and long-term care settings to physicians, with implications for hospital capacity and healthcare production (Kaestner and Kaushal, 2012; Grabowski et al., 2023; Alexander, 2026).

2 Background

2.1 Citizenship Requirement and physician licensing

Citizenship requirements in physician licensure illustrate the tension between the public- and private-interest rationales for licensing. These laws required non-U.S.-citizen physicians to become U.S. citizens before obtaining a full state medical license. This was a non-trivial hurdle for foreign-born individuals. Even for individuals who have already obtained lawful permanent residence in the United States, naturalization typically requires a substantial waiting period. Under the standard pathway, applicants must have been lawful permanent residents for at least five years before becoming eligible to apply for U.S. citizenship.[3] However, nearly half of the physicians admitted to the US in the 1960s entered on the exchange visitor visa (J-1), which additionally required them to return to their home country for two years upon completing their training. In this case, the five-year residency clock started only once they re-entered the US.

[3]This requirement may be shorter under certain circumstances. For example, lawful permanent residents married to U.S. citizens may be eligible to apply for naturalization after three years, provided that they have lived in marital union with the citizen spouse and satisfy the relevant continuous-residence and physical-presence requirements.

Historical accounts link the rise of citizenship requirement to an active campaign by the American Medical Association (AMA). In 1938, the AMA House of Delegates passed a resolution endorsing citizenship restrictions on licenses (Stevens and Vermeulen, 1973). Many states then adopted citizenship requirements that limited licensure for foreign-trained physicians. In public-interest terms, citizenship was framed as a proxy for commitment to ethical standards in a field central to public health (McIntyre, 1939). In private-interest terms, the restriction could also reduce competition from foreign physicians. These dynamics mirror a general pattern in U.S. licensing: Carollo et al. (2025) find that the founding of a state professional association substantially raises the probability of subsequent regulation, and that occupations facing greater immigrant competition were more likely to become regulated, consistent with incumbents' incentives to restrict entry.

Citizenship requirements remained largely stable until the 1960s. A clear shift towards repeal emerged in that decade and continued into the 1970s (Butter and Sweet, 1977). This policy variation was first studied by Noether (1986a,b), using a national time-series correlation between the share of states with a citizenship requirement and physician wages. To date, there is no systematic historical account that articulates what drives the removal of the citizenship requirement across states. I formalize three hypotheses that could plausibly explain the removal combining various sources of qualitative evidence I am aware of. I view those hypotheses as non-exclusive as states have different institutional features that yield different motives. For each hypothesis, I briefly discuss its potential to confound, i.e., to explain both the policy variation and the outcomes of interest. I further discuss how to empirically rule out these concerns in the empirical strategy section (Section 4.3).

Top-down doctrinal and legal shifts. The first hypothesis that is directly discussed historically points to changes in constitutional doctrine regarding alienage classifications. A series of Supreme Court decisions in the 1970s challenged citizenship-based barriers across other occupations, including lawyers, engineers, and civil service jobs.[4] Although these cases were not specific to medicine, they provided legal support for removing citizenship requirements in licensing statutes (Butter and Sweet, 1977) and were echoed in state-level attorney general opinions regarding physician licensure.[5] This mechanism poses comparatively less threat to my empirical strategy because it is less directly tied to healthcare demand and, by itself, does not predict differential changes in health outcomes across states.

Bottom-up mobilization by foreign-trained physicians. A second hypothesis emphasizes political

[4]Key cases include *Graham v. Richardson*, 403 U.S. 365 (1971); *Sugarman v. Dougall*, 413 U.S. 634 (1973); *In re Griffiths*, 413 U.S. 717 (1973); and *Examining Bd. v. Flores de Otero*, 426 U.S. 572 (1976). See https://www.law.cornell.edu/constitution-conan/amendment-14/section-1/alienage-classification for legal commentary.

[5]For example, see Michigan Attorney General opinions *OAG 1971–1972, No. 4755* and *OAG 1972–1973, No. 4776*.

pressure from immigrant physicians and their professional organizations. Systematic documentation of such organizations is limited during the early stage of my study period (1960s), partly because large-scale growth in the foreign-trained physician workforce was just beginning. Yet there are particular cases supporting this hypothesis. For example, in Florida, Cuban physicians organized through the Cuban Medical Association and petitioned for a waiver of the state's citizenship requirement, explicitly challenging alienage barriers to licensure.[6] Nonetheless, the broader role of immigrant professional mobilization becomes increasingly visible in later periods (Bhalla, 2010).

Demand-driven health manpower policy. Beginning in the 1950s, federal commissions and reports heightened concerns about a looming physician shortage. These concerns materialized as the Health Professions Educational Assistance (HPEA) Act of 1963, which expanded federal support for domestic physician training. At the same time, other major policy changes increased demand for care, including the introduction of Medicaid and Medicare in the mid-1960s and the rollout of Community Health Centers. In shortage areas, policymakers explicitly recognized that citizenship restrictions could impede recruitment.[7] This demand-based narrative is empirically relevant because rising healthcare demand could plausibly drive both the repeal of citizenship requirements and increased acceptance of foreign-trained physicians.

Figure 1 summarizes the timing of repeal across states and Table 2 lists the repeal year for each state. Figure 2 maps the geographic pattern of repeals. The geography of citizenship-requirement removals suggests both regional clustering and meaningful within-region timing variation. Early removals were relatively dispersed but appeared disproportionately in the Northeast and Upper Midwest. The largest regional cluster occurred later, between 1969 and 1973, when many Southern, Appalachian, and border states removed the requirement. Among the regions, the Midwest exhibits the most spread-out treatment variation. Midwestern states adopted across nearly the full reform window, from Illinois and Minnesota in 1962, Michigan and North Dakota in 1965, Kansas, Missouri, and Ohio around 1969–1970, Nebraska in 1972, and Wisconsin as late as 1976. Thus, while the South accounts for a prominent late-wave cluster, the Midwest is the region that contributes the most temporally dispersed identifying variation.

2.2 Licensing and Physician Labor Market Competition

Licensing is perhaps the most relevant entry threshold for FMGs to compete with USMGs. Physicians practice in a highly regulated occupation, where nearly every step of the production pipeline is regulated,

[6] "Sparking the action was a petition by the Cuban Medical Association requesting a waiver of Florida's citizenship requirement" (*The Miami News*, 1966).

[7] See, for example, a 1968 newspaper clipping in which a Nebraska senator suggests revisiting license regulation for foreign-born doctors, arguing that it is not in the state's interest to reject well-qualified physicians who have declared their intention to become U.S. citizens, especially during a physician shortage (https://www.newspapers.com/image/313627570).

with state licensure being the final legal threshold for independent practice. Becoming licensed is therefore considered the stage at which labor market competition among independent physicians occurs.[8] Licensed physicians compete with each other in wages, or patients, or both. Physician competition sometimes could be interpreted as conventional labor market competition, where the wage is the result of the market forces. For instance, in the hospital setting, a physician could be a full-time employee of the hospital. However, in most cases in the 1960s, office-based physicians work with the hospital as a platform to produce care, meanwhile contributing to hospital revenue by bringing patients to the hospital. In this case, physicians compete for the patients in the local market and also for the admission privilege of a hospital. Without a full license, physicians may still work in limited settings using less-than-full licensure,[9] but their scope of practice is restricted, and in training settings their care provision must be supervised by a licensed physician(Butter and Sweet, 1977, pp. 331–332).

Empirically, this licensing margin also appears important for foreign-trained physicians. Using the 1964 AMD data, I show that FMGs were substantially less likely than USMGs to hold state medical licenses. Appendix Figure A9 shows a snapshot of the licensing gap between foreign-trained and U.S.-trained physicians by the year since they graduated from medical school. This figure demonstrates a substantial licensing gap between foreign and US trained physicians across the career trajectory. Within 5 years of graduation, nearly 90% of USMGs are licensed, while less than 10% of FMGs are licensed. The growth in licensing rate upticks rapidly starting in year 5 and lasts until 10+ years since graduation. Admittedly, the growth in license rate could in part reflects the stayers selecting into licensing. However, the gap between USMG and FMG is still more than 20% across the experience distribution. The gap persists even after 20 years since graduation, which might be explained by low returns to licensing status in the later stages of the career.

Licensure, however, was not the only constraint facing foreign-trained physicians. Even after satisfying state licensing rules, FMGs also needed legal authorization to remain and work in the United States. Licensure itself does not require work authorization, but employment does. As a result, some licensed FMGs may have been unable to practice full-time or to remain at a facility beyond their visa expiration dates, which would attenuate the effect of licensed physicians on healthcare production.[10]

[8] In comparison, the post-graduate training stage is an earlier phase of competition among physicians; however, because the provision of care hinges on the number of residency and internship positions and the scope of practice is supervised by licensed physicians, the potential impact on care production at this stage is constrained by the supply of residency programs rather than by physicians' independent location decisions. See Gottlieb and Nicholson (2026) for the framework regarding physician competition.

[9] For example, in certain government or state hospitals, and in some contexts tied to medically underserved areas.

[10] Unlike many high-skill occupations where employer-sponsored temporary visas (e.g., H-1B) are a common pathway, foreign physicians historically relied heavily on permanent residency or on J-1 exchange-visitor visas. The J-1 pathway permits working but typically limits physicians to designated institutions and is accompanied by a home-country return requirement, making long-run career establishment difficult without adjustment of status. Public Law 91–225 (April 1970) relaxed this constraint on two margins: it narrowed the return requirement to exchange visitors who were government-financed or whose field was designated on the State

To what extent the visa constraint could mitigate the impact of citizenship removal depends on whether the facilities can leverage the full scope of practice of physicians on non-work authorization required positions. In fact, Stevens and Vermeulen (1972) document that "the line between education and staffing in relation to hospital interns and residents is not always clear" and that the exchange visitor visa "became the common vehicle for American hospitals to import foreign trained house staff, nominally on a disinterested, educational basis" (pp. 54–55). By the early 1960s, approximately 67 percent of all FMGs in internship and residency positions held J visas (p. 57 n. 18), making the J-1 pathway a de facto mechanism for hospital staffing rather than strictly an educational program.

One might consider competition with other occupations who can also provide physician service. This paper focuses on competition among allopathic physicians because, during the 1960s and 1970s, other occupations that now substitute for physician services were either small, institutionally segmented, or not yet central to medical care delivery.[11] As a result, the estimates are best interpreted as capturing the effects of FMG licensing reform within the physician labor market, with limited confounding from cross-occupation substitution.

2.3 Hospital Production Function and Market Segmentation

Hospital input can be modeled as physical capacity (staffed beds) and other non-physician staff (nurses and technicians), and they enter the production function of hospital care together with the availability of physicians. One key assumption is that the Hospital beds are not a fixed, exogenous capacity that drives utilization. They are an endogenous input, a derived demand that responds to the supply of complementary physician labor (Pauly, 1980; Harris, 1977, ch. 3).

2.3.1 Hospital-based vs office-based physicians

The distinction between licensed and unlicensed physicians has concrete implications for their role in hospital production: full licensure lifts the supervision requirement described above, expanding the independent labor input the hospital can deploy per physician on staff. These implications differ across healthcare

Department's Exchange Visitor Skills List, and it broadened the grounds for waiving the requirement (including a home-government "no objection" statement), easing conversion from exchange-visitor to immigrant status. This relaxation operated from 1970 until Congress imposed stricter, physician-specific rules in 1976 (effective January 1977), making participants in graduate medical education categorically subject to the return requirement and ineligible for the no-objection waiver (U.S. General Accounting Office, 1977, pp. 6–7).

[11] DOs are now highly integrated with MDs in the physician labor market, especially after the 2020 transition to a unified graduate medical education accreditation system. In the 1960s, however, DOs remained a small and separately organized segment of the physician workforce and were catalogued separately from the AMA's *American Medical Directory*. NPs and PAs were also still emerging: the first PA training program was established at Duke University in 1965, and NP scope-of-practice expansions largely occurred later.

sectors between FMGs and USMGs. Licensed FMGs could imply a large impact on hospital production given their predominant role in hospital-based positions. Data from the 1960 Census (5 percent sample, IPUMS) confirm this pattern. Foreign-born physicians are more than twice as concentrated in hospital-based employment as in non-hospital-based practice. 27.1 percent of physicians working in hospitals are foreign born, compared to 11.6 percent of those in non-hospital-based practice. Although hospital employment accounts for only 29 percent of physicians in health-services industries, it absorbs nearly half of all foreign-born physicians.

This sorting across employment settings has direct implications for physician competition and for where reform effects take place. Hospital-based, ward-serving physicians predominantly treat patients covered by public insurance programs such as Medicaid, while office-based physicians predominantly serve privately insured patients. Historically, 15 to 30 percent of private physicians accepted no Medicaid patients at all, with a larger fraction imposing volume limits (Sloan et al., 1978; Mitchell, 1991; Zuckerman et al., 2004). The consequence was that Medicaid enrollees continued to present at hospital outpatient departments and emergency rooms — the institutional settings they had used before Medicaid — rather than being absorbed by office-based practices. By 1972, payments for short-term care in general hospitals accounted for nearly 40 percent of total Medicaid expenditures, a figure that reflects this hospital orientation of Medicaid utilization (Davis and Schoen, 1978; Engel, 2006). These two settings constitute partially segmented markets that are driven by the price-setting feature of different payers. Reimbursement rates for publicly insured patients are determined by the state government and are thus exogenous to the local market. Prices for privately insured patients can be adjusted in the local market.

2.3.2 Short-term hospitals vs long-term hospitals

Short-term general hospitals were the core of acute inpatient care in the 1960s, treating medical and surgical conditions with average stays under 30 days. Long-term hospitals kept patients for 30 days or more and included psychiatric, tuberculosis, geriatric, and chronic-disease hospitals, as well as long-stay general facilities that were often large federal Veterans Administration complexes. Both types required organized medical staffs, but physician supply was more linked to short-term hospitals' ability to admit successive acute cases and operate staffed beds whereas physicians in long-term institutions supervised extended treatment programs for a more policy and institution determined demand (National Center for Health Statistics, 1965; Sirrocco, 1972). This is also reflected in their ownership structure. Short-term hospitals were mostly comprised of hospitals owned by non-profit organizations, while long-term general were mostly government owned. According to the AHA data, in 1965, near 68% of long-term general

beds were in state psychiatric hospitals, 4% in TB sanatorium, and in other cases are facilities treating chronic-disease or county infirmaries. Historically, state mental hospitals struggled to recruit U.S.-trained physicians to comparatively unattractive institutional positions and consequently relied heavily on foreign medical graduates (Lockett and Williams, 1974; Knesper, 1978; Hirshbein, 2022).

3 Data

3.1 License Policy and Data on License Granted

I digitized the licensing requirements for permanent licenses among FMGs and USMGs, as well as the number of licenses granted from two sources: 1) annual series of medical licensure statistics in the Journal of the American Medical Association during 1961–1973, and 2) annual series of Physician Distribution and Medical Licensure in the US during 1974–1979. Both data resources are from the same collecting agent, the AMA, with a conversion in publication form from JAMA to a book series in 1974. Starting in 1931, the Council on Medical Education and Hospitals of the American Medical Association has dedicated itself to compiling the data every year from each medical board in the US into an annual publication. Because states set their own licensing requirements and, to my knowledge, kept no public records of them, the AMA series is the most comprehensive source for this period.

Two states' removal years are ambiguous in the AMA series, and I resolved them from state statute. Montana's amendment removing the citizenship requirement for a physician's certificate to practice medicine, and adding a requirement of being able to communicate in English, is Sec. 2, Ch. 323, L. 1979, so I date Montana's removal to 1979. Wisconsin's citizenship requirement is no longer mentioned in Chapter 383, Laws of 1975, so I date its removal to 1976. Because these are the two latest removals, they form the not-yet-treated comparison pool for the medium-run estimates; Section 8 examines sensitivity to each. Table 2 lists the resulting removal year for every state, with statute-derived dates marked, and Figures 1 and 2 plot them.

I focus on collecting the removal of citizenship requirements starting from 1960, since it marks the beginning of the general trend toward removing the citizenship requirement (Butter and Sweet, 1977). The ending point year is 1979, by which all states had removed the requirement. States that removed the citizenship requirement before 1960 are considered already-removed states and will be excluded from the analysis.

To measure the number of licenses newly granted at the state level, I use the number of licenses granted to those who have secured their first medical license. Therefore, it excludes licenses for physicians already licensed and practice elsewhere, as well as the case of license renewal. I divide the count measure by the

number of residents in each state, so that I can interpret the estimates as a comparable metric across smaller and larger states.

3.2 Physician Supply Data

Prior studies generally rely on the Area Health Resource File to measure physician supply (Finkelstein, 2007; Bailey and Goodman-Bacon, 2015), which contains aggregate count of physicians at the county level. However, it lacks the information on licensing and foreign-training status, making it impossible to identify foreign-trained physicians.

To address this limitation, I digitized physician-level data using the American Medical Directory (AMD). The AMD dataset is periodically maintained by American Medical Association since 1906 as a record of the universe of physicians—those who have completed a medical degree—in the US. AMA collects the information by sending out the survey to each physician, and combining it with information obtained from other official agencies, such as medical schools and state licensing boards.[12] It is considered the most comprehensive and complete source of physician data, and has been used in both modern literature (Wasserman, 2023) as well as historical studies (Mitchener et al., 2022; Hollingsworth et al., 2024; Alsan and Neberai, 2024). Appendix Figure A10 illustrates the structure of the data - each physician is listed under the location they provided as contact, and for each physicians, it reports their name, birth year, medical school and year of graduation, year of licensing if licensed, reported major specialty and type of practice (full time patient care, residency/intern, research, faculty, administrative staff, or retired). I digitized six waves of data including 1964, 1966, 1969, 1973, 1979, 1981. In the baseline year (1964), there are 270,797 physicians in the US, and 11% of them are FMGs.

A unique feature of this dataset is it allows me to separate the physicians based on their medical school training. I define foreign-trained physicians (FMGs) as those who graduated from a non-US medical school, and US-trained physicians (USMGs) as those who graduated from a US medical school. This allows me to unpack the impact on the supply of foreign-trained physicians specifically. Because neither the AMD nor the state licensing statistics record citizenship, I use FMG status as a proxy for the non-U.S.-citizen physicians the requirements restricted. Contemporary estimates suggest the proxy is close. Among FMGs who entered the United States between 1966 and 1968, only 16 percent were U.S. citizens (6 percent by birth and 10 percent naturalized), while 68 percent were citizens of a foreign country and 16 percent could not be classified (Butter, 1971). Additionally, using the detailed licensing data for each physician, I can measure the license flow at local level based on the year of licensing, which complements the state-year licensing data

[12]The response rate in 1967 is about 87.4%

from state medical board. Finally, the panel structure of the data allows me to track individuals over time, which enables me to exploit the mechanism of migration and license renewal to explain the stock effect. Following the algorithms developed by Abramitzky, Boustan, and Eriksson (ABE, 2021), I link individuals across waves based on their first name, last name, and the medical school they graduated from. The linkage rates are about 91% for USMGs and 80% for FMGs.

Specialty groups are constructed from physicians' self-designated specialty codes in the AMD. Primary care comprises general practice, family practice, internal medicine, pediatrics, and obstetrics-gynecology. Hospital-based ("platform") specialties comprise anesthesiology, radiology, and pathology. Surgeons comprise general surgery and its subspecialties (orthopedic, otolaryngologic, urologic, plastic, colorectal, neurologic, thoracic, cardiovascular, and pediatric surgery).

The outcomes used in my analysis include the number of recently licensed physicians (flow) and number of licensed physicians (stock). State/county level outcomes are aggregated to state/county level using all individuals reported to be in the state/county. To unpack the license stock, I use whether a physician was already licensed in the state in the previous wave to measure the proportion of recent licensees that update or renew licenses. I also use retirement status to measure the number of non-recent licensees that exit the labor market. I convert all count variables to per capita terms using the state or county population from Ruggles et al. (2010) and Surveillance, Epidemiology, and End Results (SEER) Program (2009), and use them in per capita levels (per 10,000 residents) in the analysis.

3.3 Hospital Supply and Capacity Data

I use annual hospital-level data from 1956 to 1975 from the surveys of the American Hospital Association (AHA) for every AHA-registered hospital in the US. Following Finkelstein (2007), I exclude around 5% hospitals that are federally owned in the analysis, since those hospitals are rarely staffed by foreign medical graduates. To study the aggregate impact at local level, I aggregate the observations to state or county level based on the location of the hospital.

I use the total number of hospitals, beds, and employment to measure hospital supply and capacity. Following Finkelstein (2007), admissions are defined as total inpatient admissions over a twelve-month period, excluding newborns. Patient days measure inpatient utilization, constructed as the hospital's average daily census—the average number of inpatients receiving care each day during the year, excluding newborns—multiplied by 365. Both measures capture inpatient activity only and neither includes any outpatient-department care. Hospital counts are decomposed into different ownership types, including nonprofit, for-profit, and government-owned hospitals. Beds and employment are used to proxy for the capacity

of hospitals in the area. Employment data from AHA excludes physicians, and includes the employment of nurses and technicians. Because the AHA bed measure counts beds set up and staffed for use, the estimated increase reflects the activation of operating capacity through labor, not new construction.

3.4 Mortality data

To construct the elderly mortality rate, I use the annual Vital Statistics Multiple Cause of Death data from 1959 to 1979 from National Center for Health Statistics (1979), which contains the universe of death records reported by cause, age and county/state of residence. I combine the death counts at state or county level as numerator and the population of 5-year age bin group starting 65 years old from Ruggles et al. (2010) and Surveillance, Epidemiology, and End Results (SEER) Program (2009) as denominator to obtain the age-adjusted mortality rates. Appendix A.5 describes the construction of the age-standardized rates and lists the ICD code ranges defining each cause-of-death group, including the coding harmonization between two ICD codes.

3.5 Medicaid Exposure Measurement

To examine heterogeneity by local exposure to public-insurance demand, I classify counties by their pre-1965 Medicaid exposure to avoid endogeneity in participation across counties post Medicaid reform. I measure exposure using county-level participation in Aid to Families with Dependent Children (AFDC) among women in 1964, the year before Medicaid was enacted.[13] AFDC participation proxies for early Medicaid exposure because the original Medicaid program extended eligibility largely through pre-existing welfare-based (AFDC) categories (Goodman-Bacon, 2021b). I define a county as high-Medicaid-exposure if its 1964 AFDC participation among women is above the sample median, and low-Medicaid-exposure otherwise. This binary split at the median defines the county groups used throughout the heterogeneity analysis. Appendix Figure A11 shows the divergence in per-capita Medicaid transfers between the two groups over time, confirming that the 1964 AFDC split captures subsequent differences in Medicaid intensity.

3.6 County Partitions for Heterogeneity Analysis

Several results compare treatment effects across different types of local health care market. I use four partitions of the county sample, each based on a characteristic measured in 1964 so that group membership is fixed before any state removes its citizenship requirement. Three are measured directly from the county

[13]The 1964 county-level AFDC participation rate is made possible thanks to their digitization effort in Cunningham and Goodman-Bacon (2025).

data: the share of the population living in urban areas, county income, and physicians per capita. The fourth is expected exposure to the Medicaid expansion, measured by the 1964 AFDC participation rate described in Section 3.5. Because all four are determined at baseline, they partition the covariate space rather than responding to treatment.

4 Empirical Strategy

4.1 Sample Selection

The final estimation sample includes twenty-three states that removed a citizenship requirement between 1965 and 1979. I restrict the sample to states with a clean removal of a citizenship requirement observed after the first AMD wave in 1964. This restriction excludes states without complete AMD coverage, states outside the contiguous United States, states with multiple adoption and repeal episodes, early reform cohorts without a pre-reform AMD baseline, states whose reform relaxed a requirement other than citizenship (a blanket exclusion of foreign medical graduates, or permanent residency), and states that had no citizenship requirement in place by 1960. Table 2 lists every state by removal year, including the eight 1961–1964 removers dropped for lacking a pre-reform AMD baseline and the fourteen never-restricted states; its notes record the state-specific exclusions. The twenty-three analysis states are those removing between 1965 and 1979, less Alaska, for which AMD coverage is incomplete.

Excluding already-removed states is not only a conceptual issue, but also empirically plausible given the differential trend and level in certain charasteristics: compared with analysis states, already-removed states had higher incomes, more medical doctors, more foreign physicians, and more medical trainees, as shown in Table A5. They also followed different pre-period trends in FMG licensing, as shown in Appendix Figure A12. Including these states as controls could therefore violate the parallel-trends assumption, especially after the Hart–Celler Immigration and Nationality Act of 1965 increased foreign-physician inflows into states that were already open to noncitizen physicians.[14]

4.2 Estimation Model

I use the doubly-robust staggered adoption estimator of Callaway and Sant'Anna (2021), which addresses well-known issues with two-way fixed effects under heterogeneous treatment timing.[15] Each state is

[14] Because already-removed states were better positioned to absorb foreign-trained physicians after the 1965 immigration reform, using them as controls could generate a widening post-1965 gap that would be incorrectly attributed to citizenship-requirement removal.

[15] A Goodman-Bacon (2021a) decomposition of the TWFE estimator for the main outcome (Appendix Figure A13) shows that comparisons against not-yet-treated controls center around 0.1, while negatively weighted comparisons against already-treated units

assigned a treatment cohort $G_j = g$ corresponding to the year its citizenship requirement was removed, as recorded in the JAMA annual medical licensure statistics series and verified against state statutory records. The comparison group consists of not-yet-treated states ($G_j > t$): states that will eventually remove the requirement but have not yet done so by period t.

The target estimand is the average treatment effect on the treated for cohort g at time $t \geq g$:

$$ATT(g,t) = E\big[Y_{j,t}(g) - Y_{j,t}(\infty) \mid G_j = g\big] \tag{1}$$

where $Y_{j,t}(g)$ is the potential outcome for a state first treated in period g, and $Y_{j,t}(\infty)$ is the counterfactual outcome under no treatment. Under parallel trends, $ATT(g,t)$ is identified by

$$ATT(g,t) = E\big[\Delta Y_{j,g-1,t} \mid G_j = g\big] - E\big[\Delta Y_{j,g-1,t} \mid G_j > t\big] \tag{2}$$

where $\Delta Y_{j,g-1,t} \equiv Y_{j,t} - Y_{j,g-1}$ is the change in outcomes relative to the period immediately before cohort g is treated.

The licensing-flow outcomes from the JAMA series are observed annually, so t indexes calendar years. The AMD-based outcomes (physician stocks and license cohorts), by contrast, are observed only in survey waves rather than annually. Six waves were digitized (1964, 1966, 1969, 1973, 1979, and 1981); the analyses use the 1964–1973 waves, because once Wisconsin (1976) and Montana (1979) remove the requirement, no not-yet-treated comparison states remain. For these outcomes I estimate the event study in *wave time*, so that t and g index AMD waves and event time is $e = t - g$. Because the waves are unevenly spaced and a reform can occur anywhere within the calendar span a wave covers, each event-time coefficient e corresponds to a *range* of calendar years relative to treatment rather than a single year. The horizontal-axis labels in these plots report this empirical calendar-year range across the treated cohorts contributing to each bin. For example, $e = 1$ pools observations measured one AMD wave after treatment, which can correspond to anywhere from 2 to 9 calendar years after the reform. Hence, the bin is labeled $[2, 9]$. Throughout, I restrict the displayed event-study window to event times -5 to $+5$, the range over which a sufficient share of treated cohorts contributes to each estimate.[16]

Appendix A.3 quantifies what this sparse measurement costs by simulating the flow outcome at annual and four-wave frequency under the same design. The wave-based estimand is essentially unchanged unless

pull the naïve TWFE estimate to -0.01.

[16] Appendix Figure A14 plots the proportion of treated units contributing at each event time for the annual and AMD samples, with the displayed window shaded. At longer leads and lags the panel thins as fewer cohorts are observed, making the estimates less representative.

effects emerge late in the sample and the estimator remains unbiased, but power falls substantially—from 0.99 to 0.84 for a constant effect of the estimated size, and to 0.39 for an effect that begins only in 1971—so null or marginal wave-based estimates indicate imprecision rather than the absence of an effect.

The preferred specification uses the doubly-robust conditional version of this estimator, which controls for pre-treatment state characteristics X_j via regression adjustment and propensity-score reweighting (Callaway and Sant'Anna, 2021).[17] Instead of adding time-varying state-level controls as "bad controls", I use the pre-treatment characteristics following the consensus of the staggered difference-in-differences literature.[18] The baseline covariate is the ratio of FMGs among trainees in 1964, which is the major explanatory factor that predicts removal timing and outcomes. Regression adjustment models untreated outcome trends as a function of this baseline covariate among not-yet-treated states. Inverse probability weighting balances the covariate distribution across treatment cohorts and not-yet-treated controls.

Each results table reports the baseline doubly-robust specification alongside the additional baseline controls when specified. Each figure reports the baseline doubly-robust specification, and additional controls are added in the robustness checks.[19] In the robustness checks, I also vary the model by adding alternative measures of FMG presence in the physician labor market to confirm that the results are not sensitive to this choice of variable.

The aggregated event-study estimates are obtained by weighting each $ATT(g,t)$ by the share of treated states in cohort g observed at event time $t-g$, as recommended by Baker et al. (2026) to avoid the negative implicit weights of two-way fixed-effects estimators.

When outcomes are licensing statistics or physician supply, I weight observations by state population at baseline so that aggregate estimates reflect the impact on the average individual rather than the average state. For mortality outcomes, I weight by the population of the corresponding age group.

The 95% confidence intervals are clustered at the state level, using a wild cluster bootstrap to accommodate the small number of clusters. Appendix A.2 evaluates whether this inference is reliable in finite samples. Holding the actual treatment configuration fixed and simulating outcomes with a known effect, the estimator is approximately unbiased, but conventional tests reject a true null about 20 percent of the time at

[17] Figure A15 shows the physician labor market estimates mostly remain similar without the baseline control. This suggests that the baseline presence of FMGs is not correlated with subsequent changes in physician supply. By contrast, Figure A16 shows the hospital bed expansion becomes close to zero without control, suggesting the more FMG present states are correlated with a decrease in beds. Hence, controlling for it removed the omitted variable bias that downward-bias the causal estimates.

[18] Baker et al. (2026) discuss the post-treatment bias that arises if such controls themselves respond to the treatment.

[19] The reason for incorporating at most two covariates at a time is that a large proportion of the variation in outcomes is explained by two controls. I regress treatment timing as well as the main outcome of interest, FMG licensure flow, on baseline controls cumulatively in the 1964 cross-section, reporting the marginal R^2 gain from each addition. For FMG licensure flow, the baseline FMG trainee ratio alone explains 68% of the variance, and adding Medicaid adoption year and Medicare share raises this to 90%; each subsequent control adds at most 2.4 percentage points.

the nominal 5 percent level, and the distortion is not removed by bootstrap standard errors. I therefore report size-corrected critical values from the empirical null distribution: approximately $|t| = 2.8$ for annually observed outcomes and, for the flow outcome observed at the four AMD waves, $|t| = 3.7$, which I apply to the other wave-based outcomes (Appendix A.3). The results sections note which estimates exceed these thresholds.

4.3 Identifying Assumptions

The key identifying assumption is that, absent the removal of the citizenship requirement, outcomes would evolve in parallel between treated and not-yet-treated states. Since the counterfactual is unobservable, this assumption cannot be directly tested. I validate it through three complementary approaches: balance tests on potential confounders, event-study specifications testing for pre-reform parallel trends, and a conditional estimator that controls for baseline FMG exposure.

For demand-side policies, I include the year of Medicaid introduction and the share of uninsured individuals aged 65 and older (Finkelstein, 2007; Finkelstein and McKnight, 2008) as predictors of removal timing. These factors could be a threat if increased demand itself facilitated the integration of foreign physicians. According to Table 1, these factors do not significantly predict removal timing.[20] I also test whether state characteristics (income, demographics), health outcomes (infant and elderly mortality), and total physician supply per capita predict earlier removal. The results suggest these play a limited role. For county-level infrastructure shocks, including the introduction of community health centers (Goodman-Bacon, 2021b) and Hill-Burton funding (Chung et al., 2017), I conduct robustness checks by excluding counties experiencing such construction during the period of study.

For immigration policy, the Hart-Celler Immigration and Nationality Act of 1965 may have differentially increased physician inflows across states. I address this by directly estimating the effect of citizenship removal on the arrival of new foreign-trained physicians in Appendix Table A6.[21] The estimated effect is not statistically significant, suggesting that differential inflows of new arrivals are not the main channel driving divergence between early- and not-yet-removal states.

One factor does stand out as a predictor of removal timing: the existing size and share of FMGs in a state. Appendix Figure A17 indicates that the foreign physician pool explains a substantially larger share of timing variation than any other predictor, suggesting that greater FMG bargaining power or workforce reliance

[20] Appendix Figure A17 displays the R^2 for each variable, indicating that these factors explain only a small share of the timing variation.

[21] I also include the proportion of foreign-born residents as a predictor of removal timing in Table 1, which appears to be a minor factor.

facilitated the reform. I now explain why this is unlikely to threaten identification. For FMG presence to confound the estimates, it would need to independently affect outcomes even absent the citizenship reform—for instance, through relaxation of other licensing requirements or expansion of employment slots for foreign medical graduates. The evidence suggests neither channel is active. Other licensing rule changes, such as Endorsement or FLEX exam adoption in Table 1, barely predict the timing variation. The estimated effect on the stock of foreign trainees (unconditional on license status) is insignificant in Appendix Table A6. Taken together, this evidence supports the conclusion that citizenship removal is the sole policy change captured by the timing variation.

As a visual complement to the balance tests above, Figure A18 plots raw population-weighted outcomes by adoption cohort, with thicker segments marking years after each cohort group's first adoption. FMG licensing flows are initially low and similar across all groups. After adoption, treated cohorts diverge. Earlier adopters begin expanding sooner and accumulate substantially larger inflows, concentrated in the early 1970s, while later adopters see smaller increases and the latest-treated comparison group remains flat throughout except in 1971. Hospital capacity expands more broadly and gradually. All groups add beds through the 1960s, but after 1970 the treated cohorts continue expanding while the substantially higher-capacity comparison group flattens since 1969.[22] This cohort-level raw-trend plot provides support for the parallel trends assumption that the formal event studies below confirm.

These balance tests address concerns about demand-side and immigration confounders. Event-study specifications reported in the results show no significant pre-trend in license flow and hospital beds outcome, providing direct support for the parallel trends assumption. The conditional estimator further addresses potential unobservables by reweighting observations to balance the baseline FMG trainee ratio across treatment cohorts and not-yet-treated controls.

4.4 Heterogeneity by County Type

To estimate effects by county type, I apply the same staggered difference-in-differences estimator and aggregation rule separately within each predetermined subgroup across states, using the baseline partitions defined in Section 3.6. Under parallel-trends and overlap assumptions within each subgroup, the resulting estimates identify the group-specific average treatment effects of removing the citizenship requirement. Following Baker et al. (2026), this approach treats baseline county characteristics as predetermined partitions of the covariate space and estimates partition-specific treatment effects. I assess heterogeneity by testing

[22] An alternative explanation is the ceiling effect is confounding the estimates. Since the latest-treated states are also those with the highest baseline bed capacity, their growth may have been approaching a ceiling for reasons unrelated to licensing. I discuss this in Appendix A.4 using a synthetic control method and find this explanation fits the patterns in some states but not every state.

the null that these effects are equal between two partitions, using a Wald test based on a state-clustered covariance matrix that preserves covariance across subgroups (Chow, 1960; Cameron and Miller, 2015).

For the county-level analysis the sample is restricted to counties that had at least one hospital and a nonzero FMG physician stock in 1964, so that estimates are based on a sample that have healthcare infrastructure and contact with FMG. This reduces the county universe from 1,650 counties in the underlying panel to the 443 analysis counties (in 22 states). All county specifications additionally control for the county's 1964 nonwhite population share as a time-invariant covariate in the doubly-robust adjustment, since nonwhite share is the single most imbalanced pre-determined characteristic.

I report these estimates for physician supply in Section 5.2.5 and for hospital outcomes in Section 6.3.

5 Quantity and Quality Effects on the Physician Workforce

5.1 Effects on FMG Licensing Flow

5.1.1 Main Results

Table 3 reports average treatment effects on the number of first-time FMG licensees and the number of FMGs licensed in year t at the state-year level, with event study plots in Figure 3. First of all, the event-study plots in Figure 3 indicate no significant pre-reform differences between treated and not-yet-treated states, mitigating concerns about violations of the parallel trends assumption.

Columns (1)–(3) of Table 3 show the estimates for first-time FMG licensees per 10,000 population. Column (1) shows that removing the citizenship requirement significantly increases the number of first-time FMG licensees. The estimate remains positive and statistically significant when additionally controlling for the timing of Medicaid adoption in column (2), suggesting that the increase is not driven by contemporaneous changes in healthcare demand associated with Medicaid expansion. Column (3) further controls for the share of uninsured individuals aged 65 and older, and the estimated effect remains statistically significant. The overall effect also exceeds the size-corrected critical value from Appendix A.2 ($|t| = 3.4$ against 2.83), so the first-stage result is not an artifact of over-rejection with 23 clusters.

The magnitude is economically meaningful. For first-time FMG licensees, the estimated increase ranges from 0.086 to 0.116 licensees per state-year, equivalent to a 21 to 28 percent increase relative to the pre-reform annual license-flow benchmark. Based on a 80 million population among states of analysis, 0.116 corresponds to an increase of 930 physicians licensed per year. Compared to approximately 2,400 US medical graduates among the states of analysis, this expansion is about a 38% expansion in the graduation

cohort.

The gradual post-reform increase in licensing flows suggests that removing the citizenship requirement likely induced more foreign-trained physicians to prepare for and obtain state licenses, rather than simply granting licenses immediately to physicians who were otherwise fully qualified but awaiting citizenship.

When using an alternative measure on license flow, Columns (4)–(6) in Table 3 show similar evidence. The outcome is constructed using the number of FMGs licensed in year t based on the individual-level AMD directory. The estimated effects range from 0.065 to 0.086, or roughly 17 to 22 percent relative to the corresponding annual license-flow benchmark.

5.1.2 Timing Heterogeneity

This subsection examines heterogeneity in the licensing-flow effect by calendar year, to assess whether state licensing barriers interacted with federal visa constraints. Panel (a) of Figure 7 shows the Callaway–Sant'Anna group-average ATT for FMG licensing flow aggregated by calendar year. Calendar year 1972 onwards exhibits substantially larger and more persistent effects relative to earlier years.

This pattern aligns closely with a concurrent change in immigration law. As documented in Section 2.2, the J-1 exchange-visitor visa was the primary entry channel for foreign-born physicians to enter the US, but it imposed a two-year home-country return requirement that creates more uncertainty for immigrants to establish and remain the US. Public Law 91-225 (1970) limited the two-year home-residence requirement to exchange visitors in government-financed programs or fields needed by their home countries and expanded the grounds for obtaining a waiver, easing some visitors' path to permanent residence. Waiver approvals rose sharply thereafter (Mejia et al., 1980), and the number of foreign physicians adjusting from exchange-visitor to immigrant status rose from roughly 890 in 1970 to 4,140 in 1973 (Butter and Sweet, 1977). The concentration of estimated effects after 1971 provides implications that relaxing J-1 return constraints may have been an important complementary condition for citizenship requirement removals to translate into increases in FMG licensing. Section 6.2 shows the downstream hospital response follows the same calendar-year pattern.

5.2 Effects on FMG Stock

A natural question is whether the increase in licensing flows translates into a larger stock of licensed physicians. I expect a mild impact on the total stock for two reasons. First, flow is a small proportion of the stock. In 1960s, there are around 300,000 physicians in the US while one graduation cohort is around 6,000. Relative to the total stock of licensed physicians, the estimated flow increase in Section 5.1 is equivalent to

roughly 0.7 to 1.3 percent of the total licensed physician pool. Second, licensed physicians might not stay in the US, especially if the visa status is uncertain.

5.2.1 Licensed Stock

Table 4 and left panel of Figure 4 show the event-study estimates on the stock of licensed FMGs. Consistent with the early adopters having a higher presence of FMGs discussed in Section 4.3, pre-treatment coefficients suggest that even after controlling for the baseline presence of FMGs, treated states were still experiencing a statistically significant increase of 0.057 before the removal. Since this differential has the same sign as the post-removal estimate, continuation of the pre-existing trajectory would bias the estimated policy effect upward. Hence, I interpret the post-removal coefficient may indicate an acceleration following the reform, but the full 0.187 estimate cannot be attributed to citizenship removal without additional assumptions about how the pre-treatment differential would have evolved.

To evaluate the sensitivity of these estimates to violations of parallel trends, I follow Rambachan and Roth (2023) and construct robust confidence sets that allow the maximal post-treatment violation to be no larger than $\overline{M}$ times the maximal pre-treatment violation. Appendix Figure A19 reports the robust confidence sets for $\overline{M} \in \{0.25, 0.5, 0.75, 1\}$. The result suggests the largest level of violation that allows the confidence interval to exclude zero is $\overline{M} = 0.25$. In other words, a statistically significant short-run increase survives when allowing post-period parallel-trends violations up to a quarter of the largest pre-period deviation.

Compared to the total stock of licensed physicians, a 0.187 increase corresponds to about 2 percent of the total stock. The magnitude of the short-run effect remains positive and statistically significant across specifications. In comparison, the medium-run effect remains statistically insignificant across specifications, as shown in Table 4 and Appendix Figure A19.

5.2.2 Share of Physicians Licensed

Table 5 shows that removing the citizenship requirement significantly increases the share of the physician stock that is licensed. Panel A shows that across specifications, the licensed rate among FMGs increase by 12 to 13 percentage points, which is equivalent to a 28 to 30 percent increase relative to the baseline mean of 44 percent of FMGs are licensed. Panel B further shows this also affected the license rate among the total physician population. In the short-run, removing citizenship requirement increases the license rate among all physicians by 2.6 to 2.9 percentage points, or about 3 percent relative to the baseline mean of 87 percent.

5.2.3 Total FMGs and New Arrivals

The short-run rise in the licensed FMG stock is reclassification, not migration. The total number of FMGs present in a state, licensed or not, does not change after removal, and neither do two proxies for new arrivals; only licensure status moves. This subsection documents that result.

I estimate the effect on three outcomes of interest. First is total number of FMGs (including unlicensed ones) to directly test the impact on the pool of FMGs. Table A7 shows the event study coefficients are in the opposite direction of licensed outcome and statistically insignificant across specifications. Consistent with the reclassification explanation, panel A shows total number of FMGs experience no statistically significant change post-removal.

The second and third outcomes are the number of newly arrived FMGs (who can not be matched to the previous wave of AMD) and the number of FMGs in training. Both measures are a proxy for the pool of newly arrived FMGs. Panel B and C in Table A7 further validates the limited impact on the new migrants margin, i.e. no equivalent magnitude and statistically significant impact is detected for newly arrived FMG and FMGs trainees.

5.2.4 Decomposition by License Recency

This section unpacks the compositional changes among licensed FMGs to explain the medium-run null impact on licensed FMGs discussed in Section 5.2. Theoretically, the dissipation in the medium run could be driven by crowding out incumbents, recently licensed FMGs, or the exit of newly licensees.

I separate these margins by the recency of licensed status measured relative to the AMD wave dates. Specifically, newly licensed physicians were defined as those first licensed in the interval between the current wave and the previous one. Prior-cohort licensees were those licensed one interval earlier. Established physicians were licensed before these two most recent intervals. Because the AMD waves are unevenly spaced, the length of each interval varies across waves.

Appendix Table A8 and Figure A4 show the results for the stock breakdown. To understand which margins move in the medium run, I focus on the medium-run coefficient across the three groups. The results show a significant decline of 0.137 in the stock of established FMGs, while the effects for the other two groups are smaller, of opposite sign, and statistically indistinguishable from zero. Among the estimates for USMGs, there is also no equivalent decrease in any group. The medium-run decline among established FMGs is stable across robustness checks (Appendix Figure A20) but imprecisely estimated. The point estimate of -0.137 per 10,000 residents ($|t| = 2.6$) falls below the size-corrected threshold for wave-based

outcomes (Appendix A.3), and the robust confidence set excludes zero only under the smallest violation allowance ($[-0.270, -0.002]$ at $\bar{M} = 0.25$; $[-0.386, 0.119]$ at $\bar{M} = 1$). I therefore interpret the decomposition as pointing to established FMGs as the main candidate offsetting margin, rather than as a precise estimate of that offset. This reading is consistent with the trend-adjusted estimates in Appendix Table A16: the adjustment removes the pre-period coefficient for established FMGs (from 0.095 to -0.011), and the medium-run decline remains, at -0.292 with a robust confidence set that excludes zero at $\bar{M} = 1$.

5.2.5 County Heterogeneity

Does removing the citizenship requirement have differential effects across local health care markets? If it proportionally benefits markets that struggle to attract and retain physicians, the policy is welfare enhancing even if the state-level aggregate effect is modest. I test this using the baseline county partitions and the partition-specific estimator described in Section 4.4.

Table A14 show the sub-sample estimates for licensed physician stock by FMG and USMG and the Wald test for their difference. Appendix Figure A24 provides the robustness patterns for the between-county divergence. Each panel represents one category for county split. Overall, the evidence suggests a divergence in the licensed USMG stock between county types, but no detectable divergence in the FMG stock.

Column 3 and Column 6 of Table A14 show the statistical significance of the divergence in FMG and USMG licensed stock. As shown in Column 3, statistical significant divergence in FMG licensed stock is hardly detected across any subgroups. Column 6 demonstrates an economically meaningful and statistically significant divergence in USMG stock across county splits. However, the results are suggestive for the reason of pre-trend and non-robustness. Appendix Figure A25 shows the consistent pre-treatment divergence across income, urban share and expected Medicaid growth split. For low-physician-density counties, there is no pre-treatment divergence, but the post-treatment divergence becomes close to zero or even flips sign when either of the last-treated states, Wisconsin or Montana, is dropped.

5.3 Effects on USMGs

In this section, I provide evidence on the impact on USMGs.

5.3.1 Flow and Stock

Figure 3 demonstrates the impact on USMG license flow stays near zero and flat before and after the removal of citizenship requirement. Table A9 and right panel of Figure 4 show the event-study estimates on the stock of licensed USMGs.

The significant and negative pre-period coefficient suggests treated states experienced a decrease in USMG stock before the removal. Figure A21 reports the robust confidence sets consistently cover zero across $\overline{M} \in \{0.25, 0.5, 0.75, 1\}$. This is perhaps not surprising given the turnover in the pool of USMGs. The robust confidence intervals bound the effect to 4-5%. Hence, I find no detectable crowd-out of licensed USMGs in either the stock or the flow. The size-corrected inference in Appendix A.2 makes precise what the flow null can exclude: a decline in USMG licensing flow larger than about 0.14 per 10,000 residents, roughly 40 percent of the mean flow, is ruled out, whereas a decline of 15 percent of the mean flow would be detected in only 22 percent of replications.

5.3.2 Results by Specialty

Aggregate estimates may obscure heterogeneous responses across specialties and between FMGs and USMGs. Table A10 reports specialty-specific licensing rates, while Figure A5 and Table A11 report licensed stocks.

FMG licensing-rate estimates are positive across primary care, hospital-based specialties, and surgery, but only the surgical estimate remains consistently distinguishable from zero after allowing for modest violations of parallel trends ($\bar{M} = 0.25$). The corresponding FMG stock increases are modest, approximately 1.5%, 2.8%, and 1.7% of total specialty stock, respectively, and become statistically indistinguishable from zero in at least one specification. I also find no systematic evidence of crowd-out among USMGs. Specialty-specific USMG stock estimates are generally small or positive, although imprecisely estimated.

5.4 Experience and Quality of Physicians

This section speaks to the key policy debate regarding whether the physician workforce become worse as a result of removing citizenship. To do so, I provide evidence on two distinct channels. First is whether removing citizenship requirement changes who selects into licensing. Second is whether removing citizenship requirement changes the average experience and quality of physicians due to compositional change.

Quality variation between US medical schools and foreign medical schools is challenging to compare without a standardized test. I measure FMG quality using the average licensing-exam pass rate of the medical school from which each licensee graduated.[23] USMG quality is measured in two ways, using the quality rating from Cole and Lipton (1977) and today's US News & World Report ranking. Experience is measured the same way for FMGs and USMGs, as years since graduation from medical school.

[23] Because the measure is defined at the level of the school rather than the individual, it captures the average preparation of a physician's training institution rather than that physician's own exam performance, and state-level movements in it therefore reflect compositional shifts in where newly licensed FMGs were trained.

Experience. Panel A of Table 6 and Figure A6 shows the estimates on average experience of licensed FMGs and USMGs. Left panel shows a decrease in experience among licensed FMGs. The short-run estimates are not robust across control specifications. The coefficient magnitudes become smaller when controlling for other healthcare demand reforms. The medium-run coefficient is no longer significant at the 5 percent level after pre-trend adjustment.

Selection into licensing. Panel B of Table 6 and Figure A7 shows a statistically non-significant yet noisy effect on the quality of new licensees in the calendar year. The 95% robust confidence interval is not informative: the result can neither reject a decrease in the quality among licensees by 10 percentage point in license exam pass rate, nor an increase in the quality among licensees by 27 percentage point in license exam pass rate.

Quality of the stock. Panel C and D of Table 6 and Figure 5 show the coefficient on the average quality measure of licensed FMGs and USMGs. Columns 1–3 show no deterioration in quality among FMGs in the short-run: the point estimate is positive, and the Rambachan and Roth (2023) robust confidence intervals lie entirely above zero when the post-treatment violation of parallel trends is restricted to one quarter of the largest pre-treatment violation. Given the statistically significant difference in the pre-treatment period, Appendix Table A16 re-estimates the specification after removing each cohort's pre-period linear trend: the short-run effect on FMG stock quality becomes -0.02 percentage points, statistically indistinguishable from zero, and the robust confidence interval rules out declines larger than 1.85 pp at $\overline{M} = 0.25$. To put this bound in perspective, based on a cross-state standard deviations at 6.5 percentage point, a decline of that size would move a state from the median to the 39th percentile of the cross-state quality distribution. The medium-run estimate, by contrast, is inconclusive. The unadjusted estimate is 3.97 pp and significant at the 1 percent level, but its sign flips to -4.61 pp after trend adjustment, and the adjusted pre-period coefficient (2.77 pp, also significant at the 1 percent level) indicates that a linear trend fitted on the two-year 1964–1966 window extrapolates poorly over a 2–9-year horizon. The robust confidence interval for the unadjusted medium-run estimate at $\overline{M} = 1$, $[-5.44, 9.10]$, likewise includes zero. I therefore restrict the quality claim to the short run.

Columns 4–6 shows a decrease in the share of high-quality USMGs in the short-run and the estimates are statistically distinguishable from zero under robust confidence interval and across specifications. The result becomes smaller and noisier after de-trending (Appendix Table A16). The magnitude is not economically meaningful: the short-run decline of -0.58 pp in Panel C equals roughly 0.06 cross-state standard deviations (SD equals 10.5 pp).

Taken together, the results provide no robust evidence that easing licensing restrictions reduced the

measured quality of licensed FMGs, although the estimates for new licensees remain too imprecise to assess changes in selection into licensing. Similarly, the size-corrected critical value for wave-based outcomes (Appendix A.3) suggests that the short-run stock-quality estimate is not statistically significant ($|t| = 3.0$ against 3.7).

6 Hospital Adjustment in Capacity and Utilization

Physician licensing can affect not only where physicians practice but also the organizations through which medical care is produced. Classic theories of hospital organization place physicians at the center of hospital decision-making.[24] These theories suggest that physician supply and hospital capacity are jointly determined. Hence, removing citizenship requirements may make hospitals easier to staff and operate. I examine this organizational response using AHA data on the number of hospitals, non-physician employment, beds, admissions, and patient days.

I separately test the effects for short-term general hospital and long-term general hospital in consideration of the differences discussed in Section 2.3.2. For short-term general hospitals, physicians hold admission privileges and therefore have more bargaining power over hospital operation. Therefore, licensing reform might induce more licensed physicians to compete for employment in this sector. Hospitals could have capacity to expand their other inputs that are complementary to physician supply. Or capacity could remain the same if new entry of licensed physicians displaces incumbents at hospitals. For long-term general hospitals, the response could be ambiguous but for different reasons. Given the less desirable amenities at these institutions, removing citizenship restrictions might shrink the capacity of those hospitals if newly licensed physicians have better outside options. Or it may release capacity constraints by expanding the scope of practice once incumbent unlicensed physicians are licensed.

6.1 Main Results

Table 7 and Figure 6 report the effect of citizenship removal on hospital input, including capacity measures such as beds and non-physician employment, and output, including utilization measure such as admissions and patient days. Overall, the results suggest an expansion in hospital capacity in both hospital types. On utilization, short-term general hospitals show a medium-run increase in admissions, whereas the

[24] Pauly and Redisch (1973) model the nonprofit hospital as a physicians' cooperative, in which the medical staff influences the hospital's choice of complementary labor and capital. Harris (1977) similarly characterizes the hospital as comprising a physician-led "demand division" and an administrative "supply division": physicians' admission and treatment decisions generate demands for hospital resources, while administrators determine the capacity available to meet them.

utilization estimates for long-term hospitals are not robust to controlling for Medicaid.

Column 1 to 3 of Table 7 show a persistent expansion in beds capacity followed by a medium-run increase in admissions among short-term general hospitals. Column (1) of panel A shows the number of beds per 10,000 residents increase by 1.5 in the short-run and 2.5 in the medium run, with statistical significance at 5% and 1% level. Compared to the baseline mean, this corresponds to a modest 4 and 7% increase in hospital beds. When controlling for Medicaid in Column 2 and Medicare in Column 3, the point estimate remains similar while they are no longer statistically significant under Medicaid specification. One explanation is Medicaid coincides with part of the treatment timing and therefore leaving the amount of residual variation unable to identify an effect distinguishable from zero. The size-corrected critical value (Appendix A.2) suggests that the overall ATT is statistically significant ($|t| = 3.1$ against 2.8) while the short-run and medium-run estimates are not ($|t| = 2.4$ and 2.7); the capacity result therefore rests on the overall effect and the event-study pattern rather than on either window alone. Panel B shows a positive but short-lived effect on non-physician employment: 2.9 per 10,000 residents in the short run, significant at the 10 percent level (5.3 and significant at the 1 percent level with the Medicaid control; 2.4 and insignificant with the Medicare control), which dissipates in the medium run. Panel C and D show an expansion in admission in the medium run while patient days have no statistically significant changes.

For long-term general hospitals, Table 7 shows a medium-run and economically substantial expansion in non-physician employment. Panel B shows the medium run employment effect is statistically significant across specifications and the coefficients range from 24 to 42 percent relative to the baseline. It is, however, sensitive to the composition of the comparison group,[25] so I treat its magnitude as suggestive. Other outcomes are not robust when controlling for Medicaid. The coefficients move towards zero and become statistically insignificant, suggesting that Medicaid accounts for both the treatment timing variation and the changes in long-term hospital outcomes.

6.2 Timing Heterogeneity

The hospital response follows the same calendar-year pattern as the licensing flow. Panel (b) of Figure 7 shows the Callaway–Sant'Anna group-average ATT for short-term general hospital beds aggregated by calendar year. As with FMG licensing, effects from 1972 onwards are substantially larger and more persistent than in earlier years. Section 5.1.2 relates this timing to the 1970 relaxation of the J-1 home-country return requirement; that the bed response tracks the licensing response so closely is consistent with hospital

[25] As will be discussed in Section 8, the medium-run long-term employment effect attenuates substantially once Wisconsin is dropped from the estimation sample.

capacity adjusting to the physician inflow the federal change enabled, rather than to the state reform alone.

6.3 Geographic Heterogeneity in Hospital Response

This setting's simultaneous public-insurance expansion allows a further question: do markets facing different expected changes in payer mix respond differently to a supply-side shock? Section 5.2.5 showed little divergence in physician supply across county types. Hospital inputs are a separate margin: even where the licensed physician stock moves similarly, hospitals facing different expected demand may adjust capacity differently. I compare counties above and below the median on each of the four baseline partitions defined in Section 3.6, using the partition-specific estimator and Wald test described in Section 4.4.

6.3.1 County divergence in hospital input and output

Table 8 show the sub-sample estimates and the Wald test for their difference. Appendix Figure A26 provides the robustness patterns for the between-county divergence. Each panel represents one category for county split. Overall, the evidence shows that after removal of citizenship requirement, low income, less urban, and high Medicaid growth counties show a statistically significantly higher level of patient days per capita relative to their counterparts. Additionally, higher level of hospital inputs, including employment and beds, are detected for high expected-demand counties relative to low expected-demand counties.

Column 3 and Column 6 of Table 8 show the statistical significance of divergence in short-run hospital non-physician employment and beds. The robust and statistically significant divergence exists in Panel D for high versus low Medicaid growth counties. High Medicaid growth counties show a 23.8 level difference in employment and 6.4 level difference in beds, which corresponds to a quarter to a third of baseline mean for low Medicaid growth counties. A divergence in either input is not detected in the other three splits (Panels A–C). There, Column 3 shows employment divergences that are statistically significant but sensitive to specification, as shown in Appendix Figure A27, while Column 6 shows bed divergences that are small and statistically insignificant.

Column 9 and 12 of Table 8 show the divergence in hospital admissions and patient days. Similarly, Appendix Figure A26 presents the robustness of the divergence estimates. For patient days, a statistically significant at 1% divergence is robustly detected across four groups. A marginally statistically significant (at 5 or 10%) divergence in admissions is detected across multiple subgroups but the patterns exist before treatment. Hence, I cannot detect a causal post-treatment divergence.

7 Effects on Population Mortality

This section answers a key question of public interest: does removing citizenship requirement harm or benefit public health? In this section, I examine the impact on population mortality rates.[26]

7.1 All-cause Mortality Effects

The expected effects of removing the citizenship requirement on mortality depend both on the extent to which deaths are preventable and on how much the effectiveness of care changes with physician quantity, quality, and hospital capacity. Given the magnitude of responses in both physician and hospital capacity, I do not expect the effects on population mortality to be large.

Table A12 shows the event-study estimates for all-cause mortality among near-elderly (55–64) in panel A, elderly adults (65+) in panel B, and infants in panel C. Results suggest the impact on population mortality by age group could be attenuated towards zero when introduction of Medicaid is controlled. Compared to the baseline specification in column (1), the medium-run effect for near-elderly and infant mortality not only have larger confidence intervals, but also attenuates toward zero when controlling for Medicaid in column (2). These patterns suggest the timing of Medicaid introduction not only absorbs part of the treatment variation, but also to a large extent explains the divergence in mortality decrease between early versus later treated states. In contrast, controlling for the regional exposure to Medicare in column (3) yields similar estimates. This is consistent with Medicaid's introduction substantially reduced infant and child mortality, particularly among non-white children (Goodman-Bacon, 2018), whereas Medicare produced no detectable reductions in elderly mortality during its first decade (Finkelstein and McKnight, 2008).

Column (2) suggests the estimates on mortality decreases are noisy. The 95 percent confidence intervals can reject any decrease by 2 to 3 percent for the near elderly (55-64), and by 0.8 to 1.4 percent for the elderly (65+). In comparison, the treatment effect of being exposed to a community health center (CHC), is around 2.4 percent (Bailey and Goodman-Bacon, 2015). Hence, the estimates are precise enough to reject any mortality decrease equivalent to the effect of CHC, but not precise enough to reject any smaller but plausible changes driven by the removal of citizenship.

Estimates for effects on infant mortality are less precise. For comparison, Duke-funded hospital modernization reduced infant mortality by approximately 7.5 percent while increasing hospital beds by approximately 50 percent and physician supply by 11 percent (Hollingsworth et al., 2024). The capacity expansion induced by citizenship removal was substantially smaller. Even using the physician-supply response—which

[26] Unlike many studies in contemporary health settings use claims data, a physician or hospital-linked patient mortality rate is not available during the period of study.

yields the larger of the two proportional benchmarks—and imposing the strong assumption that the mortality return to additional capacity was comparable in the 1960s, proportional scaling implies an infant mortality reduction of 1.4 percent. As shown in column 3, the short-run 95 percent confidence interval is able to reject a reduction of 1.2 percent, and in the medium-run it can only reject a 10 percent decrease to a 9 percent increase. Therefore, the estimates rule out relatively large changes in infant mortality in the short-run but become too imprecise to distinguish plausibly sized effects from zero in the medium run.

7.2 Mortality by Cause of Death

All-cause mortality may dilute effects concentrated in a few treatable conditions. In this section, I further separately estimate mortality rates by causes of death. The physician labor-market and hospital results imply that any mortality response could operate through hospital inpatient setting and acute conditions. The reform increased the physician stock and hospital capacity only modestly and produced no detectable increase in inpatient admissions. This implies mortality is most likely to respond when timely diagnosis and immediate inpatient treatment can be delivered. To evaluate this prediction, I organize causes of death into infectious, cardiovascular,[27] cancer, and external cause such as motor vehicle accidents.

For older adults, the number of deaths that could have been averted by medical intervention is well documented for this period, but only a limited set of causes is likely to respond to the change in physician supply and hospital capacity. Acute respiratory conditions, such as pneumonia, could benefit substantially from inpatient hospital capacity by timely antibiotic administration for bacterial infections, oxygen and supportive care, and escalation to ventilatory support. Acute cardiac events are also likely to respond quickly if patients are admitted promptly to coronary-care units for continuous cardiac monitoring, defibrillation, and treatment. In comparison, other causes could be more ambiguous. Chronic heart disease such as hypertension is less likely to respond given the small change in primary care physician supply and the dilution by the drug compliance from patients. Cancer mortality is unlikely to respond at the aggregate level as the response in surgery survival or oncology treatment may be difficult to be precisely captured in population mortality. Changes in mortality from external causes, such as motor vehicle accidents, are mostly explained by seat belt and speed limit during this time, so only a small proportion of it could be altered by the expanded capacity through emergency department, surgical and inpatient sector.

Table A13 and Figure A8 show the results for by-cause mortality. Overall, the evidence does not support a robust and statistically precise change in mortality. For elderly adults' mortality in Columns 1 to 3, the

[27] Acute and chronic heart disease cannot be separated out over the full sample period, since ICD-7 (in use through 1967) provides no distinct rubric for acute myocardial infarction and the distinction becomes available only with the 1968 revision, after several treated states had already reformed.

effects on treatable incidents such as infectious disease become smaller and statistically insignificant when controlling for Medicaid. Appendix Figure A23 further shows the pattern does not hold across the majority of leave-one-state-out exercises. Similarly for cancer, although the medium-run increase is stable across specifications in Table A13, the estimate falls to near zero when Wisconsin is dropped. Other estimates, including cardiovascular disease and external causes, are economically small and statistically insignificant. For near elderly adults, mortality outcomes either show no responses or become non-detectable when controlling for Medicaid.

8 Robustness Checks

Alternative staggered difference-in-differences estimators. To assess sensitivity to the choice of estimator, I compare the preferred Callaway–Sant'Anna doubly-robust estimates to those obtained from alternative staggered difference-in-differences methods (Callaway and Sant'Anna, 2021; Gardner, 2021; Wooldridge, 2021; Sun and Abraham, 2021; Borusyak et al., 2024) and a stacked difference-in-differences estimator (Cengiz et al., 2019). Appendix Figure A28 presents these comparisons for FMG licensing flow and short-term general hospital admissions. Across almost all alternative methods, the principal effect on physician licensing flow is precisely estimated and stable throughout, whereas the downstream effect on hospital admissions is positive and of consistent sign and shape across all estimators but less precisely estimated. The only exception is the Sun and Abraham (2021) specification. I consider this specification more demanding as it limits the flexibility in using all not-yet-treated states, since the control group is specified as a fixed group. In this case, when only the last two treated state, Montana and Wisconsin, are used as control group, the parallel trend assumption is challenging to hold for the majority of the control states. This is revealed for the admissions outcome, where the upward pre-trend exists.

Sensitivity to alternative control variables. The preferred conditional estimator balances treated and not-yet-treated states on a single baseline covariate, the 1964 ratio of FMGs among trainees, which raises the concern that the estimates may depend on this particular measure of pre-reform FMG presence. To assess this, I re-estimate the main specifications by adding alternative measures of FMG presence in the physician labor market on top of the baseline FMG-among-trainees control. I also include the specifications when adding year of Medicaid or Medicare intensity as controls. Appendix Figures A15 and A16 report the resulting estimates. Each panel plots the point estimate under the baseline specification and under each additional control (the control ladder, shown as diamonds), for the physician labor market and hospital outcomes respectively. Across all alternative controls the point estimates remain close to the baseline estimate,

confirming that the results are not sensitive to the choice of baseline FMG covariate.

Community Health Centers. One concern is that the federal Community Health Center (CHC) program, which expanded primary-care capacity in medically underserved areas over the same period, could confound the estimates. This could constitute an omitted-variables bias if the program timing coincides with trends in the removal of citizenship requirements and independently increased the demand for foreign-trained physicians while also facilitating the expansion in hospital beds. On the other hand, it might not confound my estimates if CHCs do not rely on FMGs. To directly address this concern, I re-estimate the state level event-study model by reconstructing the state-level outcome after dropping the 59 counties that experienced a CHC expansion by 1975, using the staggered roll-out information from Bailey and Goodman-Bacon (2015). As shown in Appendix Figure A31, the post-reform increase in hospital beds remains positive and statistically significant under this exclusion.

Hill-Burton Programs. Another concern is the legacy of the Hill-Burton program, which financed hospital construction in the decades before and during my sample window. Counties with new Hill-Burton funding may be a confounder similar to CHC, so I re-estimate the model after dropping the 395 counties with Hill-Burton funding receipt during 1964-1971, using the staggered roll-out information from Chung et al. (2017). As shown in Appendix Figure A31, the post-reform increase in beds become near zero and statistically insignificant under this exclusion. However, this exclusion also eliminates more than half of the population among states of analysis. Therefore, the decrease in the identified treatment effect could be attributed to restriction to sample counties with smaller treatment effects, or attributed to the simultaneous roll-out of HB. But they can hardly be separated.

Certificate of Need Laws. Another potential concern is that Certificate of Need (CON) laws, where state regulations restricting hospital capacity expansion, adopted by most states between the mid-1960s and early 1980s, could confound the results if CON adoption timing correlates with both the citizenship reform window and FMG physician supply. I do not conduct a formal robustness test for this channel because CON laws limit hospital bed supply rather than expand it, therefore likely making my estimates more conservative.

Placebo outcomes. To further validate the identification, I estimate the baseline specification using three placebo outcomes that should not respond to the citizenship requirement removal: per capita total physician count from the Area Health Resource File (AHRF), log state population, and cumulative Hill-Burton funding for nonprofit hospitals. The AHRF-based physician count serves as an independent cross-check on the AMD-based estimates, while state population tests whether the per-capita denominator itself shifts in response to the reform. Cumulative Hill-Burton funding tests whether the reform coincided with differential federal capital subsidies to hospitals. The null result confirms that the documented bed expansion reflects

the relaxation of the physician-labor constraint rather than a simultaneous increase in funding. Appendix Figure A29 shows that none of the outcomes exhibit a significant post-reform change, supporting the validity of the identification strategy.

Leave-one-state-out sensitivity. To explore whether the results are driven by any single treated state, I re-estimate the main specifications sequentially dropping one treated state at a time. Appendix Figures A15 and A16 plot the resulting distribution of estimates alongside the control ladder: each leave-one-out estimate appears as a circle, with the Wisconsin exclusion highlighted in red. The point estimates are stable across all leave-one-out samples but occasionally sensitive for Wisconsin (WI). For physician license flow and stock, the coefficients become smaller and less statistical significant. Short-run effect among short-term hospital beds remains similar in magnitude but with wider confidence intervals. I interpret this as meaning WI plays an outsized role in the identification for two reasons. First, its relatively late treatment timing (1976) means it serves as a control group for longer-run effects in later calendar years, from which the larger post-reform point estimates arise. Second, its large population shifts the propensity-score weights toward larger states, so dropping it changes the effective comparison group. This interpretation is supported by comparing the estimates when removing Montana, which is the other state that removed the citizenship requirement after 1975 but which has a smaller population, where the point estimates remain similar. I formally test the compositional difference driven by dropping WI in Appendix A.1 and show that the coefficient shrinkage is mainly driven by the decrease in the weight of later cohort effects rather than a decrease in the magnitude of the cohort-by-time effects. The results that are not robust are those for physician experience and the medium-run hospital outcomes—short-term general beds, admissions, and long-term employment—where either the sign flips or the magnitude attenuates to near zero.

Cohort-specific trend adjustment. A remaining concern is that treatment cohorts may sit on different underlying trajectories, a concern with a specific economic basis here: immigrant inflows are geographically persistent, so states with greater baseline FMG presence may have experienced greater subsequent inflows regardless of licensing reform. To assess this, I re-estimate the main specifications after subtracting each cohort's own pre-period linear trend, following Goodman-Bacon (2021a). This is not a weaker identifying assumption than parallel trends but a different one, and neither nests the other. Appendix A.6 details the procedure, the supporting literature, and the pre-treatment trajectories by cohort. Because cohorts differ substantially in the level of FMG density but follow broadly similar pre-treatment paths, I report these estimates as a robustness exercise rather than as my preferred counterfactual model.

Appendix Table A16 reports the headline estimates with and without the adjustment. For the stock of licensed FMGs the procedure performs as expected. The pre-removal coefficient falls from 0.057 to 0.028

and is no longer statistically significant, while the short-run effect strengthens from 0.187 to 0.226 and its robust confidence interval now excludes zero even at $\overline{M} = 1$. The same pattern holds for the stock decomposed by specialty, where significant pre-trends among primary-care and hospital-based physicians are absorbed and the short-run estimates strengthen rather than attenuate. Likewise, for the established-FMG stock the adjustment absorbs the pre-period coefficient (from 0.095 to -0.011) while the medium-run decline persists and strengthens (from -0.137 to -0.292), so the offsetting margin identified in Section 5.2 does not depend on the pre-trend. The selection test at the entry margin is essentially unchanged.

The pre-trend adjustment does not behave uniformly across outcomes. For the quality of the licensed FMG stock it does not remove the pre-trend but reverses it, from -2.22 to $+2.77$, with both significant at conventional levels. Mechanically, the fitted trend overshoots: zeroing a pre-removal coefficient of -2.22 requires an adjustment of 2.22 percentage points, whereas the fitted 1964–1966 slope supplies 4.99, more than twice what is needed. Appendix A.6 documents the same failure mode for average experience and for the licensed share by specialty, where the unadjusted specifications show no pre-trend and the adjustment creates one.

Taken together, the trend adjustment confirms that the licensing and specialty-composition effects on the licensed stock are not driven by differential cohort trajectories, and leaves the no-quality-dilution result at entry intact. The divergence lies at the stock-quality outcome, whose pre-trend test fails in opposite directions with and without the adjustment. I therefore rest the quality conclusion in Section 5.4 on the bound that both specifications share rather than on the sign of either one.

9 Conclusion

This paper is the first to evaluate the causal impact of state licensing regulations on physician supply and its downstream consequences for hospital capacity and public health. I use a staggered difference-in-differences design, combining newly digitized physician-level data with state licensing and hospital records, to answer this question. I find that the reform produced a clear first stage: new FMG licenses rose 17 to 28 percent relative to the pre-reform annual flow, depending on the licensing measure and specification. However, this entry did not result in a durable stock increase. The suggestive short-run gain of roughly 2 percent of the licensed pool dissipates by the medium run, and decomposing the stock by licensure recency suggests the offset is concentrated among established FMG licensees rather than new entrants or U.S.-trained physicians. Lastly, hospitals adjusted. Short-term general bed capacity rose 4 to 7 percent; the point estimate is stable across control specifications, though precision falls when Medicaid timing is controlled for.

These findings reconcile competing claims in the current occupational licensing debate. Removing a non-skill-related licensing barrier substantially increased entry, demonstrating that state requirements can constrain access to the profession. Yet the resulting increase in the physician stock was temporary, showing that additional licenses do not necessarily translate one-for-one into a durable expansion of physician supply. Importantly, this offset occurred among established FMGs rather than through displacement of U.S.-trained physicians. This finding shifts the relevant concern from crowd-out toward the retention and geographic reallocation of foreign-trained physicians. The increase in entry also did not come at the cost of measured quality. The training quality of the licensed FMG stock did not deteriorate in the short run: I can rule out declines larger than roughly one-quarter of a cross-state standard deviation under modest violations of parallel trends. Medium-run quality estimates are too sensitive to trend assumptions to be conclusive.

The concentration of effects after the federal relaxation of the J-1 return requirement further suggests that state licensing rules and federal immigration restrictions operated as complementary constraints on foreign physicians' entry into practice. Finally, the expansion in hospital capacity shows that the consequences of licensing reform extend beyond physician headcounts to other inputs in healthcare production that adjust jointly with physician availability.

More broadly, these findings show that the integration of high-skilled immigrants depends on a layered set of institutions governing whether their existing human capital can be put to productive use. State licensing rules, federal immigration restrictions, and employer capacity jointly determine effective labor supply.

At the same time, the increase in hospital bed capacity shows that organizations respond endogenously to changes in the availability of licensed labor by adjusting complementary inputs. The effects of physician supply-side regulation therefore cannot be inferred from professional headcounts alone. Such regulation may reshape who participates in a market, where workers practice, and how firms organize production—even when its effect on the equilibrium stock of workers is modest.

References

Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., and Pérez, S. (2021). Automated linking of historical data. *Journal of Economic Literature*, 59(3):865–918.

Advisory Commission on Additional Licensing Models (2025). Guidance on implementing additional licensing pathways for internationally trained physicians. Technical report, Federation of State Medical Boards, Intealth, and ACGME. Accessed July 2026.

Alexander, J. (2026). Winning the (H-1B) lottery: Effects of high-skilled workers on firm production and employment. Working paper, University of Notre Dame. Available at SSRN.

Alsan, M., Garrick, O., and Graziani, G. (2019). Does diversity matter for health? experimental evidence from oakland. *American Economic Review*, 109(12):4071–4111.

Alsan, M. and Neberai, Y. (2024). *Why Doesn't the United States Have National Health Insurance?: The Role of the American Medical Association*. National Bureau of Economic Research.

American Osteopathic Association (2024). Comments on the FSMB advisory commission on additional licensing models draft guidance. Formal comment letter. Accessed July 2026.

Atal, J. P., Larroucau, T., Munoz, P., and Otero, C. (2025). Physicians' occupational licensing and the quantity-quality trade-off. PIER Working Paper 25-016, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Badinski, I., Finkelstein, A., Gentzkow, M., and Hull, P. (2023). Geographic variation in healthcare utilization: The role of physicians. Technical report, National Bureau of Economic Research.

Bailey, M. J. and Goodman-Bacon, A. (2015). The war on poverty's experiment in public medicine: Community health centers and the mortality of older americans. *American Economic Review*, 105(3):1067–1104.

Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A., and Sant'Anna, P. H. C. (2026). Difference-in-differences designs: A practitioner's guide. *Journal of Economic Literature*, 64(2):498–557.

Bartel, A. P. (1989). Where do the new u.s. immigrants live? *Journal of Labor Economics*, 7(4):371–391.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Bhalla, V. (2010). "we wanted to end disparities at work": physician migration, racialization, and a struggle for equality. *Journal of American Ethnic History*, 29(3):40–78.

Black, B., Hollingsworth, A., Nunes, L., and Simon, K. (2022). Simulated power analyses for observational studies: An application to the Affordable Care Act Medicaid expansion. *Journal of Public Economics*, 213:104713.

Borjas, G. J. (2005). The labor-market impact of high-skill immigration. *American Economic Review*, 95(2):56–60.

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

Braga, B., Khanna, G., and Turner, S. (2024). Migration policy and the supply of foreign physicians: Evidence from the conrad 30 waiver program. *Journal of Economic Behavior & Organization*, 226:106682.

Butter, I. (1971). The migratory flow of doctors to and from the united states. *Medical Care*, 9(1):17–31.

Butter, I. and Sweet, R. G. (1977). Licensure of foreign medical graduates: an historical perspective. *The Milbank Memorial Fund Quarterly. Health and Society*, pages 315–340.

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

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.

Campo, F., Mendola, M., Morrison, A., and Ottaviano, G. I. P. (2022). Talents and cultures: Immigrant inventors and ethnic diversity in the age of mass migration. *Journal of the European Economic Association*, 20(5):1971–2012.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.

Carollo, N. A., Hicks, J. F., Karch, A., and Kleiner, M. M. (2025). The origins and evolution of occupational licensing in the united states. Technical report, National Bureau of Economic Research.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

Chan, D. C., Gentzkow, M., and Yu, C. (2022). Selection with variation in diagnostic skill: Evidence from radiologists. *The Quarterly Journal of Economics*, 137(2):729–783.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605.

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

Cicero Institute (2025). Expanding physician access act of 2025 (model legislation). Technical report, Cicero Institute. Accessed July 2026.

Clay, K., Miller, G., Portnykh, M., and Schmick, E. J. (2025). Medical school closures, market adjustment, and mortality in the flexner report era. Technical report, National Bureau of Economic Research.

Cole, J. R. and Lipton, J. A. (1977). The reputations of american medical schools. *Social Forces*, 55(3):662–684.

Cunningham, J. P. and Goodman-Bacon, A. (2025). Changes in family structure and welfare participation since the 1960s: The role of legal services. *American Economic Journal: Applied Economics*, 17(1).

Davis, K. and Schoen, C. (1978). *Health and the War on Poverty: A Ten-Year Appraisal*. Brookings Studies in Social Economics. Brookings Institution, Washington, D.C.

Educational Commission for Foreign Medical Graduates (2025). About ECFMG: International medical graduates in the U.S. physician workforce. ECFMG. Accessed June 2026.

Engel, J. (2006). *Poor People's Medicine: Medicaid and American Charity Care since 1965*. Duke University Press, Durham.

Federation of State Medical Boards (2025). States with enacted and proposed licensure pathways for internationally trained physicians: State-by-state overview. Technical report, FSMB.

Federation of State Medical Boards and Intealth and Accreditation Council for Graduate Medical Education (2024). FSMB, Intealth, ACGME establish advisory commission to guide alternate pathways for state licensure of international medical graduates. Joint news release. Accessed July 2026.

Finkelstein, A. (2007). The aggregate effects of health insurance: Evidence from the introduction of medicare. *The quarterly journal of economics*, 122(1):1–37.

Finkelstein, A. and McKnight, R. (2008). What did medicare do? the initial impact of medicare on mortality and out of pocket medical spending. *Journal of public economics*, 92(7):1644–1668.

Friedman, M. and Kuznets, S. (1945). *Income from Independent Professional Practice*. National Bureau of Economic Research, New York.

Gardner, J. (2021). Two-stage differences in differences. Technical report, Working Paper. arXiv:2207.05943.

Gittleman, M., Klee, M. A., and Kleiner, M. M. (2018). Analyzing the labor market outcomes of occupational licensing. *Industrial Relations: A Journal of Economy and Society*, 57(1):57–100.

Goodman-Bacon, A. (2018). Public insurance and mortality: Evidence from medicaid implementation. *Journal of Political Economy*, 126(1):216–262.

Goodman-Bacon, A. (2021a). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Goodman-Bacon, A. (2021b). The long-run effects of childhood insurance coverage: Medicaid implementation, adult health, and labor market outcomes. *American Economic Review*, 111(8):2550–2593.

Gottlieb, J. D. and Nicholson, S. (2026). Physician competition: Entry and substitution. *Journal of Economic Perspectives*, 40(2):117–142.

Gottlieb, J. D., Polyakova, M., Rinz, K., Shiplett, H., and Udalova, V. (2025). The earnings and labor supply of us physicians. *The Quarterly Journal of Economics*, page qjaf001.

Grabowski, D. C., Gruber, J., and McGarry, B. (2023). Immigration, the long-term care workforce, and elder outcomes in the U.S. NBER Working Paper 30960, National Bureau of Economic Research.

Harris, J. E. (1977). The internal organization of hospitals: Some economic implications. *The Bell Journal of Economics*, 8(2):467–482.

Hirshbein, L. (2022). Foreign medical graduates and american psychiatry. *History of Psychiatry*, 33(2):163–179.

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

Kaestner, R. and Kaushal, N. (2012). Effect of immigrant nurses on labor market outcomes of US nurses. *Journal of Urban Economics*, 71(2):219–229.

Kleiner, M. M. and Krueger, A. B. (2013). Analyzing the extent and influence of occupational licensing on the labor market. *Journal of Labor Economics*, 31(S1):S173–S202.

Kleiner, M. M. and Kudrle, R. T. (2000). Does regulation affect economic outcomes? the case of dentistry. *Journal of Law and Economics*, 43(2):547–582.

Kleiner, M. M. and Soltas, E. J. (2023). A welfare analysis of occupational licensing in us states. *Review of Economic Studies*, 90(5):2481–2516.

Knesper, D. J. (1978). Psychiatric manpower for state mental hospitals: A continuing dilemma. *Archives of General Psychiatry*, 35(1):19–24.

Kulka, A. and McWeeny, D. (2019). Rural physician shortages and policy intervention. *Available at SSRN 3481777*.

Lo Sasso, A. T. (2021). Regulating high-skilled immigration: The market for medical residents. *Journal of Health Economics*, 76:102436.

Lockett, B. A. and Williams, K. N. (1974). Foreign medical graduates and physician manpower in the united states. Technical Report DHEW Publication No. (HRA) 74-30, U.S. Department of Health, Education, and Welfare, Health Resources Administration, Division of Manpower Intelligence, Washington, DC.

Mansell, L. (2025). Testimony: Creating a full license pathway for international medical graduates. Testimony to the Washington State Senate Committee on Health & Long-Term Care on SB 5185, Niskanen Center. Accessed July 2026.

McIntyre, J. E. (1939). Citizenship and medical licensure. *Federation Bulletin (Federation of State Medical Boards of the United States)*, 25(3):76–83.

McNamara, C. and Pineda-Torres, M. (2024). Medical residency subsidies and physician shortages. Technical report, IZA Discussion Papers.

Mejia, A., Pizurki, H., and Royston, E. (1980). *Foreign medical graduates: the case of the United States*. Gower Publishing Company, Limited.

Mitchell, J. B. (1991). Physician participation in Medicaid revisited. *Medical Care*, 29(7):645–653.

Mitchener, K., Esteves, R., Nencka, P., and Thomasson, M. (2022). Do pandemics change healthcare? evidence from the great influenza. Technical report, CEPR Discussion Papers.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the u.s. labor market. *The Quarterly Journal of Economics*, 118(2):549–599.

National Center for Health Statistics (1965). Development and maintenance of a national inventory of hospitals and institutions. Vital and Health Statistics, Series 1 3, PHS Publication No. 1000, U.S. Department of Health, Education, and Welfare, Public Health Service, Washington, DC.

National Center for Health Statistics (1979). Multiple Cause-of-Death Mortality Data, 1959–1979. Machine-readable data file and documentation (National Vital Statistics System). Accessed July 2022 via the NBER Public Use Data Archive.

Noether, M. (1986a). The effect of government policy changes on the supply of physicians: Expansion of a competitive fringe. *The Journal of Law and Economics*, 29(2):231–262.

Noether, M. (1986b). The growing supply of physicians: Has the market become more competitive? *Journal of Labor Economics*, 4(4):503–537.

OECD (2023). Health at a glance 2023: OECD indicators. Technical report, OECD Publishing, Paris. See indicator "International migration of doctors and nurses".

Pauly, M. and Redisch, M. (1973). The not-for-profit hospital as a physicians' cooperative. *The American Economic Review*, 63(1):87–99.

Pauly, M. V. (1980). *Doctors and Their Workshops: Economic Models of Physician Behavior*. University of Chicago Press and NBER, Chicago.

Peri, G. (2016). Immigrants, productivity, and labor markets. *Journal of economic perspectives*, 30(4):3–30.

Peri, G., Shih, K., and Sparber, C. (2015). Stem workers, h-1b visas, and productivity in us cities. *Journal of Labor Economics*, 33(S1):S225–S255.

Peterson, B. D., Pandya, S. S., and Leblang, D. (2014). Doctors with borders: Occupational licensing as an implicit barrier to high skill migration. *Public Choice*, 160(1):45–63.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591.

Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R., Schroeder, M. B., and Sobek, M. (2010). Integrated public use microdata series: Version 5.0 [machine-readable database].

Schnell, M. and Currie, J. (2018). Addressing the opioid epidemic: is there a role for physician education? *American journal of health economics*, 4(3):383–410.

Sirrocco, A. (1972). Inpatient health facilities as reported from the 1969 MFI survey. Vital and Health Statistics, Series 14 6, DHEW Publication No. (HSM) 73-1801, National Center for Health Statistics, Rockville, MD.

Sloan, F., Mitchell, J., and Cromwell, J. (1978). Physician participation in state Medicaid programs. *Journal of Human Resources*, 13(Supplement):211–245.

Stevens, R. and Vermeulen, J. (1972). Foreign trained physicians and American medicine. Technical Report DHEW Publication No. (NIH) 73-325, U.S. Department of Health, Education, and Welfare, National Institutes of Health, Bureau of Health Manpower Education, Washington, DC.

Stevens, R. and Vermeulen, J. (1973). *Foreign trained physicians and American medicine*. US National Institutes of Health, Division of Manpower Intelligence.

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

Surveillance, Epidemiology, and End Results (SEER) Program (2009). Program populations (1969–1988). National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch. Released November 2009.

U.S. General Accounting Office (1977). Interpretation of public law 94-484 as it applies to alien medical doctors. Technical Report Decision B-189451, Comptroller General of the United States, Washington, DC. December 15, 1977.

Wasserman, M. (2023). Hours constraints, occupational choice, and gender: Evidence from medical residents. *The Review of Economic Studies*, 90(3):1535–1568.

Wooldridge, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. Technical report, Working Paper. Available at SSRN 3906345.

Zuckerman, S., McFeeters, J., Cunningham, P., and Nichols, L. (2004). Changes in Medicaid physician fees, 1998–2003: Implications for physician participation. *Health Affairs*, 23:W4–374–W4–384.

Tables and figures

Table 1: Balance Table

	(1) Mean	(2) Std.	(3) OLS	(4) IPW reweighting	(5) Hazard Model
Population Characteristics					
Population (in logs)	14.72	0.88	–0.34	–2.17*	0.14
			(0.97)	(1.10)	(0.31)
Income	8.86	0.17	–3.95	–7.56***	1.43
			(3.11)	(2.64)	(1.34)
Share Non-white	0.13	0.11	1.71	4.21	–0.33
			(5.33)	(7.71)	(1.40)
Share Foreign-born	0.03	0.02	–29.85	–41.34***	12.79*
			(21.04)	(11.79)	(6.59)
Share 0–5 years old	0.13	0.01	74.14	7.94	–14.12
			(81.07)	(148.96)	(32.20)
Share 65+ years old	0.09	0.02	31.78	58.51	–0.12
			(28.92)	(48.85)	(7.69)
Number of Hospitals	0.45	0.20	2.06	9.01**	–1.11
			(4.63)	(4.31)	(1.42)
Health Outcomes/Policies					
Infant mortality (white)	22.30	2.22	0.37	0.56	–0.06
			(0.30)	(0.40)	(0.12)
Infant mortality (non-white)	41.47	4.93	0.24	0.12	0.02
			(0.14)	(0.13)	(0.02)
Elderly mortality (65+)	560.54	43.54	–0.03**	–0.03	0.01**
			(0.01)	(0.02)	(0.00)
Share elderly uninsured (region)	0.48	0.07	15.31*	17.17*	–4.07
			(8.10)	(8.56)	(2.78)
Year of Medicaid adoption	1967.61	1.70	0.27	0.03	–0.07
			(0.34)	(0.49)	(0.10)
Endorsement	1966.30	5.04	–0.05	0.17	0.01
			(0.16)	(0.13)	(0.04)
FLEX exam	1970.67	1.53	0.09	1.27*	0.05
			(0.58)	(0.67)	(0.14)
Physician Labor-Market Outcomes (per 10,000 Residents)					
# USMGs Licensed	8.81	1.33	0.42	0.12	–0.13
			(0.41)	(0.57)	(0.10)
# FMGs Licensed	0.38	0.43	–2.45***	–1.42**	0.89***
			(0.72)	(0.64)	(0.31)
# USMGs Recently Licensed	0.90	0.28	2.84**	3.65*	–1.01**
			(1.19)	(2.08)	(0.48)
# FMGs Recently Licensed	0.12	0.12	–6.31**	–9.95***	2.68**
			(2.68)	(3.28)	(1.29)
# USMGs, in Training	0.88	0.55	–0.43	–1.10	0.03
			(1.62)	(2.08)	(0.30)
# FMGs, in Training	0.30	0.30	–4.70**	–4.36***	1.68***
			(1.71)	(1.40)	(0.59)
% of FMGs among Licensed	0.04	0.04	–31.68***	–17.78**	11.25***
			(9.60)	(6.80)	(3.63)
% of FMGs among Trainee	0.24	0.24	–7.57***	–5.86***	4.53***
			(1.73)	(1.54)	(0.96)
N	23	23	23	23	153

Notes: This table shows the balance of state-level characteristics. The states of interest are those that removed a citizenship requirement after 1964 (23 eventually-treated states). Column 1 shows the mean of variables in the baseline year (1964); column 2 the standard deviation. Column 3 shows OLS estimates regressing the year of removal on each baseline characteristic (one at a time). Column 4 reweights by inverse propensity-score weights. Column 5 reports a Cox proportional-hazard model using the lagged characteristic; physician-supply covariates (AMD survey waves) are carried forward to intervening years, giving a risk window of 1964–1975 (N≈153 person-years), while demographic and policy covariates use the full 1956–1975 panel (N up to

Table 2: State Removal of Citizenship/Naturalization Requirement by Year

Removal Year	States	% of States	% of Population
1961 (1)	NH	2.0	0.3
1962 (3)	IL, MN, PA*	5.9	13.4
1963 (2)	RI*, UT	3.9	1.0
1964 (2)	NM, VT	3.9	0.7
1965 (2)	MI, ND*	3.9	4.6
1968 (2)	DE, NJ	3.9	3.8
1969 (4)	KS, OK, TN, WV	7.8	5.4
1970 (7)	AK, FL, MO, NC, OH, SC, WY	13.7	14.6
1972 (3)	KY, MS, NE	5.9	3.6
1973 (4)	AL, AR, CO, GA	7.8	6.0
1976 (1)	WI†	2.0	2.2
1979 (1)	MT†	2.0	0.4
Unrestricted States (14)	CA, CT, DC, HI, ID, IN, IA, MD, MA, NY, SD, TX, VA, WA	27.5	39.6

Notes: The table shows the states that removed the citizenship requirement by year. % of States is the group's count out of the 50 states and the District of Columbia. % of Population weights each state by its 1964 population, the baseline year used for the estimation weights; Alaska is absent from the population series. * indicates states for which the citizenship requirement was in place for no more than 3 years. † indicates a removal year established from state statute rather than the AMA series: Wisconsin, Chapter 383, Laws of 1975; Montana, Sec. 2, Ch. 323, L. 1979. A few states are not listed for the following reasons: 1) Arizona adopted and repealed the citizenship requirement multiple times during the study period; 2) Four states relaxed their FMG licensing requirements without ever having imposed a citizenship requirement. Louisiana (1968) and Nevada (1969) relaxed their requirements from a blanket exclusion of FMGs to no requirement. Maine (1975) and Oregon (1971) relaxed their requirements from permanent residency to no requirement. 3) Data for Alaska (1970) are not available across all AMD years.

Table 3: Effects on Licensing Flows

	(1)	(2)	(3)	(4)	(5)	(6)
	# of First-time FMG Licensees			FMGs Newly Licensed in Year t		
Pre-period (-5 to -1)	–0.017	0.003	0.006	–0.017	–0.009	–0.010
	(0.023)	(0.022)	(0.015)	(0.010)	(0.015)	(0.013)
Short-run (0 to 3 Yr)	0.092**	0.099**	0.075*	0.057**	0.075***	0.054**
	(0.039)	(0.043)	(0.039)	(0.026)	(0.026)	(0.025)
Medium-run (4 to 6 Yr)	0.178***	0.124***	0.135***	0.086***	0.140***	0.130***
	(0.035)	(0.041)	(0.051)	(0.025)	(0.049)	(0.042)
Overall ATT	0.116***	0.104***	0.086**	0.065***	0.086***	0.068***
	(0.034)	(0.039)	(0.034)	(0.024)	(0.026)	(0.025)
Yearly License Flow	0.410	0.410	0.410	0.387	0.387	0.387
Overall ATT / annual license flow	0.283	0.252	0.210	0.169	0.223	0.175
License Stock	9.263	9.263	9.263	9.263	9.263	9.263
Overall ATT / total license stock	0.013	0.011	0.009	0.007	0.009	0.007
Controls	FMG Ratio	+Medicaid	+Medicare	FMG Ratio	+Medicaid	+Medicare
N	315	285	285	450	423	423

Notes: The table reports difference-in-differences estimates using Callaway and Sant'Anna (2021). Pre-period, short-run, and medium-run rows average event-time effects over the indicated windows of years relative to reform; the overall ATT aggregates all post-treatment cells. Wild-bootstrap standard errors clustered at the state level are reported in parentheses. The analytic sample consists of state-level observations of eventually-treated states. The outcome in columns (1)–(3) is the number of first-time FMG licensees per 10,000 residents, from the AMA's annual medical licensure statistics; the outcome in columns (4)–(6) is the number of FMGs whose year of licensure in the American Medical Directory equals t, per 10,000 residents. Baseline control variable is the ratio of FMGs among trainees at 1964. The additional control in columns (2) and (5) is the year of Medicaid adoption (continuous); in columns (3) and (6) an indicator for an above-median share of uninsured elderly. Yearly license flow is the pre-reform-year mean of total (FMG+USMG) licenses; license stock is the wave-before-reform mean of the per-capita licensing stock. A */**/*** indicates significance at the 10/5/1% levels.

Table 4: Effects on the Stock of Licensed FMGs (per 10,000 Residents)

	(1)	(2)	(3)
Pre-Removal (≤ -2 Yr)	0.057**	0.057**	0.057**
	(0.024)	(0.025)	(0.024)
Short-run (0 to 3 Yr)	0.187**	0.175*	0.176*
	(0.094)	(0.098)	(0.097)
Honest CI ($\overline{M} = 0.25$)	[0.009, 0.361]	[–0.017, 0.363]	[–0.004, 0.353]
Honest CI ($\overline{M} = 1$)	[–0.019, 0.379]	[–0.047, 0.382]	[–0.032, 0.374]
Medium-run (2 to 9 Yr)	–0.133	–0.120	–0.114
	(0.155)	(0.149)	(0.149)
Honest CI ($\overline{M} = 0.25$)	[–0.434, 0.168]	[–0.455, 0.209]	[–0.429, 0.194]
Honest CI ($\overline{M} = 1$)	[–0.482, 0.210]	[–0.509, 0.256]	[–0.486, 0.238]
Baseline Mean	0.52	0.52	0.52
FMG trainee ratio	Yes	Yes	Yes
Medicaid adoption		Yes	
Medicare share			Yes
N	90	90	90

Notes: The table reports Callaway and Sant'Anna (2021) doubly-robust estimates for the per-capita stock of licensed foreign medical graduates, estimated at the state level and weighted by baseline state population. Estimates are binned in AMD wave time. Wild-bootstrap standard errors clustered at the state level are in parentheses. Honest confidence intervals follow Rambachan and Roth (2023), allowing the maximal post-treatment violation of parallel trends to be $\overline{M}$ times the maximal pre-treatment violation. The baseline control is the 1964 ratio of FMGs among trainees; column (2) adds the year of Medicaid adoption and column (3) the share of uninsured residents aged 65 and over. A */**/*** indicates significance at the 10/5/1% levels.

Table 5: Effects on the Share of Physicians Licensed

	(1)	(2)	(3)
Panel A. Licensed share among FMGs (percentage points)			
Pre-Removal (-5 to -2)	–0.26	–0.26	–0.26
	(3.58)	(3.73)	(3.58)
Short-run (0 to 3 Yr)	13.05***	12.49***	12.30***
	(3.95)	(4.39)	(4.40)
Honest CI ($\overline{M} = 0.25$)	[5.49, 20.76]	[3.94, 21.03]	[4.10, 20.85]
Honest CI ($\overline{M} = 1$)	[2.94, 23.94]	[1.33, 23.99]	[1.66, 23.96]
Medium-run (2 to 9 Yr)	1.66	1.68	2.37
	(8.91)	(8.62)	(8.88)
Honest CI ($\overline{M} = 0.25$)	[–18.05, 21.36]	[–18.08, 21.85]	[–17.22, 22.37]
Honest CI ($\overline{M} = 1$)	[–21.40, 25.12]	[–21.38, 25.56]	[–20.57, 26.52]
Baseline Mean	43.7	43.7	43.7
N	90	90	90
Panel B. Licensed share among all physicians (percentage points)			
Pre-Removal (-5 to -2)	0.61	0.61	0.61
	(0.51)	(0.49)	(0.51)
Short-run (0 to 3 Yr)	2.86**	2.81**	2.59**
	(1.14)	(1.10)	(1.22)
Honest CI ($\overline{M} = 0.25$)	[0.70, 5.16]	[0.62, 5.17]	[0.36, 5.05]
Honest CI ($\overline{M} = 1$)	[0.31, 6.25]	[–0.16, 6.37]	[–0.22, 6.20]
Medium-run (2 to 9 Yr)	1.19	1.00	1.02
	(1.27)	(1.39)	(1.33)
Honest CI ($\overline{M} = 0.25$)	[–1.35, 3.69]	[–1.76, 3.76]	[–1.70, 3.68]
Honest CI ($\overline{M} = 1$)	[–3.50, 5.84]	[–4.50, 6.45]	[–4.32, 6.15]
Baseline Mean	87.1	87.1	87.1
N	90	90	90
FMG trainee ratio	Yes	Yes	Yes
Medicaid adoption		Yes	
Medicare share			Yes

Notes: The table reports Callaway and Sant'Anna (2021) doubly-robust estimates for the share of physicians holding a state licence, in percentage points, estimated at the state level and weighted by baseline state population. Panel A is the licensed share among foreign medical graduates; Panel B is the licensed share among all physicians. Wild-bootstrap standard errors clustered at the state level are in parentheses. Honest confidence intervals follow Rambachan and Roth (2023), allowing the maximal post-treatment violation of parallel trends to be $\overline{M}$ times the maximal pre-treatment violation. The baseline control is the 1964 ratio of FMGs among trainees; column (2) adds the year of Medicaid adoption and column (3) the share of uninsured residents aged 65 and over. A */**/*** indicates significance at the 10/5/1% levels.

Table 6: Effects on the Experience and Quality of Licensed Physicians

	FMG			USMG		
	(1) Baseline	(2) +Medicaid	(3) +Medicare	(4) Baseline	(5) +Medicaid	(6) +Medicare
Panel A. Stock mean experience (years)						
Pre-Removal (-5 to -2)	–0.45	0.36	0.31	0.21	0.52**	0.51
	(0.85)	(0.74)	(0.87)	(0.17)	(0.20)	(0.41)
Short-run (0 to 3 Yr)	–2.50**	–0.55	–0.59	0.17	0.13	0.14
	(1.14)	(1.03)	(1.70)	(0.12)	(0.14)	(0.17)
Honest CI ($\overline{M} = 0.25$)	[–5.14, –0.09]	[–2.94, 1.84]	[–3.99, 3.07]	[–0.14, 0.44]	[–0.39, 0.53]	[–0.53, 0.49]
Honest CI ($\overline{M} = 1$)	[–7.02, 0.97]	[–5.02, 3.66]	[–5.28, 5.10]	[–0.56, 0.82]	[–1.28, 1.26]	[–1.78, 1.60]
Medium-run (2 to 9 Yr)	–3.46***	–2.52**	–3.58***	0.45***	0.60***	0.77***
	(1.29)	(1.26)	(1.01)	(0.12)	(0.14)	(0.06)
Honest CI ($\overline{M} = 0.25$)	[–6.64, 0.86]	[–5.49, 2.89]	[–4.21, 1.27]	[–0.01, 0.91]	[–0.31, 1.42]	[–0.45, 1.42]
Honest CI ($\overline{M} = 1$)	[–10.30, 4.59]	[–9.11, 6.93]	[–8.89, 6.01]	[–0.81, 1.82]	[–1.97, 3.12]	[–2.88, 3.84]
Baseline Mean		20.90			22.73	
N		92			92	
Panel B. New licensees: exam pass rate (pp; selection test, FMG only)						
Pre-period (-5 to -1)	1.09	–1.82	2.58	–	–	–
	(4.68)	(6.67)	(6.92)			
Short-run (0 to 3)	4.37	7.24	7.75	–	–	–
	(6.62)	(9.39)	(8.15)			
Honest CI ($\overline{M} = 0.25$)	[–10.42, 27.65]	[–15.56, 31.41]	[–12.58, 36.41]			
Honest CI ($\overline{M} = 1$)	[–44.31, 66.10]	[–41.62, 65.29]	[–34.87, 71.51]			
Medium-run (4 to 6)	–1.25	3.31	–3.33	–	–	–
	(3.63)	(6.00)	(4.31)			
Honest CI ($\overline{M} = 0.25$)	[–31.56, 35.10]	[–37.59, 49.85]	[–31.88, 27.26]			
Honest CI ($\overline{M} = 1$)	[–108.89, 110.24]	[–88.10, 100.01]	[–116.02, 111.39]			
Baseline Mean		70.46			–	
N		306			–	
Panel C. Stock quality (FMG: exam pass rate; USMG: US News top-10 share, pp)						
Pre-Removal (-5 to -2)	–2.22**	–2.32**	–2.31***	0.11	0.39***	0.37
	(0.94)	(0.92)	(0.90)	(0.14)	(0.14)	(0.30)
Short-run (0 to 3 Yr)	2.63***	3.85***	3.68***	–0.58***	–1.14***	–1.12**
	(0.89)	(1.01)	(1.05)	(0.22)	(0.28)	(0.47)
Honest CI ($\overline{M} = 0.25$)	[0.43, 4.75]	[1.41, 6.10]	[1.17, 6.23]	[–1.07, –0.13]	[–1.78, –0.57]	[–2.17, –0.21]
Honest CI ($\overline{M} = 1$)	[–1.59, 6.56]	[–1.09, 8.23]	[–1.14, 8.63]	[–1.21, –0.05]	[–2.16, –0.37]	[–2.69, –0.14]
Medium-run (2 to 9 Yr)	3.97***	4.15***	3.01***	–0.74**	–1.09***	–0.50
	(0.96)	(0.83)	(0.86)	(0.31)	(0.35)	(0.62)
Honest CI ($\overline{M} = 0.25$)	[–0.79, 4.77]	[–1.22, 6.06]	[–1.21, 5.85]	[–1.33, 0.33]	[–2.43, 0.29]	[–2.30, 0.07]
Honest CI ($\overline{M} = 1$)	[–5.44, 9.10]	[–6.42, 10.86]	[–6.58, 11.23]	[–1.51, 0.46]	[–2.92, 0.92]	[–3.14, 0.91]
Baseline Mean		69.89			13.85	
N		92			92	
Panel D. Stock top-10 school share, world-reputation ranking (pp; USMG only)						
Pre-Removal (-5 to -2)	–	–	–	–0.05	0.75**	0.70
				(0.24)	(0.33)	(0.94)
Short-run (0 to 3 Yr)	–	–	–	–0.63*	–1.46***	–1.43**
				(0.33)	(0.45)	(0.72)
Honest CI ($\overline{M} = 0.25$)				[–1.32, 0.08]	[–2.42, –0.61]	[–3.05, 0.02]
Honest CI ($\overline{M} = 1$)				[–1.47, 0.26]	[–2.94, –0.45]	[–3.74, 0.20]
Medium-run (2 to 9 Yr)	–	–	–	–0.38	–1.28**	–0.22
				(0.49)	(0.64)	(1.07)
Honest CI ($\overline{M} = 0.25$)				[–1.60, 0.67]	[–3.85, 0.13]	[–3.99, 0.16]
Honest CI ($\overline{M} = 1$)				[–1.84, 0.96]	[–4.47, 0.83]	[–4.84, 1.09]
Baseline Mean		–			10.53	
N		–			92	

Notes: Panel A reports mean experience of the licensed stock, measured as years since graduation from medical school. Panel B is a selection test on new licensees in the calendar year, using the average licensing-exam pass rate of the medical school from which each licensee graduated; it is defined for FMGs only. Panel C reports the quality of the licensed stock, proxied by the school-level exam pass rate for FMGs and by the share from *U.S. News* top-10 schools for USMGs. Panel D reports the top-10 share under the world-reputation ranking, defined for USMGs only. Columns (1)–(3) report foreign medical graduates and columns (4)–(6) U.S. medical graduates; within each block the first column is the baseline specification controlling for the 1964 ratio of FMGs among trainees, the second adds the year of Medicaid adoption, and the third adds the share of uninsured residents aged 65 and over. Wild-bootstrap standard errors clustered at the state level are in parentheses. A */**/*** indicates significance at the 10/5/1% levels.

Table 7: Effects on Hospital Capacity and Utilization, by Hospital Type

	Short-Term General			Long-Term & Specialty		
	(1) Baseline	(2) +Mcaid	(3) +Mcare	(1) Baseline	(2) +Mcaid	(3) +Mcare
Panel A. Beds (per 10,000)						
Pre-period (-5 to -1)	–0.42	–0.72	–0.01	0.42	–0.78	0.97
	(0.50)	(0.85)	(0.50)	(1.00)	(2.58)	(1.16)
Short-run (0 to 3)	1.51**	2.00	1.52*	1.34	0.19	2.18
	(0.62)	(1.58)	(0.81)	(1.43)	(1.99)	(1.59)
Medium-run (4 to 6)	2.46***	2.62	2.53***	6.56*	1.04	6.87*
	(0.91)	(2.28)	(0.90)	(3.89)	(4.13)	(3.66)
Overall ATT	1.77***	2.11	1.83**	2.78	0.34	3.62*
	(0.57)	(1.54)	(0.71)	(2.07)	(2.08)	(2.17)
Baseline Mean		35.7			36.9	
Panel B. Employment (per 10,000)						
Pre-period (-5 to -1)	0.94	0.58	0.75	0.24	–0.06	1.30
	(0.85)	(1.03)	(0.95)	(0.79)	(1.20)	(1.05)
Short-run (0 to 3)	2.88*	5.26***	2.43	2.94*	0.54	3.24**
	(1.71)	(1.93)	(1.83)	(1.61)	(2.38)	(1.60)
Medium-run (4 to 6)	–0.64	4.17	–0.45	7.04***	4.11**	7.22***
	(3.87)	(6.87)	(3.89)	(1.40)	(1.87)	(1.26)
Overall ATT	1.91	5.07**	1.54	4.07***	1.17	4.46***
	(2.13)	(2.39)	(2.03)	(1.47)	(2.14)	(1.43)
Baseline Mean		62.2			17.3	
Panel C. Admissions (per 10,000)						
Pre-period (-5 to -1)	–27.54	–22.08	–50.98*	–1.25	3.41	4.06
	(25.09)	(21.93)	(28.14)	(3.77)	(5.26)	(5.18)
Short-run (0 to 3)	40.13	51.71	34.58	9.29	8.50	12.22
	(24.87)	(34.48)	(25.84)	(6.78)	(9.16)	(7.48)
Medium-run (4 to 6)	60.83**	98.99***	62.44**	18.93*	–4.44	19.65**
	(29.96)	(37.47)	(28.49)	(11.25)	(10.07)	(9.94)
Overall ATT	45.86**	60.04*	43.13*	11.96	6.22	14.50*
	(22.70)	(34.89)	(23.77)	(7.66)	(8.73)	(8.06)
Baseline Mean		1293.1			64.1	
Panel D. Inpatient Days (per 10,000)						
Pre-period (-5 to -1)	–0.69*	–1.06*	–0.65	1.36	–0.49	1.03
	(0.37)	(0.64)	(0.52)	(2.25)	(2.99)	(2.61)
Short-run (0 to 3)	0.75	1.54*	0.83	1.92	–0.01	2.03
	(0.59)	(0.88)	(0.73)	(1.55)	(1.67)	(1.76)
Medium-run (4 to 6)	0.21	1.72	0.28	4.67*	0.10	4.86**
	(1.70)	(2.62)	(1.65)	(2.39)	(3.82)	(2.17)
Overall ATT	0.60	1.57	0.66	2.68	0.01	2.90
	(0.82)	(1.06)	(0.91)	(1.76)	(1.83)	(1.92)
Baseline Mean		26.4			30.8	
N			363			

Notes: Estimates from the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021), using not-yet-treated states as the comparison group. Columns (1)–(3) within each block add controls cumulatively: the baseline 1964 FMG-among-trainees ratio, then Medicaid adoption timing, then Medicare intensity. Short-run pools event years 0–3 and medium-run pools event years 4–6. Standard errors clustered at the state level in parentheses. A */**/*** indicates significance at the 10/5/1% level.

Table 8: County Heterogeneity in Hospital Input and Output

	STG employment			STG beds			STG admissions			STG patient days		
	Low	High	Diff.	Low	High	Diff.	Low	High	Diff.	Low	High	Diff.
Panel A. Split by Baseline median income												
Pre-period	4.0	4.9	–0.9	–0.4	1.7	–2.2	109	–11	119*	72	237	–165
	(1.8)	(1.9)	[0.665]	(1.5)	(1.6)	[0.428]	(71)	(34)	[0.083]	(473)	(442)	[0.840]
Short-run (0 to 3 Yr)	4.4	–7.4	11.9**	2.5	1.1	1.4	93	–80	172**	3655	–569	4224***
	(2.5)	(4.0)	[0.017]	(0.8)	(1.8)	[0.419]	(55)	(40)	[0.029]	(784)	(388)	[0.000]
N counties	221	222		221	222		221	222		221	222	
N observations	3757	3774		3757	3774		3757	3774		3757	3774	
Panel B. Split by Baseline physician density												
Pre-period	3.7	4.3	–0.6	–0.9	1.4	–2.3	39	–18	57*	–295	452	–747
	(1.5)	(2.6)	[0.836]	(1.3)	(1.3)	[0.316]	(29)	(35)	[0.094]	(493)	(506)	[0.415]
Short-run (0 to 3 Yr)	0.2	–12.9	13.0*	0.3	0.7	–0.3	26	–124	150*	533	–838	1371**
	(2.6)	(5.7)	[0.061]	(1.0)	(1.5)	[0.831]	(40)	(48)	[0.051]	(340)	(418)	[0.026]
N counties	221	222		221	222		221	222		221	222	
N observations	3757	3774		3757	3774		3757	3774		3757	3774	
Panel C. Split by Baseline urban share												
Pre-period	1.9	5.2	–3.3	1.0	2.0	–0.9	7	–4	11	164	581	–417
	(1.2)	(2.7)	[0.230]	(0.9)	(1.7)	[0.588]	(21)	(41)	[0.819]	(220)	(591)	[0.527]
Short-run (0 to 3 Yr)	–1.5	–12.7	11.2*	1.1	0.8	0.3	–40	–115	75	378	–761	1139***
	(1.2)	(5.9)	[0.073]	(0.7)	(1.7)	[0.822]	(21)	(48)	[0.130]	(309)	(391)	[0.004]
N counties	221	222		221	222		221	222		221	222	
N observations	3757	3774		3757	3774		3757	3774		3757	3774	
Panel D. Split by baseline Medicaid exposure												
	High	Low	Diff.	High	Low	Diff.	High	Low	Diff.	High	Low	Diff.
Pre-period	1.8	4.5	–2.7	0.0	1.3	–1.3	16	–1	17	110	574	–465
	(2.4)	(3.6)	[0.533]	(3.9)	(1.7)	[0.751]	(40)	(61)	[0.805]	(246)	(828)	[0.617]
Short-run (0 to 3 Yr)	7.4	–16.4	23.8***	5.2	–1.2	6.4**	68	–144	212*	1669	–1024	2693***
	(3.1)	(7.1)	[0.006]	(2.7)	(1.1)	[0.015]	(67)	(63)	[0.054]	(509)	(438)	[0.001]
N counties	172	271		172	271		172	271		172	271	
N observations	2924	4607		2924	4607		2924	4607		2924	4607	
FMG trainee ratio							Yes					
Baseline nonwhite population ratio							Yes					

Notes: Estimates are obtained by applying the Callaway and Sant'Anna (2021) estimator separately within each predetermined county partition, for short-term general (STG) hospitals. Each outcome block reports the two groups and their difference, computed as the first column minus the second: columns (1)–(3) non-physician employment, (4)–(6) beds, (7)–(9) admissions, and (10)–(12) patient days. Panels A–C list the low group first; Panel D lists the high-Medicaid-exposure group first, so that a positive difference indicates a larger response in high-exposure counties. Standard errors clustered at the state level are in parentheses; Wald-test p-values for equality across partitions are in brackets. Panels split counties at the 1964 median of household income, physician density, urban share, and AFDC caseload rate. Arkansas is omitted because county-level characteristics are not available in the 1973 AMD year. A */**/*** indicates significance at the 10/5/1% levels.

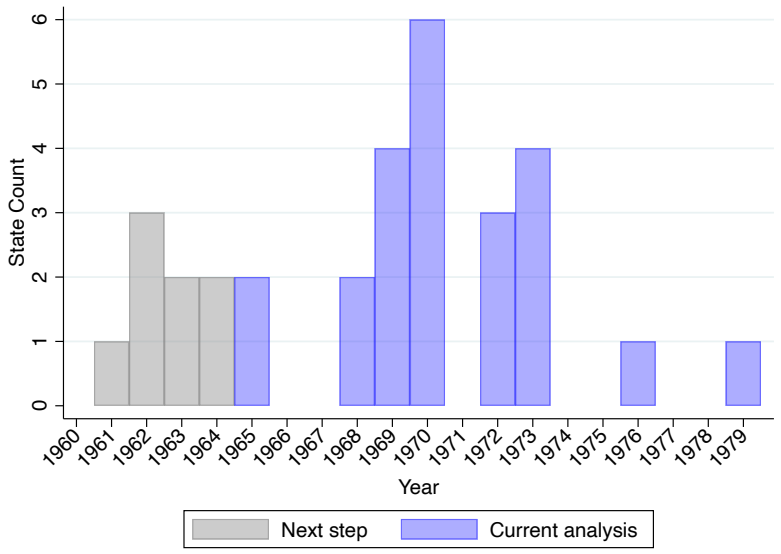


Figure 1: Removal Timing across years

Source: Data from Journal of the American Medical Association, 1960-1980.

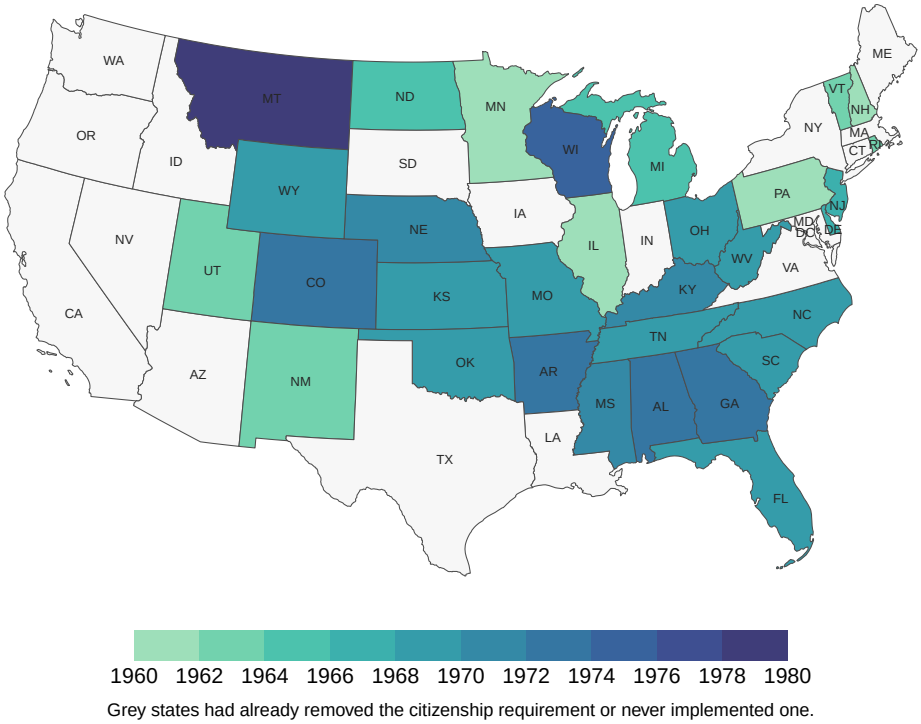


Figure 2: Removal Timing across US States

Source: Data from Journal of the American Medical Association, 1960-1980.

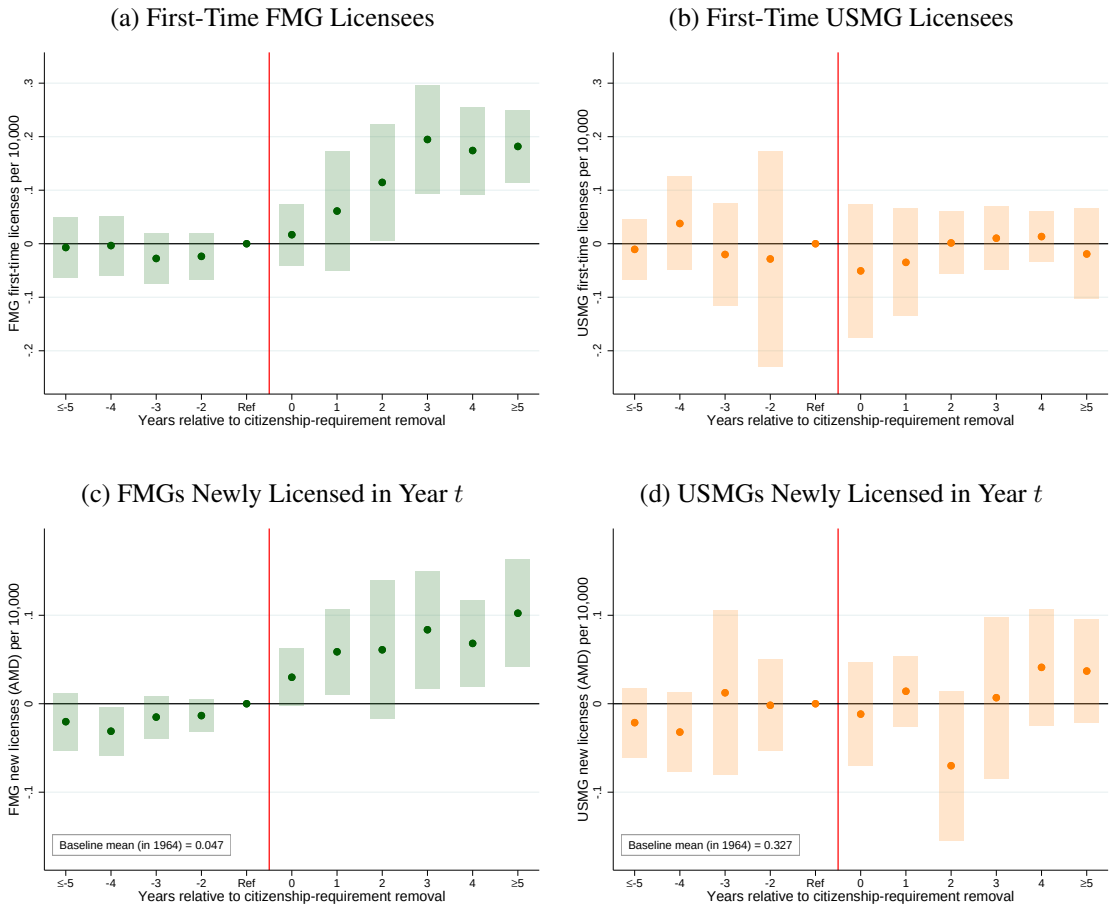
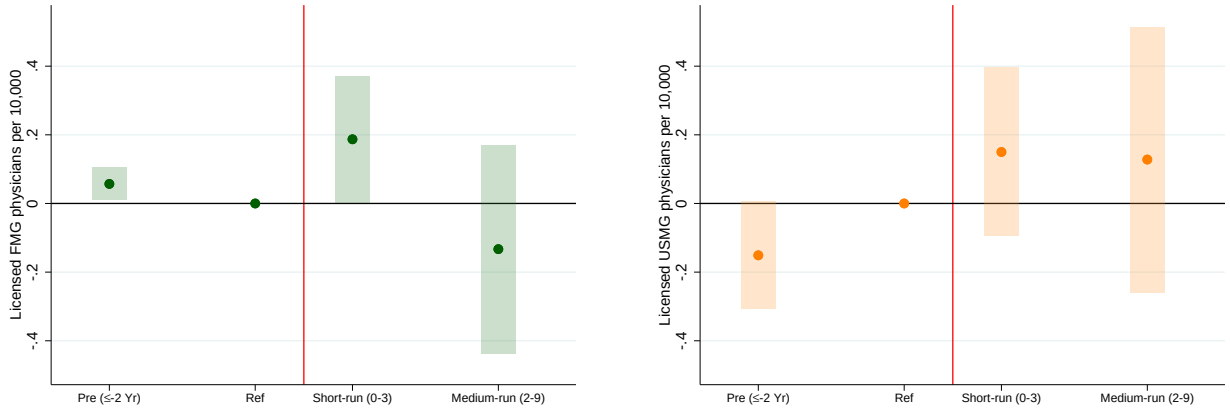


Figure 3: Event Study on Licensing Flows: FMGs and USMGs

Notes: Event-study estimates of the effect of removing the citizenship requirement on licensing flows per 10,000 residents, estimated with the Callaway and Sant'Anna (2021) doubly-robust estimator using not-yet-treated states as controls and the 1964 ratio of FMGs among trainees as the baseline covariate; specifications correspond to columns (1) and (4) of Table 3. Panels (a) and (b) use the number of first-time licensees from the AMA's annual medical licensure statistics; panels (c) and (d) use the number of physicians whose year of licensure in the American Medical Directory equals t. Confidence intervals (95%) are based on wild-bootstrap standard errors clustered at the state level.



(a) Foreign-trained physicians (FMGs)

(b) U.S.-trained physicians (USMGs)

Figure 4: Event Study on the Stock of Licensed Physicians, by Place of Training

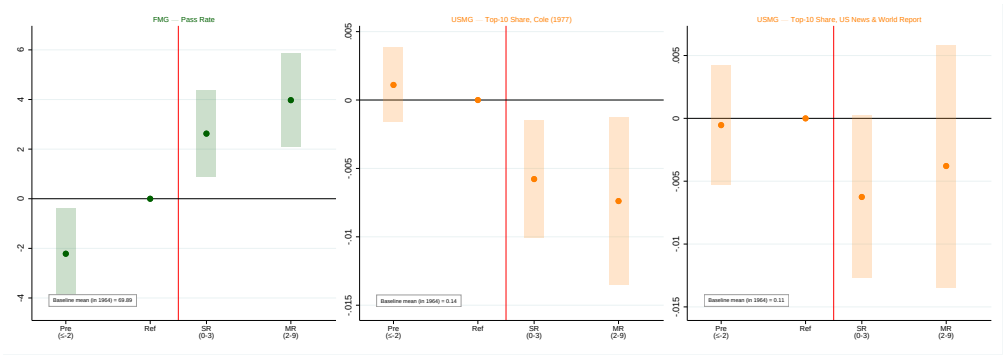
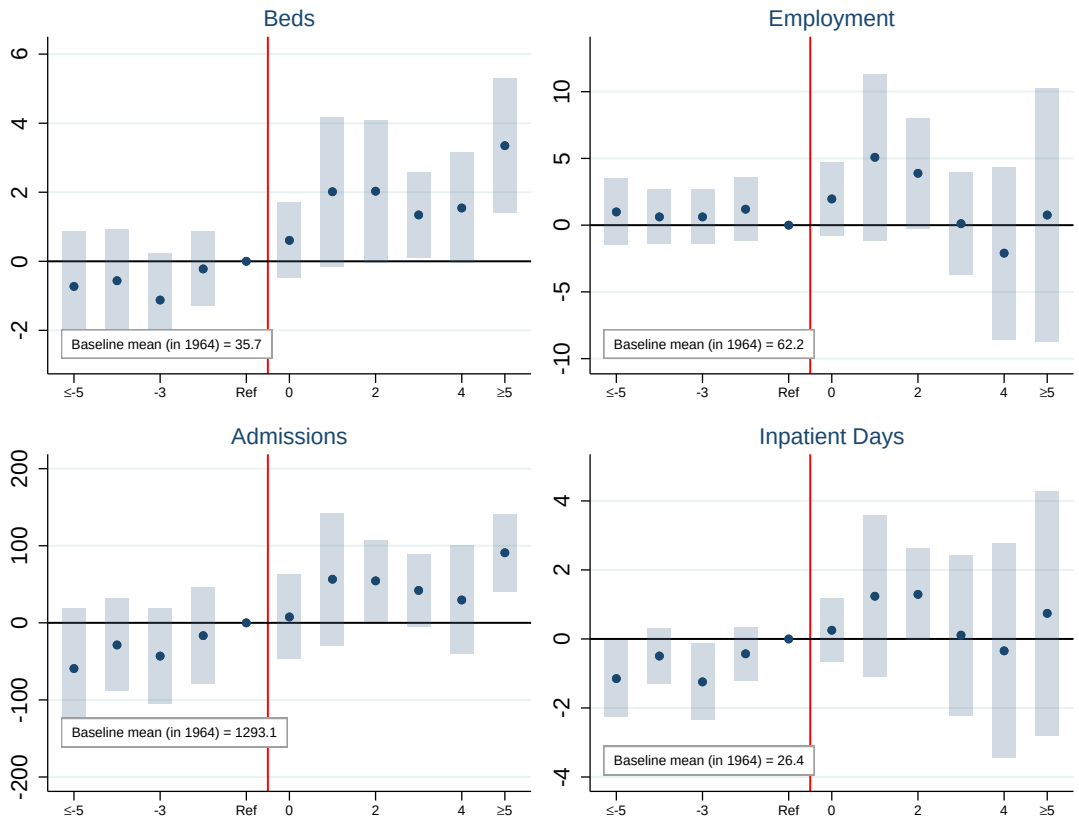
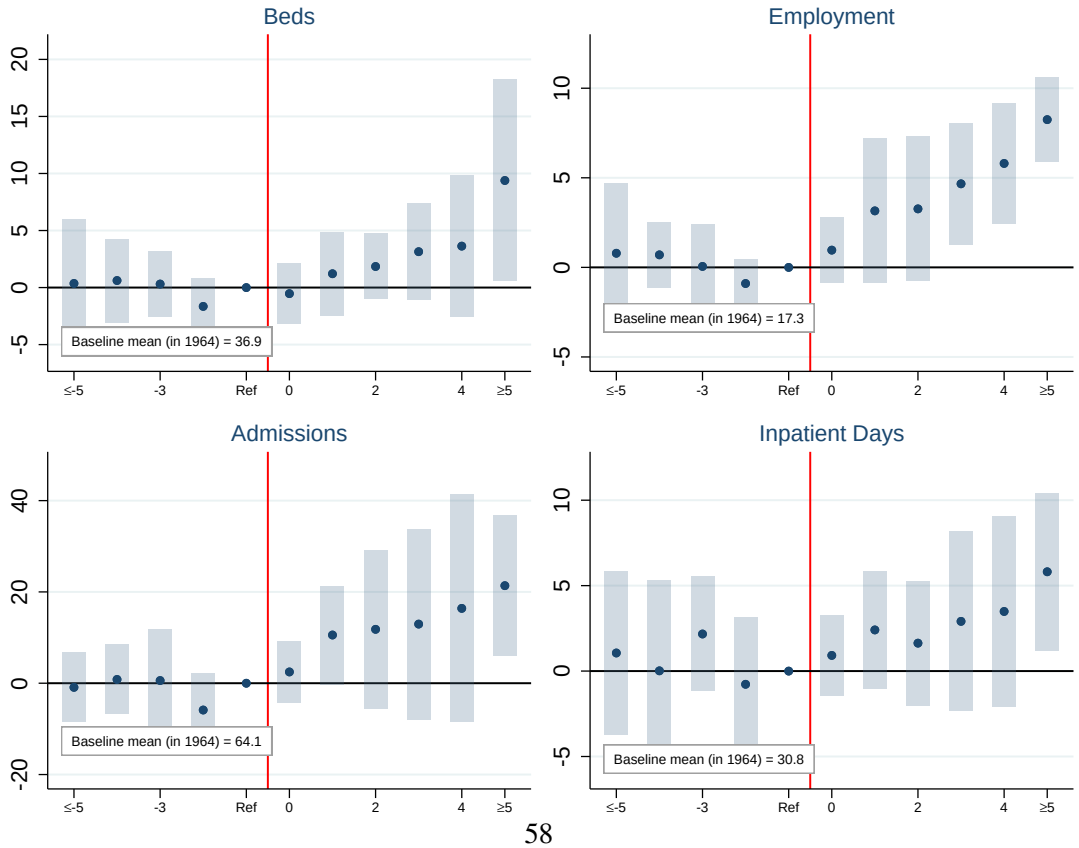


Figure 5: Average Quality of the Licensed Physician Stock: FMGs and USMGs

Notes: FMG quality is the average licensing-exam pass rate of the licensee's medical school. USMG quality uses the Cole and Lipton (1977) rating and the present-day *U.S. News & World Report* ranking.



(a) Short-Term General Hospitals



(b) Long-Term General Hospitals

Figure 6: Event Study on Hospital Capacity and Utilization

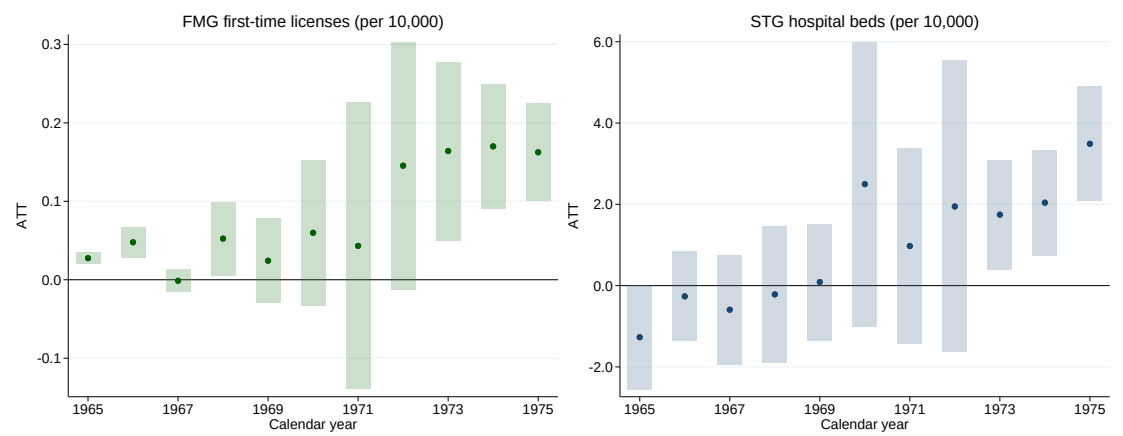


Figure 7: Treatment Effect Heterogeneity by Calendar Year of Reform

Notes: Panel (a), at left, reports FMG first-time licenses; panel (b), at right, reports short-term general hospital beds. Callaway–Sant'Anna group-average ATTs aggregated by calendar year of citizenship-requirement removal, estimated on the analysis-state sample, for FMG licensing flow and short-term general hospital beds. Bars are 95% confidence intervals, clustered by state. The dashed vertical line marks 1971, the first full year after Public Law 91-225 (April 1970) relaxed the J-1 home-country return requirement (Butter and Sweet, 1977).

A Appendix

A.1 Does Wisconsin Explain All the Treatment Effect?

Leave-one-state-out estimates in the Robustness section shows the sensitivity of estimates when dropping Wisconsin. One potential concern is that these estimates completely relies on Wisconsin, i.e. any cohort treatment effect is identified only when their control group contains Wisconsin. This could weaken the robustness of the finding if the identification recovers an unusual change in the control state Wisconsin rather than the removal of citizenship requirement.

To determine the source of this sensitivity, I decompose the aggregate treatment effect into its underlying cohort-by-calendar-year treatment effects and aggregation weights. Table A1 reports, for each cohort-by-timing group, its aggregation weight and its contribution to the pooled ATT in the full sample (Columns 1–2), together with the corresponding values when Wisconsin is excluded (Columns 3–4). This decomposition distinguishes sensitivity arising from the loss of identifying support from that arising through changes in the estimated treatment effects.

As shown in Column 5 of panel A, holding the full-sample aggregation weights fixed and using the leave-Wisconsin-out estimates for cells that remain estimable, these cells contribute 0.087 to the full-sample post-reform ATT of 0.116.[28] Thus, 75 percent of the aggregate effect is recovered from comparisons that do not require Wisconsin. The remaining 25 percent is contributed by post-1970 cells for which Wisconsin is the only eligible not-yet-treated comparison state. For hospital beds, panel B suggests dropping WI even increases some cohort by time estimates so that the aggregate estimate becomes larger than the full sample one.

The decomposition exercise therefore does not support that the finding is generated entirely by Wisconsin's outcome trajectory. Instead, Wisconsin provides crucial identifying support in later calendar years.

A.2 Design-Calibrated Monte Carlo Diagnostics

Questions. How reliable are the paper's estimation and inference procedures in finite samples? With 23 treated states, an anchored comparison group (identification leans on two late-treated states, Wisconsin and Montana), and a common pre-trend, it is not obvious that the point estimates are unbiased or that the standard errors are correctly sized. I answer three questions with a design-calibrated Monte Carlo diagnostic, which fixes the treatment effect by construction: (i) does the estimator recover it without bias; (ii) does

[28] Columns are cell-level aggregates: column (5) combines the weights underlying column (1) with the estimates underlying column (4) cell by cell, so it does not equal (1)×(4)/(3) at the group level; cells unidentifiable without Wisconsin enter as zero.

Table A1: Decomposition of the Aggregate Effect With and Without Wisconsin

Cohort × timing group	Full sample		Dropping WI		WI-free est.
	Weight	Contrib.	Weight	Contrib.	(preserving weight)
	(1)	(2)	(3)	(4)	(5)
Panel A. FMG first-time licenses (per 10,000)					
$g \leq 1970, t \leq 1970$	0.269	0.011	0.525	0.017	0.009
$g \leq 1970, t > 1970$	0.552	0.093	0.436	0.047	0.079
$g > 1970, t \leq 1970$			*empty by construction*		
$g > 1970, t > 1970$	0.179	0.012	0.039	−0.003	−0.001
Total (= pooled ATT)	1.000	0.116	1.000	0.062	0.086
Estimation N	294		247		
Panel B. STG hospital beds (per 10,000)					
$g \leq 1970, t \leq 1970$	0.269	0.220	0.525	0.488	0.251
$g \leq 1970, t > 1970$	0.552	1.148	0.436	0.790	1.901
$g > 1970, t \leq 1970$			*empty by construction*		
$g > 1970, t > 1970$	0.179	0.402	0.039	−0.013	−0.007
Total (= pooled ATT)	1.000	1.770	1.000	1.265	2.145
Estimation N	363		313		

Notes: Columns (1)–(2) report each cohort-by-timing group's aggregation weight and its contribution to the pooled ATT in the full sample; columns (3)–(4) report the corresponding values when Wisconsin is excluded. Column (5) combines the full-sample weights with the leave-Wisconsin-out estimates cell by cell, so it does not equal (1)×(4)/(3) at the group level; cells that are unidentifiable without Wisconsin enter as zero. Panel A reports FMG licensing flow and Panel B hospital beds.

inference reject a true null at the nominal rate; and (iii) does the design have power to detect an effect of the estimated size? The procedure shares the logic of Bertrand et al. (2004)—evaluating inference by how often a known null is rejected, letting the data's own dependence structure generate the null distribution—but inverts the implementation. Rather than assigning placebo policies to real outcomes, I hold the actual treatment configuration fixed and simulate outcomes from a process calibrated to the data.

Approach. *Data-generating process.* Each of 500 replications constructs a synthetic state–year outcome panel by overlaying a treatment effect, known by construction, on the estimated deterministic and stochastic structure of the actual data:

$$Y^*_{st} = \hat{\alpha}_s + \hat{\gamma}_t + \hat{\lambda}\, e_{st} + \hat{\kappa}\, e_{st}\, \mathbf{1}\{s = \text{WI}\} + \tau\, \mathbf{1}\{e_{st} \geq 0\} + u_{st}, \qquad u_{st} = \hat{\rho}\, u_{s,t-1} + \varepsilon_{st}, \tag{3}$$

where $e_{st} = t - G_s$ is event time relative to state s's actual reform year and $\varepsilon_{st} \sim N(0, \hat{\sigma}^2)$. The stochastic and trend components are calibrated on *pre-treatment information only* (not-yet-treated observations, $t < G_s$), so the process embeds the specific threats present in my panel—a common differential pre-trend (λ), Wisconsin's offsetting trend (κ), and serially correlated state shocks (ρ, σ)—without letting post-treatment outcomes influence the calibration. The parameter of interest is τ, the treatment effect I control: each replication applies the exact baseline estimation procedure to Y^*, and the distribution of the resulting estimates and t-statistics across replications answers the three questions above.

Parameter specification. The fixed effects $\hat{\alpha}_s$ and $\hat{\gamma}_t$ come from a population-weighted regression of the outcome on state and year indicators over the full sample (they set the outcome levels and are absorbed by the estimator). The trend parameters $\hat{\lambda}$ and $\hat{\kappa}$ come from regressing on event time and its Wisconsin interaction the residuals of a second, separate fixed-effects regression estimated on not-yet-treated observations only; $\hat{\rho}$ and $\hat{\sigma}$ come from an AR(1) fit to the same within-state residuals ($\hat{\sigma}$ is the innovation standard deviation), with stationary initialization of u_{s0}. The calibrated values are reported in the notes to Table A2; serial correlation is modest for the USMG flow and essentially absent for the FMG flow, and both differential trends are small relative to the injected effects. The truth τ is set by design: $\tau = 0$ for the size experiments (run with and without the trend terms, and repeated with multiplier-bootstrap standard errors to test whether the bootstrap corrects any distortion); $\tau = +0.116$ (the estimated FMG baseline effect) and $\tau = -0.05$ ($\approx$15% of the mean USMG flow) for the power experiments.

Data structure. The simulated panels are identical in structure to the estimation samples behind the state-level results: the same 23 states and reform years listed in Table 2 and mapped in Figures 1 and 2,

the same annual coverage (1962–1979 for the FMG flow, balanced with $N = 414$; 1959–1979 for the USMG flow, $N = 478$ with one state unobserved in 1973–77), the same baseline covariate, population weights, and missingness pattern, and the same role for Wisconsin (1976) and Montana (1979) as late-treated cohorts anchoring the not-yet-treated comparisons. Nothing about the design is simulated; only the outcome realizations are.

The estimates are approximately unbiased. Table A2 reports the Monte Carlo results. Across all configurations, the mean estimate is within 0.005 of the true effect, including under the calibrated differential trends, the Wisconsin trend, the anchored comparison group, and population weighting. These features therefore generate little bias in the point estimates.

Inference over-rejects, but the main FMG result remains statistically detectable. Conventional inference is oversized: under the null, the design rejects about 20% of the time at the nominal 5% level because reported standard errors understate the true sampling variation. The distortion is unchanged when the calibrated trend terms are removed and when multiplier-bootstrap standard errors replace analytic ones (Table A2), so it reflects the small number of clusters rather than the trend structure or the variance estimator. The two variance estimators are also numerically interchangeable in this design: in the USMG null runs, mean standard errors are 0.030 under both, null rejection rates are 19.4% (analytic) and 17.8% (bootstrap), and the bootstrap-derived critical value is $|t| = 2.80$ against the analytic 2.98. Critical values derived from either procedure therefore apply to the bootstrap-based t-statistics reported in the paper's tables. I therefore use the empirical null distribution to calculate outcome-specific critical values that deliver a 5% rejection rate. One caution on magnitudes: because the data-generating process is calibrated to pre-treatment residuals, its noise level is quieter than the actual data's (the observed FMG standard error is 0.033, against a true sampling SD of 0.023 inside the simulation). The portable quantity is therefore not the absolute SD but the ratio of reported to true variability—about 0.65—which is why the correction is expressed as unitless critical values rather than corrected standard errors.

For the FMG flow, power is essentially one at both the conventional and size-corrected thresholds, and the observed estimate ($t = 3.49$) exceeds the size-corrected critical value of $|t| = 2.83$. The main FMG result therefore remains statistically detectable after accounting for the design's over-rejection.

For the USMG flow, the size-corrected critical value is $|t| = 2.98$ and the corresponding confidence interval is about 1.5 times wider than the conventional interval. The null estimate therefore does not rule out modest crowd-out: effects smaller than roughly 0.14 per 10,000, about 40% of mean USMG flow, could

remain undetected. Consistent with this bound, when a crowd-out effect of -0.05 (15% of the mean flow) is injected, the design detects it in only 22% of replications at the size-corrected threshold (Table A2, last USMG row). The same understatement extends to the Rambachan–Roth robust confidence sets reported for the stock and quality outcomes, which are constructed from the same estimated variances; their reported breakdown values should accordingly be read as modestly optimistic. Because this simulation uses annual outcomes and a homogeneous treatment effect, the consequences of sparse wave-based measurement are examined separately.

Table A2: Design-Calibrated Monte Carlo: Bias, Size, and Power

Outcome	Configuration	Mean est.	True SD	Mean SE	Rej. @1.96	95th pct. $\lvert t\rvert$	Adj. power
USMG flow	$\tau = 0$, trends, analytic	0.005	0.046	0.030	0.194	2.98	–
USMG flow	$\tau = 0$, no trends, analytic	0.002	0.046	0.030	0.186	3.05	–
USMG flow	$\tau = 0$, trends, bootstrap	0.004	0.043	0.030	0.178	2.80	–
USMG flow	$\tau = -0.05$, trends, analytic	-0.045	0.046	0.030	0.418	–	0.215
FMG flow	$\tau = 0$, trends, analytic	-0.003	0.023	0.015	0.208	2.83	–
FMG flow	$\tau = 0$, no trends, analytic	0.001	0.023	0.015	0.208	2.80	–
FMG flow	$\tau = +0.116$, trends, analytic	0.113	0.023	0.015	0.992	–	0.984

Notes: 500 replications per configuration. "Mean est." is the average estimate across replications (the truth is τ); "True SD" the standard deviation of the estimates; "Mean SE" the average reported standard error; "Rej. @1.96" the share of replications with $|t| > 1.96$ (size in the $\tau = 0$ rows, nominal power otherwise); "95th pct. $|t|$" the size-corrected 5% critical value from the null replications; "Adj. power" the size-adjusted power: the rejection rate judged at that corrected threshold. A dash marks cells where the statistic is not defined: corrected critical values exist only for null ($\tau = 0$) configurations, and size-adjusted power only for configurations with a nonzero injected effect. Calibrated data-generating-process parameters $(\hat{\rho}, \hat{\sigma}, \hat{\lambda}, \hat{\kappa})$: USMG flow $(0.195,\ 0.066,\ +0.0015,\ -0.0011)$; FMG flow $(-0.05,\ 0.034,\ -0.0009,\ +0.0013)$. Trends refers to the calibrated differential pre-trend (λ) and Wisconsin slope (κ) in the data-generating process.

A.3 Bias and Precision Costs of Observing Outcomes at Four AMD Waves

Questions. Several outcomes in this study—the licensed physician stock, its composition, and its quality—are observed only at the AMD survey waves of 1964, 1966, 1969, and 1973 rather than annually. This exercise asks what that sparse outcome measurement, by itself, does to the staggered difference-in-differences estimates along two dimensions. First, does the wave-based procedure recover the treatment effect it targets, and how does that target differ from the annual design's when treatment effects evolve with exposure? Second, how much precision and power are lost when outcomes are observed at only four waves, and does inference remain correctly sized?

Sparse observation dates alter the design mechanically even before any identification concern arises. States adopting between the same survey waves cannot be distinguished by exposure length at those observed dates, which coarsens treatment cohorts when effects are dynamic. Sparse measurement also reduces the within-state variation available for clustered inference, potentially worsening finite-sample distortions.

I quantify these effects by simulation. I generate annual outcomes using the process in Appendix A.2, impose a known treatment effect, and apply the same estimation procedure to both the annual and four-wave samples. Each replication records the estimate, the estimand implied by the design, and the true average post-treatment effect, allowing me to separate estimator bias from changes in what the design targets. Null simulations are then used to assess test size, critical values, and power.

Approach. *Design comparison.* Each simulated outcome draw is estimated twice using the same baseline CS-DiD procedure, including the same controls, weights, and not-yet-treated comparison groups. The annual design (L0) uses the full 1962–1975 panel and exact adoption years, while the wave design (LW) retains only the 1964, 1966, 1969, and 1973 observations.

Because the estimation window ends in 1975, Wisconsin (1976) and Montana (1979) serve as never-treated controls in *both* designs here, whereas in Appendix A.2 (panel through 1979) they enter as late-treated cohorts. This is the comparison-group role that the paper's Wisconsin robustness analysis examines, so the two appendices bracket both conventions.

Sparse outcome measurement mechanically coarsens treatment timing. For example, 1968 and 1969 adopters are both untreated in 1966 and treated by 1969, so the four-wave data cannot separately identify their effects at different exposure lengths without additional assumptions. This matters when treatment effects evolve with time since adoption, but not when effects are constant.

Both designs are applied to the same simulated outcome draw in each replication. The comparison is therefore paired, so differences in bias, precision, size, or power reflect outcome availability rather than different realizations of the simulated data.

Data-generating processes and effect paths. I generate annual outcomes using the data-generating process in Appendix A.2 under two environments. The first imposes parallel trends by construction. The second uses the fixed effects, differential trends, and error process calibrated from the 1962–1975 FMG-flow panel.

I consider three treatment-effect paths, each defined relative to a state's actual adoption year: (i) a constant effect of 0.116; (ii) an effect that increases with exposure, from 0.029 to 0.203 over six years; and (iii) an effect of 0.116 that begins only in 1971. I also simulate a zero-effect case in each environment to evaluate test size and calibrate critical values.

Targets and performance metrics. For each design, I compute its design-implied estimand by applying the estimation procedure to effectively noise-free outcomes. This separates two sources of discrepancy: the estimate–estimand gap measures estimator bias, while the estimand–true-effect gap captures differences in the effect targeted by the design.

I define the true effect as the population-weighted average treatment effect across all treated state-years. I also compute the corresponding average using only the four observed AMD waves while retaining exact adoption dates. Comparing these targets helps distinguish the effect of sparse outcome measurement from the additional effect of cohort coarsening.

I summarize finite-sample performance using bias, empirical standard deviation, test size at the conventional critical value, the critical value required to obtain the nominal size, and power evaluated using that size-corrected critical value.

The wave design is approximately unbiased for its estimand, with limited changes in the estimand except for late-emerging effects. Estimator bias is negligible in both designs (Table A3, Panel B): below 0.001 for the annual design and about 0.0007 for the wave design, in each case smaller than the Monte Carlo standard error of the mean estimate (0.001 and 0.0016 with 500 replications). Moving from annual to sparse outcome measurement therefore introduces no detectable estimator bias.

Sparse measurement also changes the estimand relatively little when effects are constant or gradually increasing. With a constant effect, the wave estimand equals the true average effect (0.116) under parallel trends; under calibrated trends both designs fall 0.003–0.005 short, the deterministic bias the differential trends impose. Under the ramp, the annual and wave estimands are 0.096 and 0.093, respectively, relative to a true average effect of 0.112. The difference is larger when effects emerge only after 1971: the wave estimand is 0.075, about 12% below the true average effect of 0.085, because 1973 is the only observed wave after the effect begins. The "Wave-sampled" column helps separate the sources of estimand movement, though the two paths differ. Under the ramp, the annual design's own estimand (0.096) already sits below the population-weighted truth (0.112), because the estimator's event-time aggregation weights exposure lengths differently from a treated-state-year average; that gap is a property of the target, present with annual data, not a cost of sparse measurement. The wave design adds only 0.003 to it. For the late-emerging effect the annual estimand equals the truth (0.085), so the wave design's shortfall is entirely a measurement cost: observing outcomes only at the waves moves the target to 0.078, and pooling adopters within wave intervals a further 0.003 to 0.075—observation timing, not cohort coarsening, accounts for most of it.

The main cost of sparse measurement is lower precision and power. In the parallel-trends environment (the calibrated environment is similar throughout; Table A3, Panel B), the empirical standard deviation rises from about 0.023 under annual measurement to 0.036 under the four-wave design. Conventional clustered inference also becomes more oversized: the null rejection rate at the nominal 5% level rises from about 20%

to 32%, and the corresponding size-corrected critical value rises from approximately 2.7–2.8 under annual measurement to 3.7–3.8 under the four-wave design (in units of $|t|$).

Using these size-corrected critical values, power falls from 0.99 to 0.84 for a constant effect, from 0.97 to 0.71 for a gradually increasing effect, and from 0.91 to 0.39 for a late-emerging effect. Sparse measurement therefore matters primarily through reduced precision and detection power, especially when treatment effects emerge late.

Table A3: Monte Carlo with Wave-Sampled Outcomes: Estimands and Replication Performance

Panel A. Exact estimands by effect path

Effect path	Truth	Wave-sampled	Annual (flat)	Wave (flat)	Annual (cal.)	Wave (cal.)
Constant	0.116	0.116	0.116	0.116	0.113	0.111
Ramp	0.112	0.102	0.096	0.093	0.093	0.088
Late ($\geq$1971)	0.085	0.078	0.085	0.075	0.081	0.070

Panel B. Replication performance (500 reps)

Environment	Design	Bias ($\times 10^3$): const / ramp / late	True SD	Mean SE	Size @1.96	Corr. crit	Adj. power: const / ramp / late
Flat trends	Annual	0.1 / 0.1 / 0.1	0.023	0.015	0.193	2.71	0.99 / 0.97 / 0.91
Flat trends	Wave	0.7 / 0.7 / 0.7	0.036	0.019	0.323	3.84	0.84 / 0.71 / 0.39
Calibrated	Annual	0.2 / $-$0.0 / 0.1	0.023	0.015	0.202	2.81	0.98 / 0.94 / 0.88
Calibrated	Wave	0.7 / 0.7 / 0.7	0.036	0.019	0.326	3.74	0.82 / 0.68 / 0.37

Notes: Effect paths, each starting at the state's actual adoption year: constant $= 0.116$; ramp rising from 0.029 to 0.203 over six years of exposure; late $= 0.116$ in calendar years ≥ 1971 only. Panel A reports exact design estimands from runs with negligible noise; "Truth" is the population-weighted average true effect over treated state-years and "Wave-sampled" the same average over treated state-waves; calibrated columns include the deterministic bias implied by the estimated differential trends. Panel B: "Bias" is the mean estimate minus the design's own Panel A estimand in the same environment (reported $\times 10^3$), so it isolates estimator bias from estimand movement; "True SD" is the standard deviation of estimates across replications; "Mean SE" the average reported standard error in the null configuration; "Size @1.96" the rejection rate under $\tau = 0$; "Corr. crit" the 95th percentile of the null $|t|$; "Adj. power" the rejection rate at that corrected threshold under each effect path. All comparisons are paired: both designs estimate the same simulated draws. Both designs use the annual panel restricted to 1962–1975—deliberately, so that the annual benchmark ends with the wave grid's last observation (1973) rather than Appendix A.2's 1979; as a consequence, Wisconsin (1976) and Montana (1979) serve as never-treated controls in both designs here, whereas in Appendix A.2 they are late-treated cohorts. The data-generating-process parameters are re-estimated on this window and are essentially unchanged from those in the notes to Table A2 (values reported in the Approach).

A.4 Does the Ceiling Effect Explain Estimates for Hospital Beds?

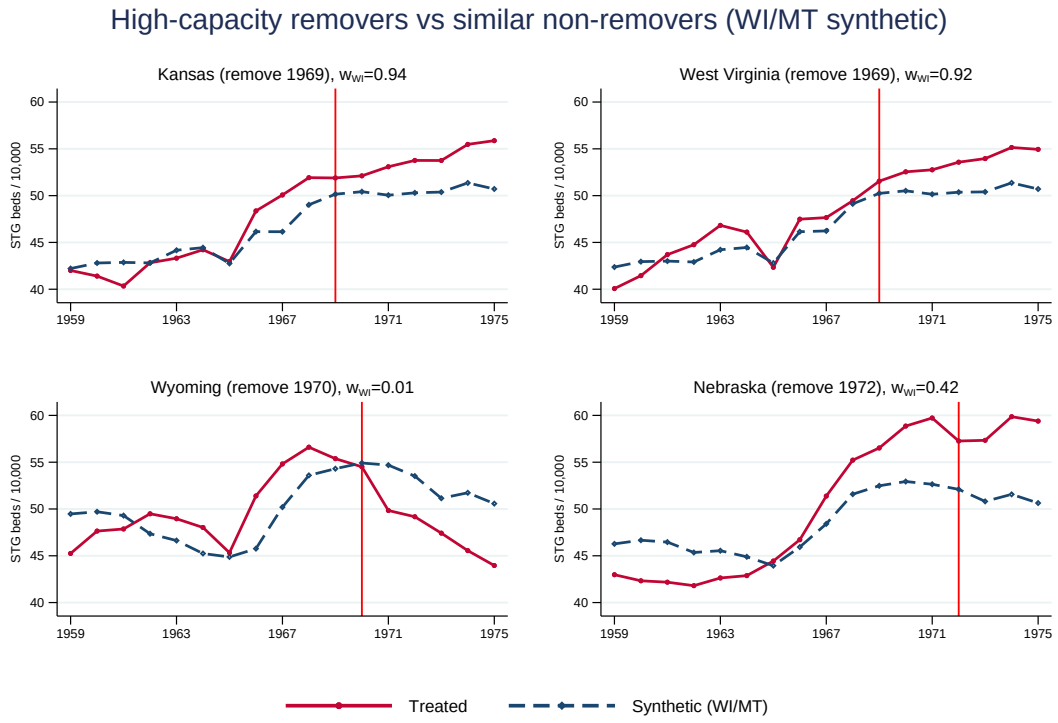


Figure A1: Synthetic Control for Hospital Beds in High-Capacity Treated States

Notes: Each panel plots the observed bed trajectory for one high-baseline-capacity treated state against a synthetic counterfactual constructed from the latest-treated group (Wisconsin and Montana). If the estimated effect merely reflected a ceiling on high-capacity states, the treated series would flatten in step with its counterfactual.

If the results is completely explained by an unusual change in the control state Wisconsin rather than the comparison between other treatment and not yet treated group, then it is impossible to differentiate whether the treatment effect to the removal of citizenship, or other alternative explanations. For instance, based on the raw trend in Figure A18, an alternative explanation for the expansion in beds is a ceiling effect, where later treated states are those with the highest baseline beds capacity, therefore its growth is capped in coincidence with the timing of the treatment.

To evaluate this explanation, I conduct synthetic control exercise where I focus on couple of high capacity treatment states and construct counterfactual group using the last treated group (Wisconsin and Montana). If the treatment effect is confounded with the flattening out of high capacity states, then the flattening pattern should emerge similarly for the treated states as it is for the last control group even before its removal of citizenship requirement. Appendix Figure A1 show the treatment state trend and their counterfactual control

group separately for each treated state. The evidence is mixed for different state. For Kansas and Nebraska, the flattening occurs around the same time with its counterfactual, but its post-treatment period still shows a faster growth in beds. This divergence is most substantial for West Virginia. However, for Wyoming, the beds show a faster decrease compared to the counterfactual. These evidence suggests the ceiling pattern exists among high capacity states, but it does not always cancel out the divergence post-reform.[29]

A.5 Mortality Data and Cause-of-Death Coding

A.5.1 Construction of the Older-Adult Mortality Rate

Mortality rates are built from the Vital Statistics Multiple Cause of Death files in two steps. The age-specific rate in each bin is deaths in that bin divided by the population at risk, where the denominator is the bin's population share multiplied by total population, both interpolated between census years from IPUMS and SEER.

The age-standardized rate is then the weighted sum of those age-specific rates, using age weights fixed at their base-year values, so that movements in the rate reflect changes in the likelihood of dying rather than changes in the age structure of the population. Rates are expressed per 10,000 residents. The rate for adults aged 65 and over is constructed the same way over the 65–74, 75–84, and 85-and-over bins, with the denominator restricted to the 65-and-over population, while the rate for adults aged 55–64 is a crude rate within that single ten-year band. Cause-specific rates change only the numerator. Hence, a cause-specific coefficient can be read directly as a share of the all-cause rate without rescaling.

A.5.2 Cause-of-Death Groups

Table A4 reports the code ranges defining each cause group, harmonized across the two revisions of the International Classification of Diseases that fall inside the sample period, in 1968 and 1979.

The infectious group comprises tuberculosis, syphilis, and the remaining infective and parasitic diseases, together with *pneumonia and influenza*; following Bailey and Goodman-Bacon (2015), these respiratory infections are grouped with infectious disease rather than with the respiratory chapter, since they were amenable to antibiotic treatment over this period.

The cardiovascular group covers rheumatic, hypertensive, and ischemic heart disease, other forms of heart disease, cerebrovascular disease, generalized arteriosclerosis, and other diseases of the arteries. Note

[29] Another intuitive way to test this is to control for the baseline beds capacity. However, this design requires a balanced beds capacity as control group for each treatment group. This is hardly feasible for early treated states, which are mostly low capacity states. When imposing baseline beds capacity as control, I find that the early treated cohorts reveal a 11 units treatment estimate in beds, driven by an over-extrapolated from the limited variation between the high-capacity and not-yet-treated states.

that the Seventh Revision files cerebrovascular disease under the nervous system chapter at codes 330–334 rather than under the circulatory chapter, and those codes are included explicitly for the pre-1968 years. Splitting cardiovascular disease by acute versus chronic incidents is infeasible because the Seventh Revision provides no distinct rubric for acute myocardial infarction while a separate acute category becomes available only with the Eighth Revision in 1968, after several treated states had already removed the citizenship requirement. Cancer comprises all malignant and other neoplasms. The external group comprises injury deaths, together with transport accidents, fire, poisoning, suicide, and homicide. I intentionally exclude deaths attributed to misadventures during medical or surgical care and adverse effects of drugs administered in therapeutic use, because these deaths are caused by contact with the medical system.

Table A4: Cause-of-Death Groups by ICD Revision

Cause group	ICD-7 (1959–1967)	ICDA-8 (1968–1978)	ICD-9 (1979–1980)
Infectious, incl. pneumonia & influenza	001–138, 480–483, 490–493	000–136, 470–486	001–139, 480–487
Cancer	140–239	140–239	140–239
Cardiovascular disease	330–334, 400–468	390–458	390–459
External causes[†]	800–999	800–999	800–999

Notes: The listed groups are not comprehensive. Together they account for approximately 87 percent of deaths at ages 65 and over.

A.5.3 Trends and the ICD Revisions

Figure A2 plots age-standardized mortality at ages 65 and over for all causes and for each cause group, with vertical lines marking the two ICD revisions. Cancer, cardiovascular disease, and external causes pass through both revisions without a visible step, and all-cause mortality tracks its long decline undisturbed.

The infectious group rises by roughly 19 percent at 1968, reflecting the reorganization of the respiratory chapter between the Seventh and Eighth Revisions together with the 1968 influenza pandemic. This movement is consistent with the corresponding series published by Bailey and Goodman-Bacon (2015), where it rises by 19.6 percent at the same revision against 19.2 percent here. If the change is consistent across states of different treatment timing, it will be absorbed by calendar-year fixed effects and leaves the estimates unaffected.

A.6 Cohort-Specific Trend Adjustment

This appendix details the trend-adjustment exercise summarized in Section 8. For each treatment cohort I use pre-treatment outcomes to estimate a linear trend and subtract the extrapolated trend from the outcome

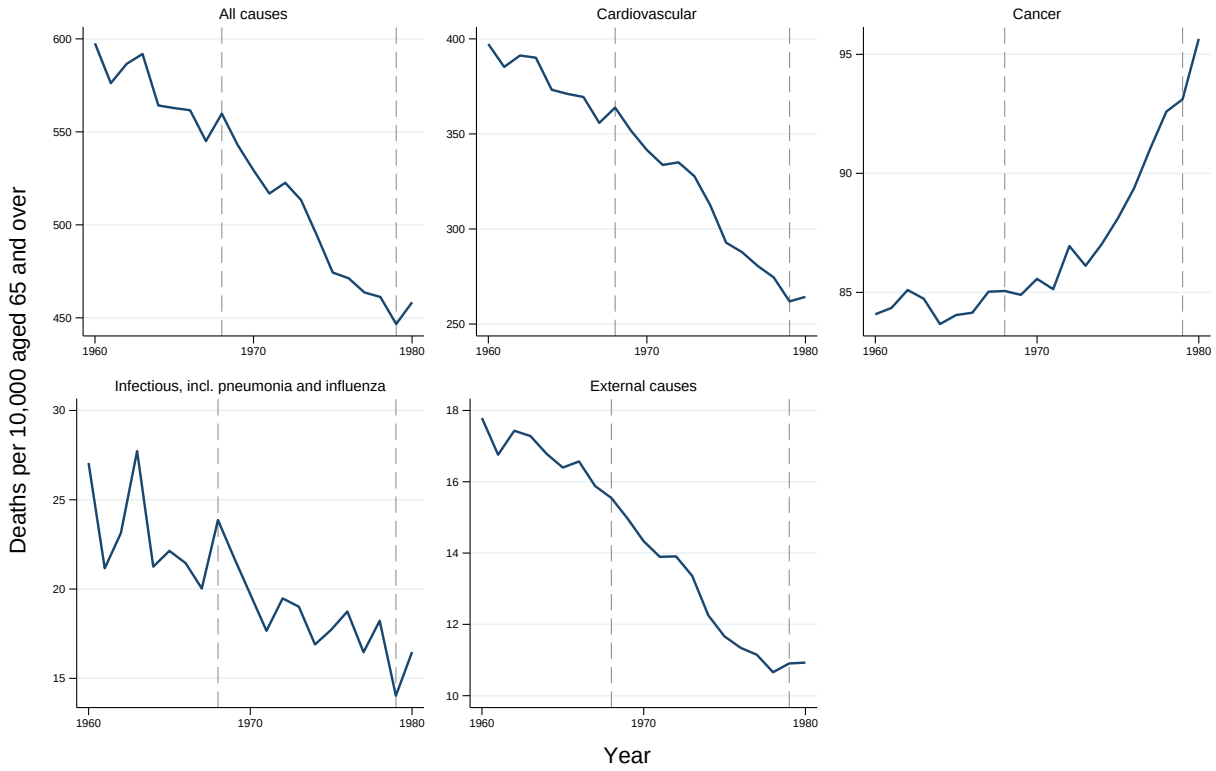


Figure A2: Age-Standardized Mortality by Cause, Ages 65 and Over

Notes: Population-weighted means across the 23 analysis states, per 10,000 residents aged 65 and over. Dashed vertical lines mark the 1968 and 1979 revisions of the International Classification of Diseases. Each panel uses its own vertical scale.

before estimation. The trend is fitted on the common 1964–1966 pre-period and weighted by baseline state population. Because the procedure requires an estimable pre-period, Michigan and North Dakota—which removed the citizenship requirement in 1965 and so are already treated at the 1966 AMD wave, leaving no post-1964 pre-treatment observation—drop out, so the adjusted and unadjusted estimates are not computed on identical samples. This procedure allows treatment cohorts to follow different underlying trajectories, but in doing so it does not impose a weaker identifying assumption than parallel trends; it imposes a different one, and neither nests the other. Where parallel trends requires that treated and comparison states would have moved together absent treatment, the trend adjustment requires that whatever differential movement is visible before treatment would have persisted, linearly, throughout the post-treatment window.

There is an economic reason to consider such differential trends when studying the stock of licensed FMGs. A longstanding immigration literature documents substantial persistence in the geographic distribution of immigrant inflows: new immigrants disproportionately locate in areas with established immigrant populations and co-ethnic networks (Bartel, 1989; Card, 2001).[30] Consequently, states with greater initial immigrant or FMG presence may have experienced systematically greater subsequent FMG inflows even in the absence of licensing reform. This concern may be particularly relevant following the 1965 Immigration and Nationality Act, when an expansion in national immigrant inflows could have generated differential local inflows according to states' pre-existing immigrant networks. In my setting, baseline FMG presence differs substantially across treatment cohorts and strongly predicts reform timing, making such differential underlying trajectories a potential concern.

Empirically, however, I find limited evidence that these differences in baseline FMG presence translated into substantially different pre-treatment trends. As shown in Appendix Figure A3, treatment cohorts differ considerably in their levels of FMG physician density, but their trajectories over the observable pre-treatment period are relatively similar. I therefore do not treat cohort-specific linear trends as my preferred counterfactual model. Instead, I report the de-trended estimates as a robustness exercise that allows for persistent cohort-specific trends under the stronger assumption that these trends would have continued after treatment.

The exercise is also informative about its own limits. Applied across all outcomes measured on the AMD wave clock, the adjustment absorbs pre-trends where one was present but introduces pre-trends where none was. It removes the significant pre-removal coefficients for the licensed FMG stock in total and among primary-care and hospital-based physicians. It leaves a significant pre-trend, of the opposite sign, for the stock pass rate. And for three outcomes whose unadjusted specifications show no pre-trend at all—average

[30] Related causal evidence shows that these networks facilitate immigrants' access to employment (Munshi, 2003), while recent work documents an important role for co-ethnic networks in the location decisions of highly skilled immigrant inventors (Campo et al., 2022).

Table A5: Baseline characteristics: states that removed vs. never had a citizenship requirement

	Removed req.	Never had req.	Difference
Population characteristics			
Population (log)	14.721	14.941	–0.220
Total income (log)	8.857	8.997	–0.140***
Share non-white	0.133	0.107	0.026
Share female	0.510	0.510	–0.001
Share foreign-born	0.028	0.055	–0.027**
Share aged 0–5	0.126	0.122	0.004
Share aged 65+	0.095	0.095	–0.000
Relevant policy			
Medicaid (adopted)	0.000	0.000	0.000
Endorsement (adopted)	0.391	0.400	–0.009
FLEX exam (adopted)	0.000	0.000	0.000
FMG physician supply (per 10k)			
FMGs recently licensed	0.123	0.221	–0.098*
% recently licensed among FMGs	0.115	0.162	–0.047
FMGs recently licensed, patient-care	0.094	0.155	–0.061*
FMGs recently graduated	0.097	0.190	–0.093
FMGs in training	0.302	0.636	–0.334
FMGs, total stock	0.828	1.882	–1.054*
Number of states	23	15	

Notes: Baseline (1964) state characteristics for states that ever removed a physician-licensure citizenship requirement (year_t_large non-missing) versus states that never imposed such a requirement (never-treated). Columns 1–2 report group means; column 3 reports the treated-minus-never-treated difference, with stars from a state-clustered regression of each characteristic on the treated indicator. */**/*** denote significance at the 10/5/1% levels. FMG supply measures are per 10,000 population.

experience of the licensed stock and the licensed share in each specialty group—the adjustment creates one significant at the one percent level. Average experience is the clearest case: the unadjusted pre-removal coefficient is -0.45 and statistically indistinguishable from zero, while the adjusted coefficient is 3.47 and significant at the one percent level, with the post-treatment estimate inflating by a factor of roughly two and a half as a direct consequence. The common explanation is the length of the window used to fit the trend. A slope estimated on three pre-period observations and extrapolated across a fifteen-year post-treatment horizon is not well disciplined, and where the underlying series is not close to linear the extrapolation overshoots. Appendix Figure A3 shows this directly for the quality outcomes: the not-yet-treated comparison pool declines over the pre-period without the adjustment and rises after it.

A.7 Additional Tables and Figures

Table A6: Effects on New FMG Arrival and FMG Trainees

	(1)	(2)	(3)
Panel A. FMGs not matched to the previous wave of AMD			
Year -10 to -4 (T_{-2})	0.003	0.003	0.003
	(0.028)	(0.028)	(0.028)
Year 0 to 3 (T_0)	0.034	0.041	0.059
	(0.035)	(0.035)	(0.036)
Year 4 to 9 (T_1)	0.191	0.191	0.191
	(0.134)	(0.171)	(0.134)
N	63	63	63
Panel B. FMG physicians in training			
Year -10 to -4 (T_{-2})	0.012	0.016	0.014
	(0.019)	(0.021)	(0.021)
Year 0 to 3 (T_0)	-0.009	-0.011	-0.013
	(0.041)	(0.042)	(0.036)
Year 2 to 9 (T_1)	0.088	0.091	0.089
	(0.076)	(0.126)	(0.083)
N	90	90	90
Controls	FMG Ratio	+Medicaid	+Medicare

Notes: The table reports difference-in-differences estimates using Callaway and Sant'Anna (2021). Panel A is FMGs not matched to the previous AMD wave (new arrivals); Panel B is FMG physicians in training. The control variable in column (1) is the ratio of FMGs among trainees in the baseline year. Column (2) additionally controls for Medicaid timing, entered as a binned indicator for Medicaid adoption after 1966. Column (3) additionally controls for the Medicare proxy (share of uninsured 65+). Wild-bootstrap standard errors clustered at the state level are reported in parentheses. T_{-1} is the reference period and is normalized to zero. */**/*** indicates significance at the 10/5/1% levels.

Table A7: Reclassification versus Expansion: Total FMGs, New Arrivals, and Trainees

	(1)	(2)	(3)
Panel A. FMG Physicians Present			
Pre-Removal (≤ -2 Yr)	–0.031	–0.031	–0.031
	(0.061)	(0.068)	(0.061)
Short-run (0 to 3 Yr)	–0.023	–0.036	–0.029
	(0.047)	(0.067)	(0.048)
Medium-run (2 to 9 Yr)	0.061	0.102	0.122
	(0.116)	(0.252)	(0.169)
Baseline Mean	1.10	1.10	1.10
N	90	90	90
Panel B. Newly Observed FMG Physicians			
Pre-Removal (≤ -2 Yr)	0.003	0.003	0.003
	(0.028)	(0.026)	(0.028)
Short-run (0 to 3 Yr)	0.034	0.041	0.059
	(0.033)	(0.035)	(0.039)
Medium-run (2 to 9 Yr)	0.191	0.191	0.191
	(0.174)	(0.187)	(0.174)
Baseline Mean (1966)	0.49	0.49	0.49
N	63	63	63
Panel C. FMG Trainees			
Pre-Removal (≤ -2 Yr)	–0.036*	–0.036*	–0.036*
	(0.021)	(0.021)	(0.021)
Short-run (0 to 3 Yr)	–0.009	–0.011	–0.013
	(0.041)	(0.045)	(0.036)
Medium-run (2 to 9 Yr)	0.088	0.091	0.089
	(0.068)	(0.124)	(0.071)
Baseline Mean	0.42	0.42	0.42
N	90	90	90
FMG trainee ratio	Yes	Yes	Yes
Medicaid adoption		Yes	
Medicare share			Yes

Notes: Outcomes are per 10,000 residents. Panel A is the total number of FMGs including unlicensed physicians; Panel B is newly observed FMGs, defined as those who cannot be matched to the previous AMD wave; Panel C is FMGs in training. Null effects across all three panels indicate that the increase in licensed FMGs reflects reclassification of physicians already present rather than new migration. Wild-bootstrap standard errors clustered at the state level are in parentheses. The baseline control is the 1964 ratio of FMGs among trainees; column (2) adds the year of Medicaid adoption and column (3) the share of uninsured residents aged 65 and over. A */**/*** indicates significance at the 10/5/1% levels.

Table A8: Licensed Physician Stock by Recency of Licensure: FMGs and USMGs

	Newly Licensed	Prior Cohort	Established
Panel A. FMG			
Pre-Removal (-5 to -2)	–0.014	–0.024*	0.095**
	(0.058)	(0.014)	(0.045)
Short-run (0 to 3 Yr)	0.202**	0.013	–0.028
	(0.093)	(0.017)	(0.040)
Honest CI ($\overline{M} = 0.25$)	[0.024, 0.377]	[–0.023, 0.061]	[–0.102, 0.059]
Honest CI ($\overline{M} = 1$)	[–0.019, 0.406]	[–0.059, 0.116]	[–0.135, 0.127]
Medium-run (2 to 9 Yr)	0.084	–0.080	–0.137**
	(0.055)	(0.063)	(0.053)
Honest CI ($\overline{M} = 0.25$)	[–0.048, 0.207]	[–0.264, 0.100]	[–0.270, –0.002]
Honest CI ($\overline{M} = 1$)	[–0.147, 0.292]	[–0.339, 0.161]	[–0.386, 0.119]
Baseline Mean	0.14	0.12	0.27
N	90	90	90
Panel B. USMG			
Pre-Removal (-5 to -2)	–0.105	–0.011	–0.035
	(0.093)	(0.057)	(0.084)
Short-run (0 to 3 Yr)	0.004	–0.022	0.168
	(0.075)	(0.027)	(0.102)
Honest CI ($\overline{M} = 0.25$)	[–0.178, 0.180]	[–0.085, 0.048]	[–0.060, 0.367]
Honest CI ($\overline{M} = 1$)	[–0.273, 0.264]	[–0.147, 0.136]	[–0.171, 0.431]
Medium-run (2 to 9 Yr)	0.268	–0.029	–0.111
	(0.261)	(0.030)	(0.171)
Honest CI ($\overline{M} = 0.25$)	[–0.293, 0.839]	[–0.116, 0.072]	[–0.519, 0.282]
Honest CI ($\overline{M} = 1$)	[–0.411, 1.003]	[–0.249, 0.255]	[–0.704, 0.442]
Baseline Mean	0.88	0.83	7.31
N	90	90	90
FMG trainee ratio	Yes	Yes	Yes

Notes: The licensed physician stock is decomposed by recency of licensure relative to the AMD wave dates. Newly licensed physicians were first licensed in the interval between the current wave and the previous one; prior-cohort physicians were licensed one interval earlier; established physicians were licensed before these two most recent intervals. Panel A reports FMGs and Panel B USMGs. Wild-bootstrap standard errors clustered at the state level are in parentheses. A */**/*** indicates significance at the 10/5/1% levels.

Table A9: Effects on the Stock of Licensed USMGs

	(1)	(2)	(3)
Pre-Removal (≤ -2 Yr)	–0.151*	–0.151*	–0.151*
	(0.080)	(0.081)	(0.080)
Short-run (0 to 3 Yr)	0.150	0.142	0.107
	(0.125)	(0.135)	(0.120)
Honest CI ($\overline{M} = 0.25$)	[–0.097, 0.416]	[–0.122, 0.427]	[–0.136, 0.375]
Honest CI ($\overline{M} = 1$)	[–0.198, 0.569]	[–0.224, 0.579]	[–0.244, 0.543]
Medium-run (2 to 9 Yr)	0.128	0.134	0.164
	(0.197)	(0.312)	(0.182)
Honest CI ($\overline{M} = 0.25$)	[–0.316, 0.555]	[–0.484, 0.752]	[–0.233, 0.553]
Honest CI ($\overline{M} = 1$)	[–0.603, 0.815]	[–0.726, 0.970]	[–0.549, 0.854]
Baseline Mean	9.03	9.03	9.03
FMG trainee ratio	Yes	Yes	Yes
Medicaid adoption		Yes	
Medicare share			Yes
N	90	90	90

Notes: The table reports Callaway and Sant'Anna (2021) doubly-robust estimates for the per-capita stock of licensed U.S. medical graduates, estimated at the state level and weighted by baseline state population. Estimates are binned in AMD wave time. Wild-bootstrap standard errors clustered at the state level are in parentheses. Honest confidence intervals follow Rambachan and Roth (2023). The baseline control is the 1964 ratio of FMGs among trainees; column (2) adds the year of Medicaid adoption and column (3) the share of uninsured residents aged 65 and over. A */**/*** indicates significance at the 10/5/1% levels.

Table A10: Effects on the Share of Physicians Licensed, by Specialty Group

	FMG			USMG		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Primary Care						
Pre-Removal (-5 to -2)	–5.330	–5.330	–5.330	–0.322	–0.322	–0.322
	(3.573)	(3.490)	(3.573)	(0.575)	(0.657)	(0.575)
Short-run (0 to 3 Yr)	11.021***	10.204**	9.772**	0.392	0.546	0.607
	(4.154)	(5.012)	(4.565)	(0.564)	(0.623)	(0.601)
Honest CI ($\overline{M} = 0.25$)	[1.033, 20.155]	[–1.314, 20.542]	[–1.077, 19.879]	[–0.833, 1.640]	[–0.728, 1.846]	[–0.645, 1.858]
Honest CI ($\overline{M} = 1$)	[–5.477, 25.201]	[–9.282, 25.154]	[–8.711, 25.352]	[–1.259, 1.970]	[–1.110, 2.254]	[–1.108, 2.222]
Medium-run (2 to 9 Yr)	7.097	6.662	7.185	2.585***	2.352***	2.190**
	(11.100)	(11.065)	(10.743)	(0.726)	(0.739)	(0.935)
Honest CI ($\overline{M} = 0.25$)	[–18.136, 31.849]	[–19.859, 31.237]	[–18.476, 31.407]	[1.016, 4.064]	[0.796, 3.970]	[0.430, 3.951]
Honest CI ($\overline{M} = 1$)	[–27.366, 39.576]	[–32.494, 39.250]	[–30.250, 40.157]	[0.035, 5.134]	[–0.232, 4.809]	[–0.461, 4.913]
Baseline Mean		52.2			96.3	
N		90			90	
Panel B. Hospital-Based (Platform)						
Pre-Removal (-5 to -2)	–8.183*	–8.183*	–8.183*	–1.393	–1.393	–1.393
	(4.743)	(4.501)	(4.743)	(1.143)	(1.319)	(1.143)
Short-run (0 to 3 Yr)	15.026**	13.837*	14.032*	–1.005	–0.997	–0.856
	(6.585)	(7.331)	(7.205)	(0.724)	(0.729)	(0.619)
Honest CI ($\overline{M} = 0.25$)	[0.225, 31.851]	[–1.987, 31.620]	[–1.136, 31.934]	[–2.774, 0.736]	[–2.789, 0.796]	[–2.468, 0.913]
Honest CI ($\overline{M} = 1$)	[–9.766, 50.159]	[–11.726, 49.332]	[–9.709, 51.586]	[–4.393, 2.353]	[–4.445, 2.545]	[–4.361, 2.995]
Medium-run (2 to 9 Yr)	12.704	13.729*	15.225**	–0.163	–0.141	–0.120
	(8.301)	(8.051)	(7.333)	(1.775)	(1.876)	(1.776)
Honest CI ($\overline{M} = 0.25$)	[–9.076, 34.851]	[–6.904, 34.033]	[–3.972, 35.595]	[–4.715, 4.476]	[–5.319, 5.135]	[–5.048, 4.898]
Honest CI ($\overline{M} = 1$)	[–42.141, 68.721]	[–41.368, 68.118]	[–41.934, 74.638]	[–7.371, 7.227]	[–8.000, 7.922]	[–8.493, 8.252]
Baseline Mean		48.5			93.2	
N		90			90	
Panel C. Surgeons						
Pre-Removal (-5 to -2)	–3.299	–3.299	–3.299	–1.102*	–1.102	–1.102*
	(5.648)	(4.818)	(5.648)	(0.662)	(0.713)	(0.662)
Short-run (0 to 3 Yr)	16.186***	16.590***	15.489**	–0.272	–0.304	–0.487
	(5.922)	(5.250)	(6.217)	(0.526)	(0.478)	(0.405)
Honest CI ($\overline{M} = 0.25$)	[3.701, 29.333]	[4.430, 29.395]	[2.471, 28.968]	[–1.432, 0.952]	[–1.353, 0.862]	[–1.429, 0.591]
Honest CI ($\overline{M} = 1$)	[–3.644, 39.301]	[–2.749, 39.361]	[–5.201, 38.873]	[–2.331, 1.921]	[–2.251, 1.911]	[–2.297, 1.773]
Medium-run (2 to 9 Yr)	–16.118	–16.384	–14.731	0.529	0.316	0.211
	(25.167)	(19.682)	(24.699)	(0.751)	(0.947)	(0.697)
Honest CI ($\overline{M} = 0.25$)	[–65.328, 32.137]	[–66.224, 32.488]	[–64.502, 34.074]	[–1.549, 2.429]	[–1.807, 2.516]	[–1.625, 2.147]
Honest CI ($\overline{M} = 1$)	[–77.552, 44.334]	[–78.574, 43.815]	[–77.934, 45.486]	[–3.761, 4.369]	[–3.649, 4.765]	[–3.744, 4.412]
Baseline Mean		44.7			94.5	
N		90			90	
FMG trainee ratio	Yes	Yes	Yes	Yes	Yes	Yes
Medicaid adoption		Yes			Yes	
Medicare share			Yes			Yes

Notes: The table reports Callaway and Sant'Anna (2021) doubly-robust estimates for the share of physicians holding a state licence, in percentage points, estimated at the state level and weighted by baseline state population. Panels A–C are primary care, hospital-based, and surgical specialties. Columns (1)–(3) report foreign medical graduates and columns (4)–(6) U.S. medical graduates. Within each block the first column is the baseline specification controlling for the 1964 ratio of FMGs among trainees, the second adds the year of Medicaid adoption, and the third adds the share of uninsured residents aged 65 and over. Wild-bootstrap standard errors clustered at the state level are in parentheses. Honest confidence intervals follow Rambachan and Roth (2023), allowing the maximal post-treatment violation of parallel trends to be $\overline{M}$ times the maximal pre-treatment violation. A */**/*** indicates significance at the 10/5/1% levels.

Table A11: Effects on the Licensed Physician Stock by Specialty Group

	FMG			USMG		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Primary Care						
Pre-Removal (-5 to -2)	0.043***	0.043***	0.043***	–0.164***	–0.164***	–0.164***
	(0.015)	(0.015)	(0.015)	(0.044)	(0.051)	(0.044)
Short-run (0 to 3 Yr)	0.082*	0.076*	0.076*	0.190**	0.195**	0.188***
	(0.042)	(0.045)	(0.042)	(0.074)	(0.077)	(0.071)
Honest CI ($\overline{M} = 0.25$)	[0.003, 0.163]	[–0.013, 0.166]	[–0.005, 0.160]	[0.031, 0.363]	[0.030, 0.375]	[0.027, 0.363]
Honest CI ($\overline{M} = 1$)	[–0.021, 0.194]	[–0.037, 0.192]	[–0.025, 0.187]	[–0.059, 0.493]	[–0.070, 0.508]	[–0.084, 0.506]
Medium-run (2 to 9 Yr)	–0.110	–0.103	–0.098	0.200	0.210	0.221*
	(0.116)	(0.095)	(0.105)	(0.135)	(0.180)	(0.118)
Honest CI ($\overline{M} = 0.25$)	[–0.338, 0.118]	[–0.348, 0.142]	[–0.333, 0.141]	[–0.116, 0.500]	[–0.157, 0.576]	[–0.064, 0.506]
Honest CI ($\overline{M} = 1$)	[–0.375, 0.160]	[–0.382, 0.181]	[–0.371, 0.179]	[–0.355, 0.709]	[–0.386, 0.798]	[–0.338, 0.770]
Baseline Mean		0.27			5.22	
N		90			90	
Panel B. Hospital-Based (Platform)						
Pre-Removal (-5 to -2)	0.017***	0.017***	0.017***	0.006	0.006	0.006
	(0.006)	(0.006)	(0.006)	(0.023)	(0.025)	(0.023)
Short-run (0 to 3 Yr)	0.025**	0.023**	0.023*	0.009	0.007	0.004
	(0.012)	(0.012)	(0.012)	(0.035)	(0.030)	(0.032)
Honest CI ($\overline{M} = 0.25$)	[0.003, 0.049]	[0.000, 0.048]	[–0.000, 0.047]	[–0.059, 0.072]	[–0.056, 0.065]	[–0.063, 0.064]
Honest CI ($\overline{M} = 1$)	[–0.003, 0.059]	[–0.004, 0.059]	[–0.005, 0.058]	[–0.089, 0.083]	[–0.084, 0.080]	[–0.094, 0.082]
Medium-run (2 to 9 Yr)	0.008	0.008	0.009	0.089	0.093	0.103
	(0.015)	(0.020)	(0.014)	(0.099)	(0.112)	(0.093)
Honest CI ($\overline{M} = 0.25$)	[–0.023, 0.039]	[–0.026, 0.043]	[–0.022, 0.040]	[–0.133, 0.306]	[–0.141, 0.322]	[–0.107, 0.313]
Honest CI ($\overline{M} = 1$)	[–0.037, 0.056]	[–0.042, 0.059]	[–0.039, 0.058]	[–0.162, 0.326]	[–0.168, 0.348]	[–0.141, 0.346]
Baseline Mean		0.08			0.79	
N		90			90	
Panel C. Surgeons						
Pre-Removal (-5 to -2)	0.008	0.008	0.008	–0.105***	–0.105***	–0.105***
	(0.009)	(0.010)	(0.009)	(0.024)	(0.024)	(0.024)
Short-run (0 to 3 Yr)	0.044**	0.042*	0.043*	0.018	0.017	0.010
	(0.021)	(0.022)	(0.022)	(0.029)	(0.028)	(0.027)
Honest CI ($\overline{M} = 0.25$)	[0.004, 0.084]	[–0.018, 0.100]	[0.002, 0.084]	[–0.059, 0.091]	[–0.061, 0.091]	[–0.065, 0.082]
Honest CI ($\overline{M} = 1$)	[–0.001, 0.088]	[–0.008, 0.090]	[–0.002, 0.087]	[–0.135, 0.159]	[–0.139, 0.161]	[–0.151, 0.188]
Medium-run (2 to 9 Yr)	–0.053	–0.050	–0.050	–0.027	–0.025	–0.016
	(0.052)	(0.051)	(0.051)	(0.086)	(0.108)	(0.077)
Honest CI ($\overline{M} = 0.25$)	[–0.154, 0.048]	[–0.162, 0.062]	[–0.153, 0.052]	[–0.238, 0.177]	[–0.285, 0.236]	[–0.214, 0.182]
Honest CI ($\overline{M} = 1$)	[–0.159, 0.055]	[–0.170, 0.068]	[–0.160, 0.057]	[–0.370, 0.301]	[–0.414, 0.350]	[–0.363, 0.328]
Baseline Mean		0.11			2.39	
N		90			90	
FMG trainee ratio	Yes	Yes	Yes	Yes	Yes	Yes
Medicaid adoption		Yes			Yes	
Medicare share			Yes			Yes

Notes: The table reports average treatment effects on the per-capita stock of licensed physicians by specialty group, separately for FMGs (Panels A–C) and USMGs (Panels D–F). Specialty groups are primary care, hospital-based, and surgery. Wild-bootstrap standard errors clustered at the state level are in parentheses. Honest confidence intervals follow Rambachan and Roth (2023). The baseline control is the 1964 ratio of FMGs among trainees; column (2) adds the year of Medicaid adoption and column (3) the share of uninsured residents aged 65 and over. A */**/*** indicates significance at the 10/5/1% levels.

Table A12: Effects on All-Cause Mortality by Age Group

	(1)	(2)	(3)
	Baseline	+Control Medicaid	+Control Medicare
Panel A. All-Cause Mortality 55–64 per 10,000			
Pre-period (-5 to -1)	–0.456	–2.330	–1.464
	(1.122)	(2.642)	(1.304)
Short-run (0 to 3 Yr)	–0.291	–1.576	–0.924
	(1.034)	(2.060)	(0.788)
Medium-run (4 to 6 Yr)	–3.167*	–0.467	–2.943*
	(1.792)	(1.689)	(1.675)
Overall ATT	–1.096	–1.379	–1.547*
	(1.085)	(1.679)	(0.922)
Baseline Mean	169.8	169.8	169.8
N	340	340	340
Panel B. All-Cause Mortality 65+ per 10,000			
Pre-period (-5 to -1)	–1.170	–3.679	–0.103
	(3.617)	(7.663)	(4.955)
Short-run (0 to 3 Yr)	–0.287	–1.641	1.792
	(2.672)	(3.315)	(2.978)
Medium-run (4 to 6 Yr)	0.381	7.824	1.107
	(3.574)	(6.440)	(3.270)
Overall ATT	–0.099	0.013	1.580
	(2.500)	(3.740)	(2.583)
Baseline Mean	565.1	565.1	565.1
N	340	340	340
Panel C. Infant Mortality (all causes) per 1,000 births			
Pre-period (-5 to -1)	–0.721	–0.294	–1.771
	(1.209)	(1.829)	(1.328)
Short-run (0 to 3 Yr)	0.347	0.405	0.256
	(0.256)	(0.369)	(0.288)
Medium-run (4 to 6 Yr)	–1.244**	–0.098	–1.353***
	(0.547)	(1.282)	(0.485)
Overall ATT	–0.084	0.319	–0.232
	(0.285)	(0.433)	(0.285)
Baseline Mean	26.2	26.2	26.2
N	294	294	294

Notes: Panel A reports all-cause mortality for adults aged 55–64 (crude, per 10,000); Panel B age-standardized all-cause mortality for adults aged 65 and over (per 10,000); Panel C infant mortality from all causes (per 1,000 live births). Column (1) is the baseline doubly-robust specification controlling for the 1964 ratio of FMGs among trainees; column (2) adds the year of Medicaid adoption; column (3) adds the share of uninsured residents aged 65 and over. Wild-bootstrap standard errors clustered at the state level are in parentheses. A */**/*** indicates significance at the 10/5/1% levels.

Table A13: Effects on Mortality by Cause of Death

	Ages 65+ (age-standardized)			Ages 55–64 (crude)		
	(1) Baseline	(2) +Mcaid	(3) +Mcare	(1) Baseline	(2) +Mcaid	(3) +Mcare
Panel A. Infectious disease (per 10,000)						
Pre-period (-5 to -1)	–0.01	–2.05	–0.72	0.00	0.06	–0.37
	(0.81)	(1.75)	(0.99)	(0.27)	(0.43)	(0.25)
Short-run (0 to 3 Yr)	0.31	0.51	0.64	–0.47	0.31	–0.40
	(0.50)	(0.99)	(0.53)	(0.44)	(0.41)	(0.46)
Medium-run (4 to 6 Yr)	–2.63***	–0.42	–2.68***	–1.81***	–0.58	–1.79***
	(0.81)	(1.25)	(0.79)	(0.64)	(0.52)	(0.66)
Overall ATT	–0.52	0.35	–0.38	–0.85*	0.15	–0.83*
	(0.49)	(0.92)	(0.49)	(0.47)	(0.36)	(0.47)
Baseline Mean		21.2			4.9	
Panel B. Cardiovascular disease (per 10,000)						
Pre-period (-5 to -1)	0.9	1.4	1.4	0.3	–0.1	1.1
	(2.8)	(4.3)	(2.8)	(1.1)	(1.4)	(1.0)
Short-run (0 to 3 Yr)	–1.6	–0.3	–0.5	0.3	–0.1	0.4
	(1.9)	(2.5)	(2.1)	(0.8)	(1.7)	(0.9)
Medium-run (4 to 6 Yr)	–2.7	2.9	–2.1	0.3	–0.1	0.3
	(2.8)	(5.6)	(2.7)	(1.5)	(0.7)	(1.5)
Overall ATT	–1.9	0.3	–1.0	0.3	–0.1	0.4
	(1.9)	(2.6)	(2.0)	(0.9)	(1.6)	(1.0)
Baseline Mean		370.7			90.0	
Panel C. Cancer (per 10,000)						
Pre-period (-5 to -1)	–1.24	–1.25	–0.26	–0.04	–1.17	–0.88
	(0.77)	(0.87)	(0.96)	(0.65)	(0.80)	(0.63)
Short-run (0 to 3 Yr)	1.34	1.09	1.58*	–0.05	–1.24*	–0.65
	(0.91)	(1.52)	(0.94)	(0.54)	(0.68)	(0.54)
Medium-run (4 to 6 Yr)	2.84*	4.04	2.93**	–1.14	–0.74	–1.18
	(1.46)	(3.03)	(1.43)	(0.78)	(1.54)	(0.76)
Overall ATT	1.76**	1.61	2.00***	–0.36	–1.15	–0.82
	(0.79)	(1.41)	(0.73)	(0.50)	(0.70)	(0.51)
Baseline Mean		83.8			38.8	
Panel D. External causes (per 10,000)						
Pre-period (-5 to -1)	–0.52	–0.96	–0.98	–0.02	–0.27	–0.14
	(0.74)	(0.89)	(0.76)	(0.23)	(0.44)	(0.22)
Short-run (0 to 3 Yr)	0.02	–0.34	–0.20	–0.12	–0.29	–0.16
	(0.44)	(0.63)	(0.52)	(0.30)	(0.43)	(0.39)
Medium-run (4 to 6 Yr)	0.32	–0.15	0.26	0.43	0.45	0.44
	(0.52)	(0.69)	(0.60)	(0.42)	(0.34)	(0.41)
Overall ATT	0.10	–0.30	–0.06	0.04	–0.16	0.03
	(0.35)	(0.52)	(0.46)	(0.30)	(0.42)	(0.35)
Baseline Mean		16.7			9.1	
N			340			
FMG trainee ratio			Yes			

Notes: Columns (1)–(3) report age-standardized mortality for adults aged 65 and over; columns (1)–(3) of the right block report crude mortality for adults aged 55–64. Both are per 10,000 residents. Panels correspond to the cause groups defined in Appendix Table A4. Within each block, the first column is the baseline specification controlling for the 1964 ratio of FMGs among trainees, the second adds the year of Medicaid adoption, and the third adds the share of uninsured residents aged 65 and over. Wild-bootstrap standard errors clustered at the state level are in parentheses. A */**/*** indicates significance at the 10/5/1% levels.

Table A14: County Heterogeneity in the Licensed Physician Stock

	FMG licensing stock			USMG licensing stock		
	Low	High	Diff.	Low	High	Diff.
Panel A. Split by Baseline median income						
Pre-period	–0.66	–0.26	–0.40***	0.98	–0.13	1.11***
	(0.10)	(0.10)	[0.001]	(0.21)	(0.11)	[0.000]
Short-run (0 to 3 Yr)	0.64	0.30	0.35*	–0.76	0.84	–1.60***
	(0.13)	(0.16)	[0.062]	(0.17)	(0.34)	[0.001]
N counties	221	222		221	222	
N observations	1105	1110		1105	1110	
Panel B. Split by Baseline physician density						
Pre-period	0.05	–0.45	0.51***	0.05	–0.18	0.23
	(0.04)	(0.13)	[0.002]	(0.10)	(0.08)	[0.134]
Short-run (0 to 3 Yr)	0.01	0.28	–0.27	–0.06	1.17	–1.22**
	(0.04)	(0.15)	[0.114]	(0.07)	(0.40)	[0.011]
N counties	221	222		221	222	
N observations	1105	1110		1105	1110	
Panel C. Split by Baseline urban share						
Pre-period	–0.01	–0.36	0.34**	0.05	–0.25	0.31**
	(0.04)	(0.12)	[0.012]	(0.10)	(0.09)	[0.033]
Short-run (0 to 3 Yr)	0.05	0.28	–0.22	–0.01	1.17	–1.18**
	(0.04)	(0.15)	[0.178]	(0.08)	(0.41)	[0.015]
N counties	221	222		221	222	
N observations	1105	1110		1105	1110	
Panel D. Split by baseline Medicaid exposure						
	High	Low	Diff.	High	Low	Diff.
Pre-period	–0.39	–0.31	–0.08	0.84	–0.09	0.93***
	(0.14)	(0.17)	[0.653]	(0.18)	(0.16)	[0.001]
Short-run (0 to 3 Yr)	0.34	0.24	0.10	–0.44	1.22	–1.66***
	(0.15)	(0.15)	[0.472]	(0.20)	(0.44)	[0.005]
N counties	172	271		172	271	
N observations	860	1355		860	1355	
FMG trainee ratio			Yes			
Baseline nonwhite population ratio			Yes			

Notes: Estimates are obtained by applying the Callaway and Sant'Anna (2021) estimator separately within each predetermined county partition. Columns (1)–(3) report the FMG licensed stock for low- and high-group counties and their difference; columns (4)–(6) report the same for the USMG licensed stock. Standard errors clustered at the state level are in parentheses; p-values from a Wald test of equality across partitions, using a state-clustered covariance matrix that preserves covariance across subgroups, are in brackets. Panels split counties at the 1964 median of household income, physician density, urban share, and AFDC caseload rate. Arkansas is omitted because county-level characteristics are not available in the 1973 AMD year. A */**/*** indicates significance at the 10/5/1% levels.

Table A15: Decomposition of the Aggregate Hospital Effect Excluding Hill-Burton-Active Counties

Cohort	Full sample		Excl. HB-active		Retained	Expected contrib.
	Weight (1)	Contrib. (2)	Weight (3)	Contrib. (4)	pop. share (5)	(pop.-shrunk) (6) = (2)×(5)
$g1965$	0.131	−0.020	0.131	−0.004	0.448	−0.009
$g1968$	0.000	0.000	0.000	0.000	0.103	0.000
$g1969$	0.215	0.649	0.215	0.130	0.498	0.323
$g1970$	0.489	0.902	0.489	0.618	0.474	0.428
$g1972$	0.082	0.116	0.082	0.006	0.621	0.072
$g1973$	0.083	0.249	0.083	0.110	0.460	0.115
Total (= pooled ATT)	1.000	1.898	1.000	0.859		0.930
Estimation N	346		337			

Notes: The aggregate effect is decomposed into cohort-level contributions. Columns (1)–(2) report each adoption cohort's aggregation weight and its contribution to the pooled ATT in the full sample; columns (3)–(4) report the corresponding values when counties with Hill-Burton funding receipt during 1964–1971 are excluded (Chung et al., 2017). Column (5) is the share of each cohort's baseline population retained under that exclusion, and column (6) scales the full-sample contribution by that share, giving the contribution expected from population loss alone. Comparing columns (4) and (6) separates attenuation caused by losing identifying population from attenuation caused by a change in the underlying cohort effects. Panel A reports short-term general hospital beds and Panel B short-term general admissions.

Table A16: Estimates With and Without Cohort-Specific Trend Adjustment

	Licensed FMG stock (per 10,000)					FMG quality (percentage points)	
	Total	Primary care	Platform	Surgeons	Established	Flow pass rate	Stock pass rate
Panel A. Baseline specification							
Pre-Removal (≤ -2)	0.057**	0.043***	0.017***	0.008	0.095**	1.09	–2.22**
	(0.024)	(0.015)	(0.006)	(0.009)	(0.045)	(4.68)	(0.94)
Short-run (0 to 3 Yr)	0.187**	0.082*	0.025**	0.044**	–0.028	4.37	2.63***
	(0.094)	(0.042)	(0.012)	(0.021)	(0.040)	(6.62)	(0.89)
Honest CI ($\overline{M} = 0.25$)	[0.009, 0.361]	[0.003, 0.163]	[0.003, 0.049]	[0.004, 0.084]	[–0.102, 0.059]	[–10.42, 27.65]	[0.43, 4.75]
Honest CI ($\overline{M} = 1$)	[–0.019, 0.379]	[–0.021, 0.194]	[–0.003, 0.059]	[–0.001, 0.088]	[–0.135, 0.127]	[–44.31, 66.10]	[–1.59, 6.56]
Medium-run (2 to 9 Yr)	–0.133	–0.110	0.008	–0.053	–0.137**	–1.25	3.97***
	(0.155)	(0.116)	(0.015)	(0.052)	(0.053)	(3.63)	(0.96)
Honest CI ($\overline{M} = 0.25$)	[–0.434, 0.168]	[–0.338, 0.118]	[–0.023, 0.039]	[–0.154, 0.048]	[–0.270, –0.002]	[–31.56, 35.10]	[–0.79, 4.77]
Honest CI ($\overline{M} = 1$)	[–0.482, 0.210]	[–0.375, 0.160]	[–0.037, 0.056]	[–0.159, 0.055]	[–0.386, 0.119]	[–108.89, 110.24]	[–5.44, 9.10]
Panel B. Trend-adjusted							
Pre-Removal (≤ -2)	0.028	–0.024	–0.010	0.004	–0.011	1.87	2.77***
	(0.025)	(0.016)	(0.006)	(0.009)	(0.049)	(7.59)	(0.90)
Short-run (0 to 3 Yr)	0.226**	0.134***	0.043***	0.046*	0.008	5.59	–0.02
	(0.097)	(0.045)	(0.013)	(0.024)	(0.048)	(5.86)	(0.81)
Honest CI ($\overline{M} = 0.25$)	[0.029, 0.420]	[0.042, 0.230]	[0.014, 0.071]	[–0.003, 0.095]	[–0.084, 0.094]	[–10.99, 29.37]	[–1.85, 1.93]
Honest CI ($\overline{M} = 1$)	[0.003, 0.430]	[0.030, 0.253]	[0.006, 0.077]	[–0.024, 0.111]	[–0.119, 0.118]	[–52.13, 72.46]	[–3.48, 3.86]
Medium-run (2 to 9 Yr)	–0.135	–0.024	0.058**	–0.101	–0.292***	2.40	–4.61***
	(0.249)	(0.170)	(0.023)	(0.085)	(0.088)	(5.24)	(1.29)
Honest CI ($\overline{M} = 0.25$)	[–0.551, 0.274]	[–0.313, 0.270]	[0.015, 0.100]	[–0.243, 0.041]	[–0.438, –0.148]	[–30.00, 45.52]	[–8.18, –1.05]
Honest CI ($\overline{M} = 1$)	[–0.589, 0.303]	[–0.333, 0.297]	[0.005, 0.110]	[–0.252, 0.048]	[–0.483, –0.112]	[–110.05, 124.91]	[–11.79, 2.56]

Notes: Panel A reproduces the baseline Callaway and Sant'Anna (2021) estimates reported in the main text; Panel B re-estimates the same specifications after subtracting each treatment cohort's own pre-period linear trend from the outcome. The trend is estimated on the common 1964–1966 pre-period and weighted by baseline state population. Columns (1)–(4) report the per-capita stock of licensed FMGs in total and by specialty group; column (5) the stock of established licensed FMGs, those licensed before the two most recent inter-wave intervals; column (6) reports the licensing-exam pass rate of newly licensed FMGs (the entry margin) and column (7) the pass rate of the licensed FMG stock. All columns use the baseline specification controlling for the 1964 ratio of FMGs among trainees. Because the trend adjustment requires an estimable 1964–1966 pre-period, Michigan and North Dakota—which removed the requirement in 1965 and are therefore already treated at the 1966 AMD wave—are dropped from Panel B; the two panels are therefore not estimated on identical samples. Honest confidence intervals follow Rambachan and Roth (2023), allowing the maximal post-treatment violation of parallel trends to be $\overline{M}$ times the maximal pre-treatment violation. Wild-bootstrap standard errors clustered at the state level are in parentheses. A */**/*** indicates significance at the 10/5/1% levels.

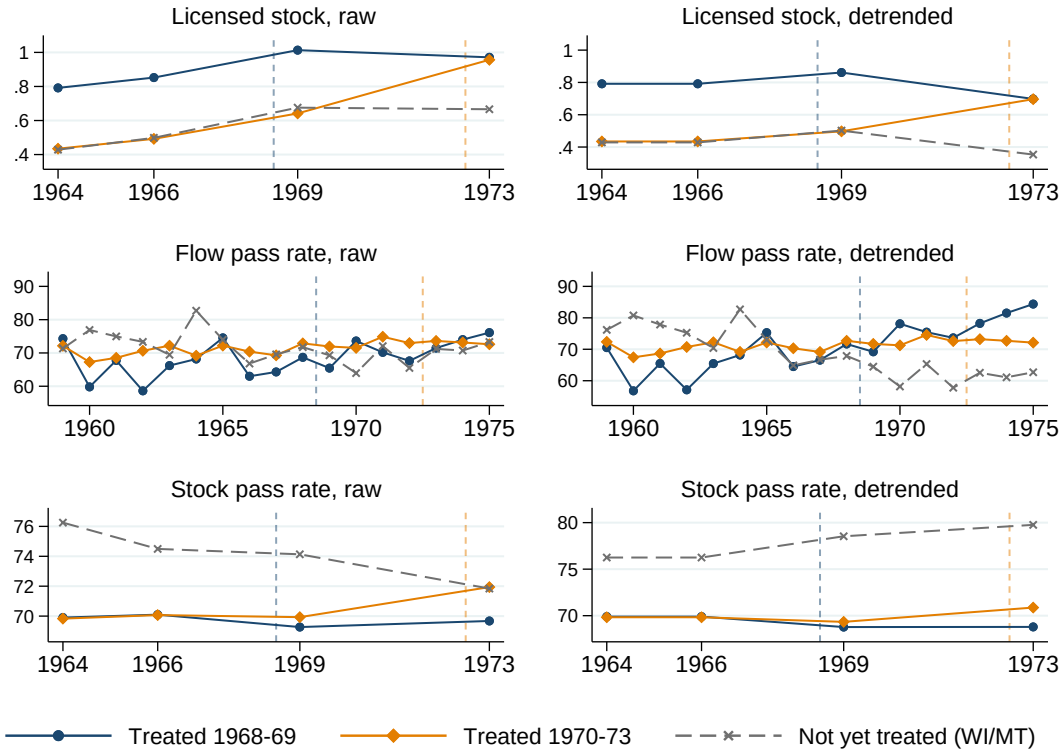


Figure A3: Group Means With and Without Cohort-Specific Trend Adjustment

Notes: Population-weighted group means for three outcomes, shown without the trend adjustment (left column) and after it (right column). Rows are the per-capita stock of licensed FMGs, the licensing-exam pass rate of newly licensed FMGs (the entry margin), and the pass rate of the licensed FMG stock. States are grouped by year of citizenship-requirement removal into an early cohort (1968–69) and a later cohort (1970–73), with Wisconsin and Montana, removed after 1975, shown as the not-yet-treated comparison pool; vertical dashed lines mark the two removal dates. The top-left panel is the basis for treating cohort-specific linear trends as a robustness exercise rather than as the preferred counterfactual model: cohorts differ substantially in the level of FMG density but follow broadly similar pre-treatment paths. The bottom row shows why the adjustment is not innocuous for the quality outcomes—the comparison pool declines over the pre-period without adjustment and rises after it, which is the source of the pre-trend sign reversal discussed in Section 8. Corresponding estimates appear in Appendix Table A16.

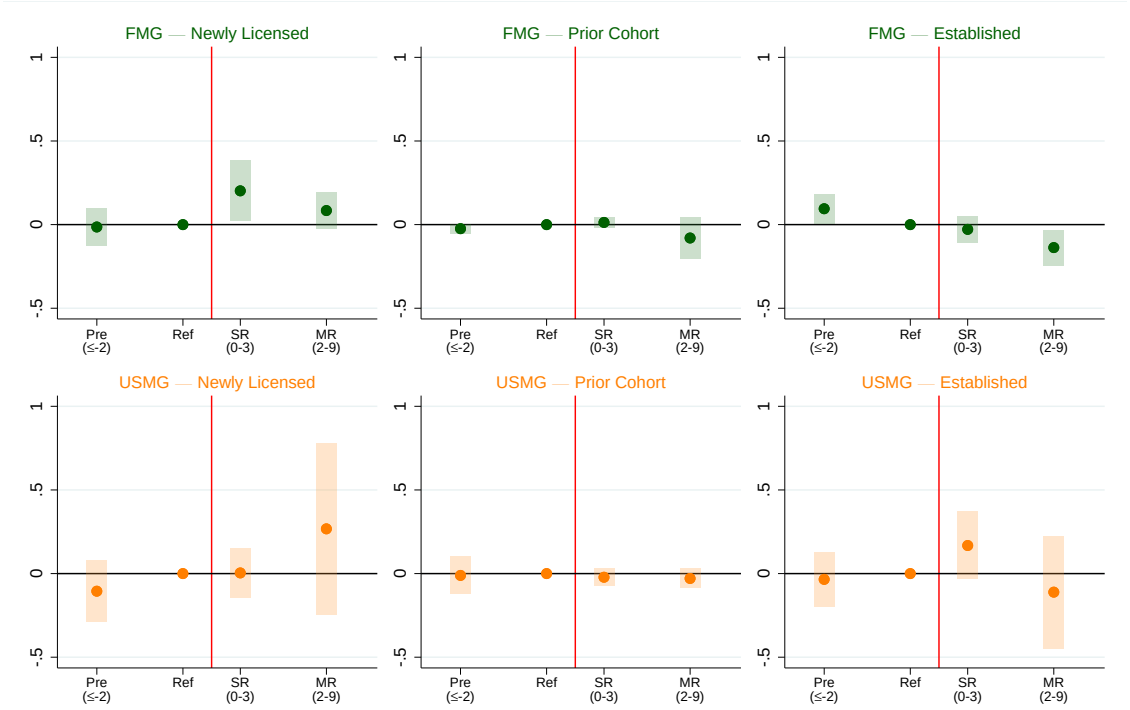


Figure A4: Licensed FMG Stock by Recency of Licensure

Notes: Event-study estimates for the licensed FMG stock split by recency of licensure relative to the AMD wave dates. Newly licensed physicians were first licensed in the interval between the current and previous wave; prior-cohort physicians one interval earlier; established physicians before those two intervals. Corresponding estimates appear in Table A8.

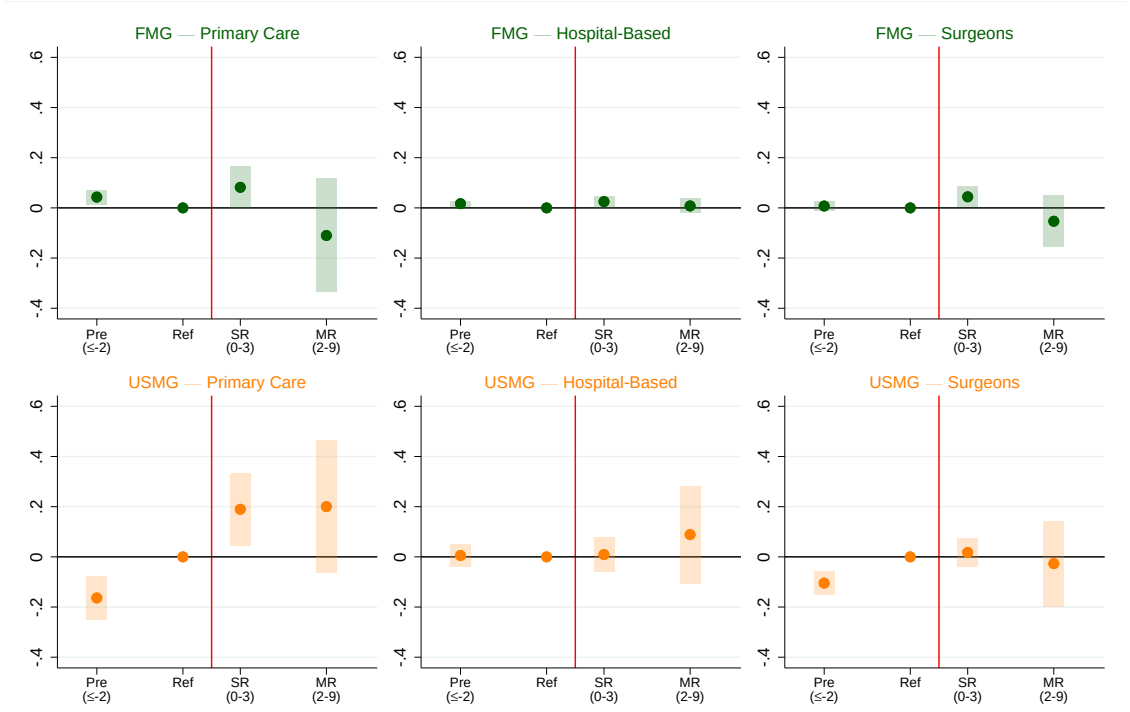


Figure A5: Licensed Physician Stock by Specialty Group: FMGs and USMGs

Notes: Event-study estimates for the licensed physician stock by specialty group (primary care, hospital-based, and surgery), estimated separately for FMGs and USMGs. Rambachan–Roth sensitivity appears in Appendix Figure A22.

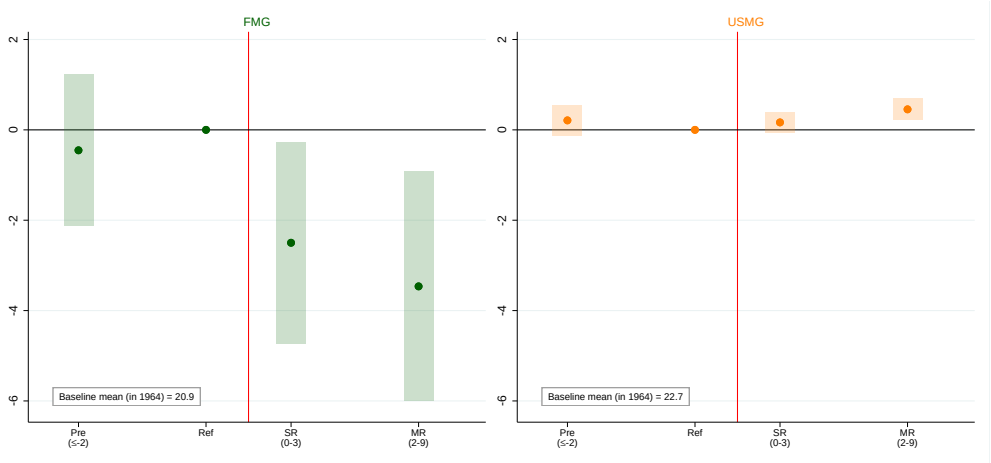


Figure A6: Average Experience of Licensed Physicians: FMGs and USMGs

Notes: Experience is measured as years since graduation from medical school. The left panel reports licensed FMGs and the right panel licensed USMGs.

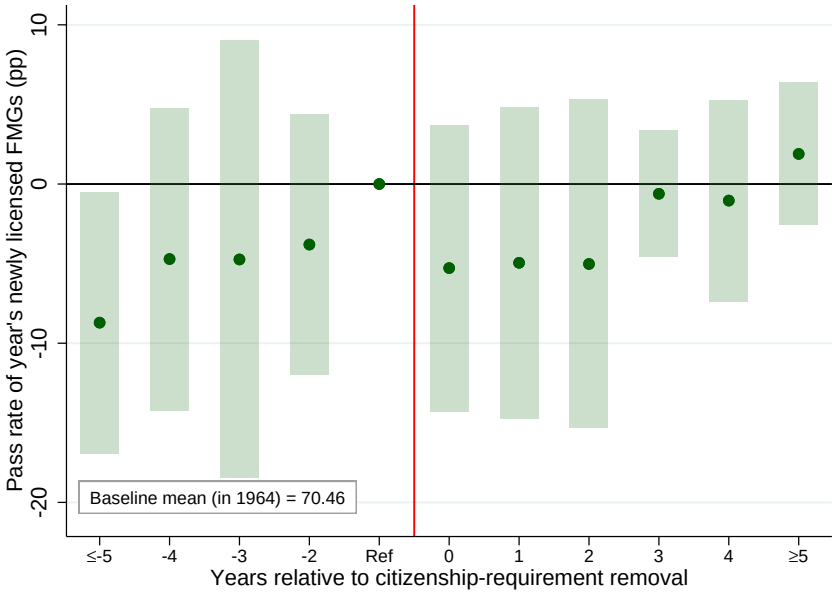


Figure A7: Selection into Licensure: Average Exam Pass Rate of New FMG Licensees

Notes: FMG quality is proxied by the average licensing-exam pass rate constructed from the medical school each licensee graduated from. The outcome is measured over new licensees in the calendar year, so the coefficient tests whether citizenship removal changed who selected into licensure.

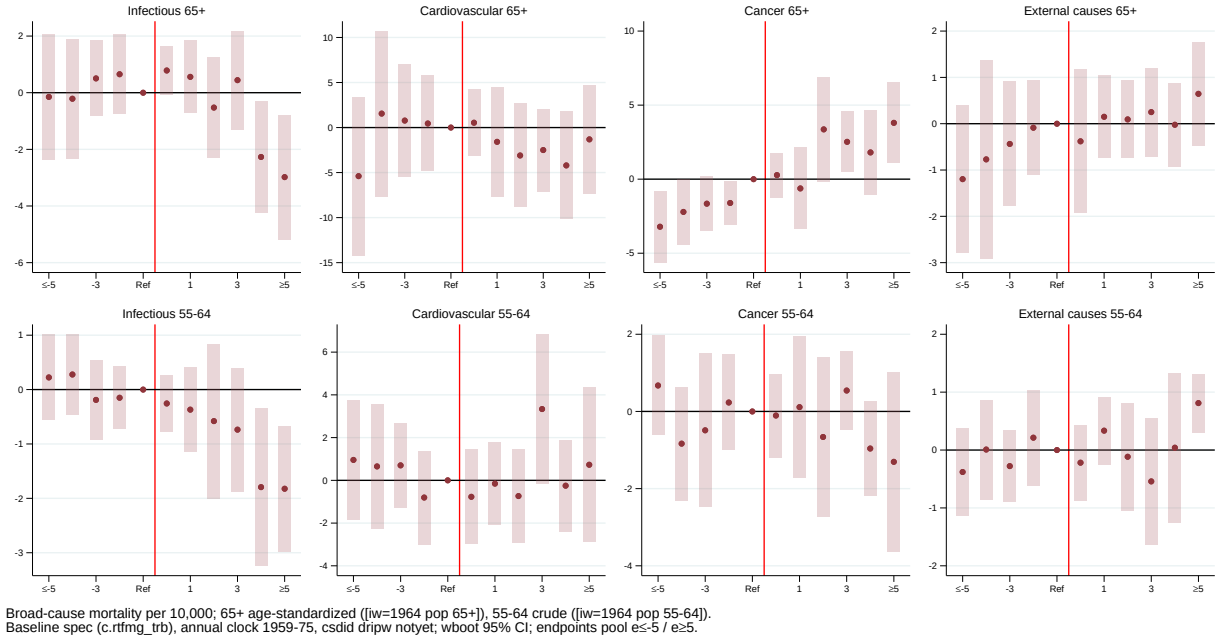


Figure A8: Event Study on Mortality by Cause of Death and Age Group

Notes: Age-standardized mortality by cause group (infectious including pneumonia and influenza, cardiovascular, cancer, and external causes), per 10,000 residents. The upper row reports estimates for ages 65 and over; the lower row reports ages 55–64. Code ranges defining each group are listed in Appendix Table A4. Robustness appears in Appendix Figure A30.

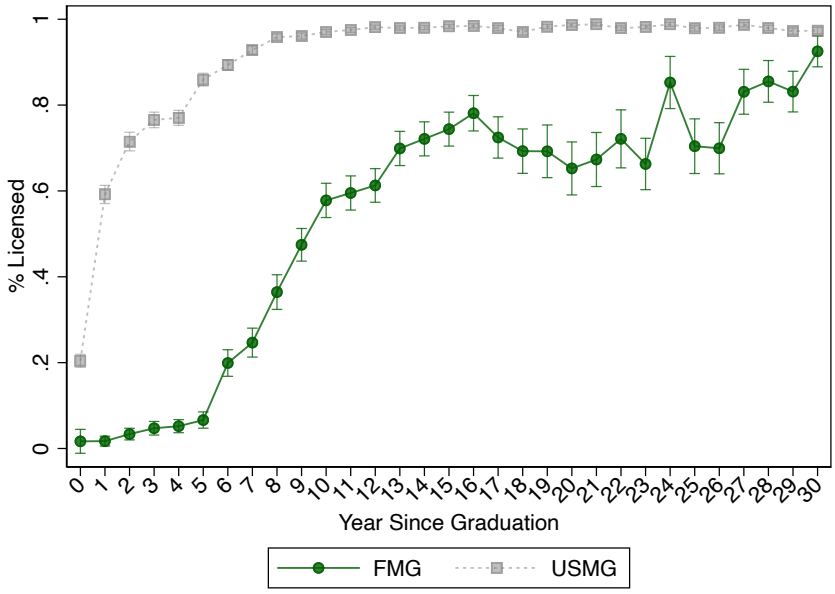


Figure A9: Licensing Rate by Years since Graduation: FMGs vs. USMGs

Notes: Share of physicians holding a full state license, by years elapsed since medical school graduation, computed from the 1964 American Medical Directory (analysis states only). FMG = foreign medical graduate; USMG = U.S. medical graduate.



CALIFORNIA

LOS ALTOS

MIKKELSEN, Peter C• 444 LASSEN APT 3 94022• B78 ME023-04-07 00 TOP11
MILLER, Duane C• 746 ALTOS OAKS DR 94022• B34 ME039-01-80 L62 P TOP01 SS516
MONTGOMERY, Peter Terrill• 966 CLINTON RD 94022• B36 ME051-04-62 NB63 GS TOP05
MURPHY, Vincent Timothy• 794 ALTOS OAKS DR 94022• B23 ME010-02-47 L52 NB49 AB30 OBG TOP01
NELSON, Algot Reginald• 762 UNIVERSITY AVE 94022• B02 ME008-01-28 L46 AB55 PD, PH TOP09 SS312
NELSON, Robt C• 762 ALTOS OAKS DR 94022• B17 ME038-06-50 L57 AB35 OPH TOP01 SS108
NORBERG, Carl Einar• 141 JORDAN AVE 94022• B93 ME026-04-28 00 TOP11
ORECCHIA-CUBICCIOTTI, Ida• 1125 LAS FLORES CT 94022• B78 ME041-07-07 L32 00 TOP11
OTT, Chas Edward• B11 ALTOS OAKS DR 94022• B27 ME016-02-57 L58 GS TOP01
PETERSON, Carl Andrew• 763 ALTOS OAKS DR 94022• B17 ME026-04-43 L50 AB50 PATH TOP01 SS828,848
PHILLIPS, Donald F• 795 ALTOS OAKS DR 94022• B26 ME024-07-49 L56 NB50 AB85 GS,CD TOP01 SS336
POPPY, James Henry• 794 ALTOS OAKS DR 94022• B27 ME025-01-54 L59 AB30 OBG TOP01 SS300
PRITCHARD, Jacob Le Roy• 439 DEODORA DR 94022• B85 ME005-02-18 L16 00 TOP11
REINERTSEN, Bernhard R• 1187 LAURELES DR 94022• B93 ME016-01-20 L24 AB70 00 TOP11 SS024
RICHARDS, John Clarence• 1000 FREMONT AVE 94022• B22 ME038-06-46 L49 AB55 PD TOP01
RICHARDS, Richard Kohn• 1580 WAKEFIELD TERR 94022• B04 ME407-21-31 ANES TOP05 SS636,792,312,084
RICHERT, Joel Hiram• P O BOX 1136 94022• B24 ME016-11-48 L55 AB20 PUD, IM TOP01 SS696,312
RONEY, James G, Jr• 600 UNIVERSITY 94022• B18 ME056-06-48 L49 NB50 AB70 GPM,99 TOP10 SS318
SHAHINIAN, Lee E• 75 2ND ST 94022• B09 ME005-11-36 L36 AB45 OTO TOP01 SS420
SMALL, Marvin Myron• 390 DOMINGO WAY 94022• B36 ME023-01-61 L65 PD TOP01
THOMPSON, Jack E• 1545 KENSINGTON CIRCLE 94022• B26 ME056-06-52 L53 AB05 ANES TOP01 SS636
VERBARG, Lydia L• BOX 342 94022• B08 ME035-09-37 L38 NB38 PH TOP09
VOGAN, Guy Sedgwick• 1000 RILMA LANE 94022• B91 ME041-12-16 00 TOP11
WALLACE, Seymour• 2261 VIA MADEROS 94022• B29 ME035-01-59 L61 NB60 ANES TOP01 SS636
WELLAND, Frederick Hurd• 35 SYLVIAN WAY 94022• B33 ME024-01-59 L63 NB60 AB20 IM TOP01
WESTOVER, Raymond P• 1079 LOS ALTOS 94022• B93 ME030-05-18 00 TOP11

LOS ALTOS

WHEAT, Richard Pierce• 731 ALTOS OAKS DR 94022• B24 ME023-07-48 L49 NB49 AB20 IM TOP01 SS654,312,024

LOS ALTOS HILLS- SANTA CLARA

ABILDGAARD, Wm Harry• 10520 MADGADENA 94022• B21 ME005-12-52 L52 00 TOP12
DOANE, Allan R• 24180 DAWNRIDGE DR 94022• B13 ME064-01-42 GPM TOP10
DUISENBERG, Chas Edward• 13023 LA CRESTA DR 94022• B18 ME005-11-44 L44 AB80 R TOP01 SS912,324
DUMAS, Kenneth J• 25446 ALTAMONT RD 94022• B26 ME035-09-52 L62 NB53 00 TOP10
FINLAYSON, Norris Bryner• 27881 ELENA RD 94022• B32 ME024-01-57 L59 NB58 AB50 PATH TOP08 PA05-02
JOHNS, Clara Ritchie• 12882 VISCAINO 94022• B16 ME016-02-41 L57 AB70 GPM, PD TOP05
KARCH, Robt Wm• 26101 EUCALYPTUS LANE 94022• B30 ME017-20-55 L56 ANES TOP01 SS636
LE DREW, Lloyd S• 26792 TAAFFE RD 94022• B32 ME021-01-59 L65 OBG TOP01 SS300
MC KEE, Wm Dinwiddie• 13888 FREMONT PINES LANE 94022• B31 ME024-01-57 L58 NB58 AB20 A,IM TOP01 SS036
MC PADDEN, Arthur Jos• 25881 ESTACADA DR 94022• B27 ME050-02-55 L61 NB56 AB55 PD,99 TOP01 SS132
SOLOMON, Geo Freeman• 26871 MOODY RD 94022• B31 ME005-11-55 L56 AB75 P TOP06 SS516,576,552 PA05-11
TABER, Ben-Zion• 12374 MELODY LANE 94022• B27 ME024-01-52 L53 NB53 AB30 OBG TOP01 SS300

LOS ANGELES- LOS ANGELES

AABERG, E Lindeman• 10921 WILSHIRE BLVD 90024• B07 ME016-06-35 L48 AB85 GS TOP01 SS888,630,390
AABERG, Monrad E• 1001 GAYLEY AVE 90024• B06 ME016-06-33 L46 AB30 OBG TOP01 PA05-14
AARON, Ruth• 10444 SANTA MONICA BLVD 90025• B10 ME016-02-37 L37 P TOP01 SS516,008 PA05-06
ABARBANEL, Abraham Robt• 6515 WILSHIRE BLVD 90048• B12 ME035-08-36 L45 OBG,99 TOP01 SS888,630
ABBOTT, Kenneth H• 1710 BROOKLYN AVE 90033• B11 ME005-12-36 L36 NB36 AB75, 25 NS,N TOP01 SS840,888, 336,852,084 PA05-12
ABDO, Francis Jos• 544 N ROSSMORE 90004• B10 ME016-01-37 L39 GP TOP01
ABDULIAN, Dani Hovhannes• 6753 HOLLYWOOD BLVD 90028• B27 ME016-11-58 L59 IM,CD TOP01
ABDULIAN, Hagop J• 6753 N HOLLYWOOD BLVD 90028• B23 ME605-01-49 L51 NB52 AB85 GS TOP01 SS336

LOS ANGELES

ABOWITZ, Murray• 6333 WILSHIRE BLVD 90048• B11 ME154-07-37 L40 IM TOP01 SS576
ABRAHAM, Edward Arnold• 1212 W JEFFERSON BLVD 90007• B36 ME005-15-63 L64 GP TOP01
ABRAHAM, Saml Victor• 9201 SUNSET BLVD 90069• B99 ME016-11-28 L31 AB35 OPH TOP01 SS108,732,336,888,495
ABRAHAMS, David• 10921 WILSHIRE BLVD 90024• B07 ME016-06-35 L48 AB85 GS TOP01 SS888,630,390
ADAMS, Joel Lee• 1200 N STATE ST 90033• B37 ME048-12-63 L64 IM TOP04
ADAMS, John M• UNIV CALIF MED CENTER 90024• B05 ME035-01-33 L50 NB35 AB55 PD TOP06 SS492,132,276,792, 600,612 PA05-14
ADAMS, Patricia V Orm• BOX 42145 90042• B21 ME038-40-46 L54 ANES TOP01
ADAMS, Richard Earle• 1200 N STATE ST 90033• B37 ME005-06-65 L66 OBG TOP04
ADAMS, Robt Maurice, Jr• LOS ANGELES COUNTY HOSP 90033• B41 ME021-01-66 00 TOP03
ADAMS, Wm S• UNIV CALIF MED CENTER 90024• B19 ME035-45-43 L49 NB48 AB20 IM TOP05 SS312 PA05-14
ADAMSON, France Marshall• 1200 N STATE ST 90033• B38 ME005-06-84 L65 PD TOP04
ADELSTEIN, Leo Joel• 1930 WILSHIRE BLVD 90057• B02 ME024-07-25 L27 NS,N TOP01 SS888,336,084
ADEY, Wm Ross• UNIV CALIF HLTH SCIENCES CTR 90024• B22 ME143-01-43 N,99 TOP06 SS024 PA05-14

LOS ANGELES

ADHAM, Nabil Faik• U-SOUTHERN CALIF DEPT OF MED 90033• B33 ME605-01-58 IM TOP05
ADKINS, Morris Pardee• 5510 DUARTE ST 90058• B25 ME005-15-62 L58 GP,GS TOP02
ADLER, Denis Ciklej• 2010 WILSHIRE BLVD 90057• B14 ME473-02-38 L45 NB44 AB80 R TOP01 SS588,912,324 PA05-12
AINLAY, Geo Wm, Sr• 11633 SAN VICENTE BLVD 90049• B96 ME030-05-27 L43 IM TOP01 SS390
AKASHI, Ronald Hidetoshi• WHITE MEM HOSP-CLINIC 90033• B35 ME005-15-66 00 TOP03
AKHLAGHI, Iraj Ray• 1200 N STATE ST UNIT II 90033• B35 ME005-15-63 L64 IM TOP04
AKIYAMA, Yoshio• 584 N IRVING BLVD 90004• B30 ME005-14-56 L57 AB55 PD TOP01
ALBAECK, Karl L• 3919 BEVERLY BLVD 90004• B11 ME005-15-62 L43 NS,N TOP01
ALBAUGH, Clarence Henry• 321 N LARCHMONT BLVD 90004• B13 ME024-01-36 L41 NB37 AB35 OPH TOP01 SS108,888, 336 PA05-06
ALBAUGH, Clarence Jacob• 321 N LARCHMONT BLVD 90004• B78 ME041-01-08 L42 OPH TOP01
ALBAUGH, Kathryn L• 321 LARCHMONT BLVD 90004• B16 ME005-06-49 L49 NB49 AB35 OPH TOP01 SS108 PA05-12
ALBERTI, John Birch• UNIV OF CALIF HOSP DEPT OF RAD 90024• B32 ME056-06-59 L60 AB80 R TOP04

LOS ANGELES

ALBIN, David Lawrence• 8016 COWAN AVE 90045• B37 ME005-14-63 L64 P TOP04
ALBIN, Meyer Stanley• 8618 S SEPULVEDA BLVD 90045• B13 ME016-11-39 L45 GP TOP01 SS060
ALBORI, Martin• 739 N HIGHLAND 90038• B14 ME581-01-41 L51 OM,ORS TOP01 SS864
ALBRITTON, David Richards• UNIV OF CALIF HOSP 90024• B40 TOP01 SS060
ALLEN, Howard Harold• 4524 SAN PEDRO PL 90011• B02 ME010-03-28 L29 GP,OBG TOP02
ALLEN, Howard Norman• 3685 MENTONE 90034• B36 ME016-11-60 L62 NB61 IM TOP04
ALLEN, John Paul• UNIV OF SOUTHERN CALIF 90033• B29 ME033-05-60 L62 PD,N TOP06 PA05-06
ALLEN, John William• 1720 BROOKLYN AVE 90033• B34 ME005-17-62 L62 IM TOP05
ALLEN, Raymond Archie• UNIV CALIF MED CENTER 90024• B21 ME020-02-46 L54 AB50 PATH TOP01 SS876,828,204, 612 PA05-14
ALLEN, Thos Gorman• LOS ANGELES COUNTY HOSP BX 32 90033• B35 ME005-06-81 L62 AB80 R TOP05 SS324
ALLEN, Wm Clyde• 221 E VERNON AVE 90011• B02 ME010-03-32 L40 GP,OTO TOP02
ALLEN, Wm Leroy, Jr• 3701 STOCKER ST 90008• B19 ME020-02-47 L48 GP,GS TOP02 SS060

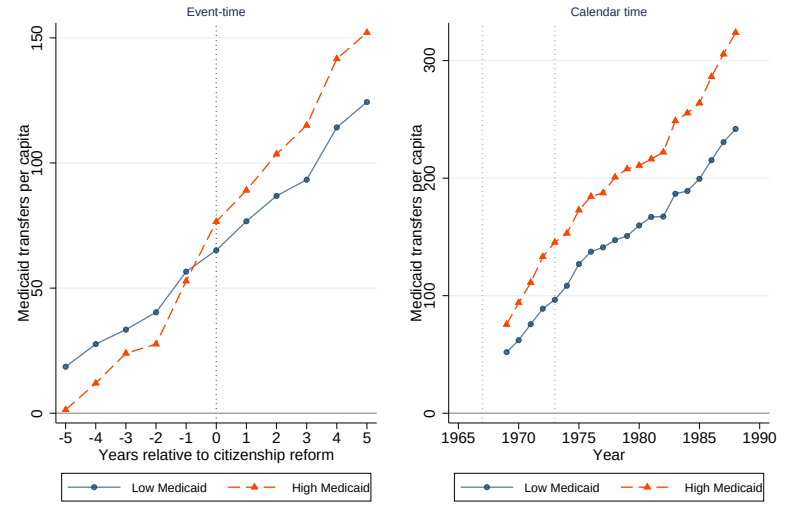
1222

LOS ANGELES- LOS ANGELES

AABERG, E Lindeman• 10921 WILSHIRE BLVD 90024• B07 ME016-06-35 L48 AB85 GS TOP01 SS888,630,390
AABERG, Monrad E• 1001 GAYLEY AVE 90024• B06 ME016-06-33 L46 AB30 OBG TOP01 PA05-14
AARON, Ruth• 10444 SANTA MONICA BLVD 90025• B10 ME016-02-37 L37 P TOP01 SS516,008 PA05-06

Figure A10: American Medical Directory

Source: Data comes from American Medical Directory (1967).



Note: Medicaid spending is defined as per-capita Medicaid transfer payments from the BEA Regional Economic Information System (REIS), measured as total Medicaid medical-vendor payments received by a county's residents divided by county population. Available from 1969 onward.

Figure A11: Raw Trends in Per-Capita Medicaid Transfers by County Exposure Group

Notes: Raw trends in per-capita Medicaid transfers for counties classified as high- versus low-Medicaid-exposure, based on above- versus below-median 1964 AFDC participation among women. The divergence after Medicaid enactment confirms that the pre-reform AFDC split captures subsequent differences in Medicaid intensity.

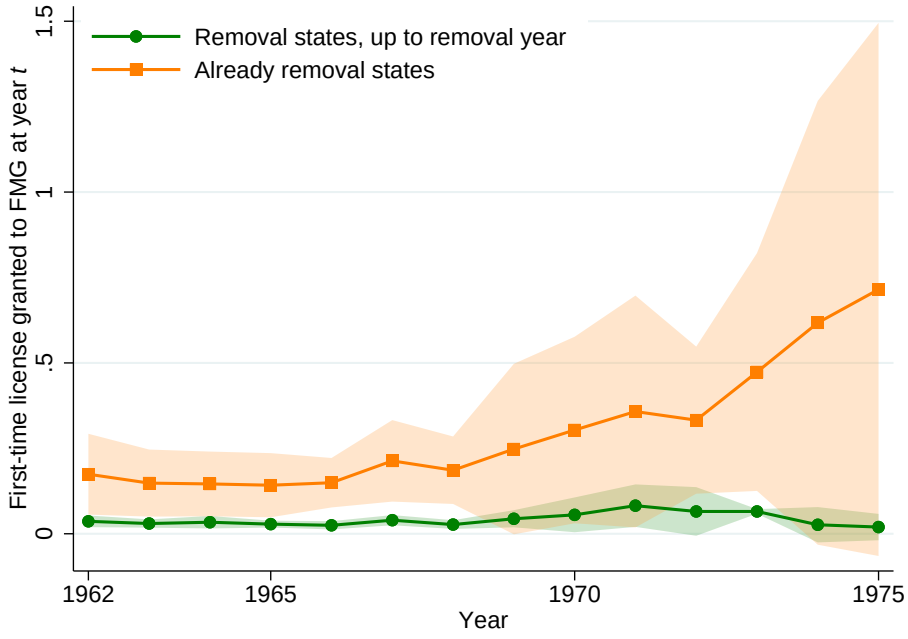


Figure A12: First-time FMG licensure: removal vs. already-removed states

Notes: The figure plots the group-year mean of first-time licenses granted to FMGs (per 10,000 population) with 95% confidence bands. The sample comprises 38 states: 23 removal states that removed the citizenship requirement between 1965 and 1979, and 15 already-removed states that never imposed a citizenship requirement. Removal states (green) are shown only for calendar years strictly before each state's own removal year; consequently the green group thins mechanically at later years (by 1975 it rests on only the latest-removing states), which is reflected in the widening confidence band. Already-removal states (orange) are shown for all available years. Alaska, Arizona, and Hawaii, states that removed the requirement at or before the 1964 baseline, and states flagged for ambiguous reform timing are excluded.

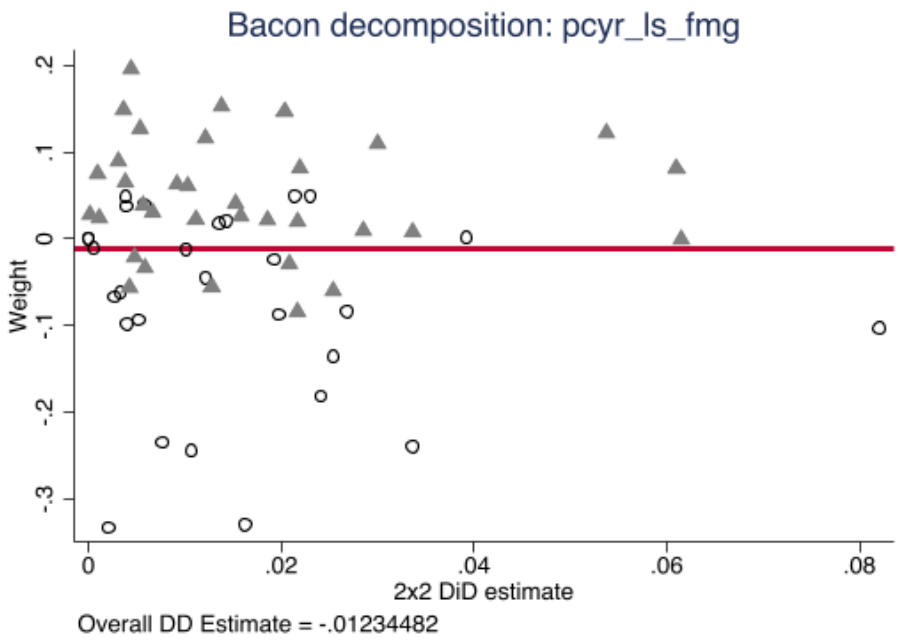
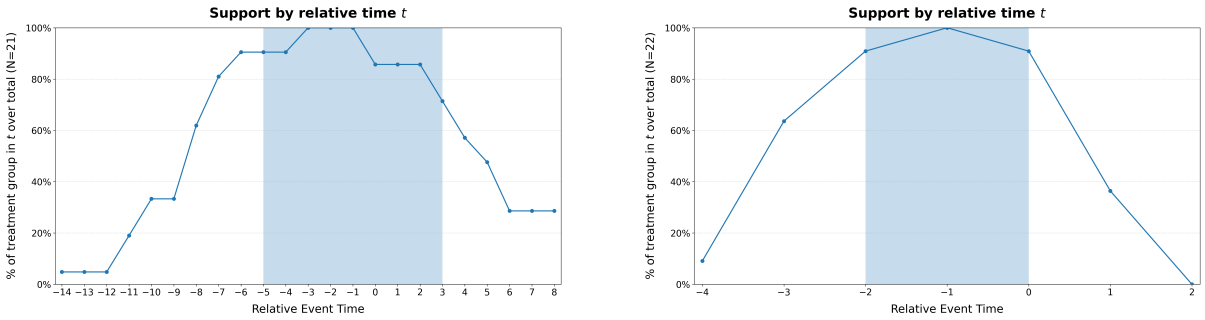


Figure A13: Goodman-Bacon Decomposition of TWFE Estimates: FMGs Newly Licensed in Year t

Notes: Each point represents one pairwise 2×2 comparison underlying the naïve two-way fixed-effects (TWFE) estimate. Circles denote comparisons between a treated cohort and not-yet-treated states (the valid comparison group); triangles denote comparisons where an already-treated cohort serves as the control (the inappropriate comparison group). The size of each point reflects its weight in the TWFE decomposition. The attenuated TWFE estimate (-0.01) is pulled toward zero by the negative-weighted triangle comparisons, while point estimates from the valid circle comparisons cluster around 0.1, consistent with the preferred Callaway–Sant'Anna estimates.

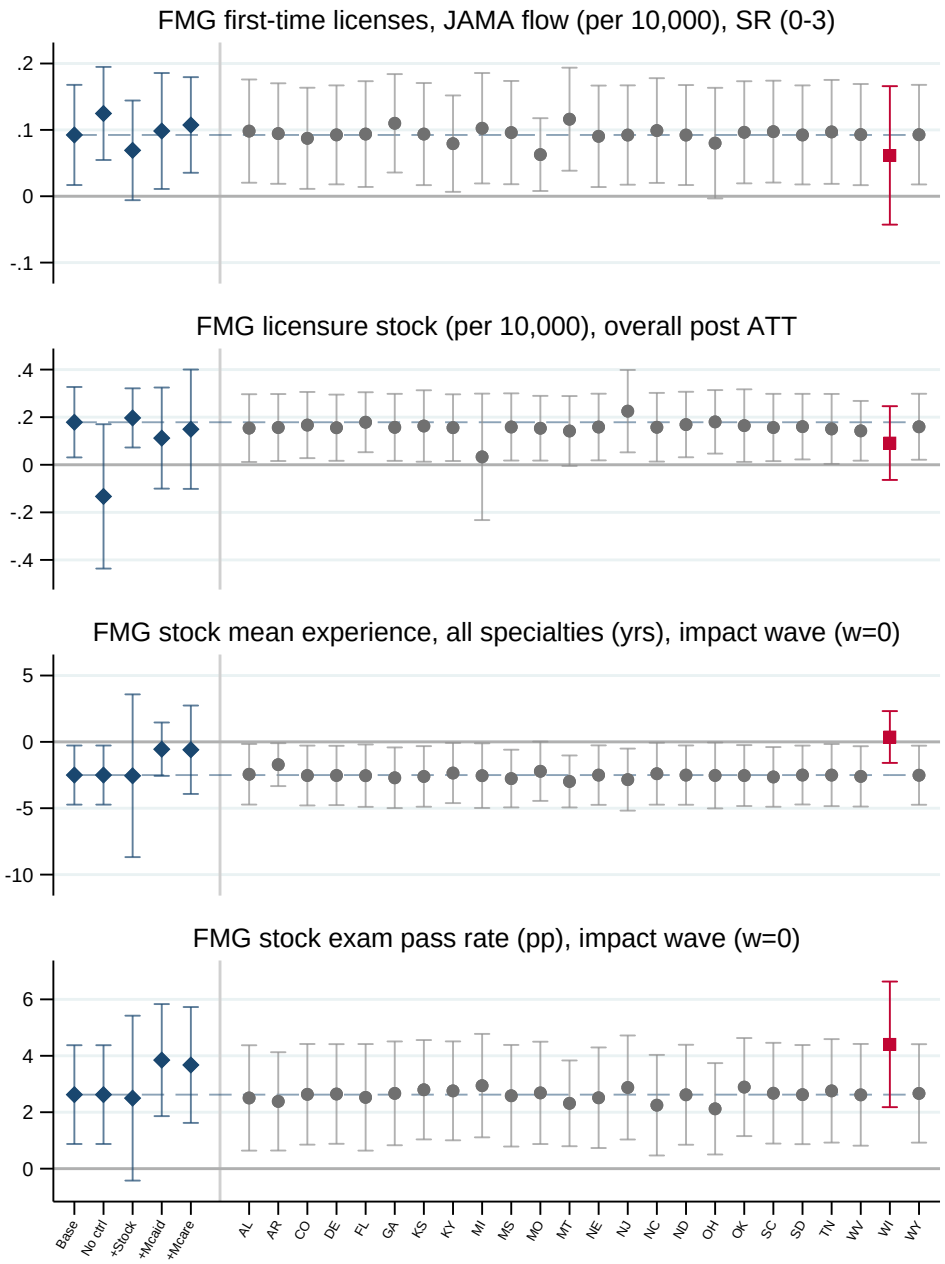


(a) Annual data

(b) AMD data

Figure A14: Proportion of Treated Units by Event Time

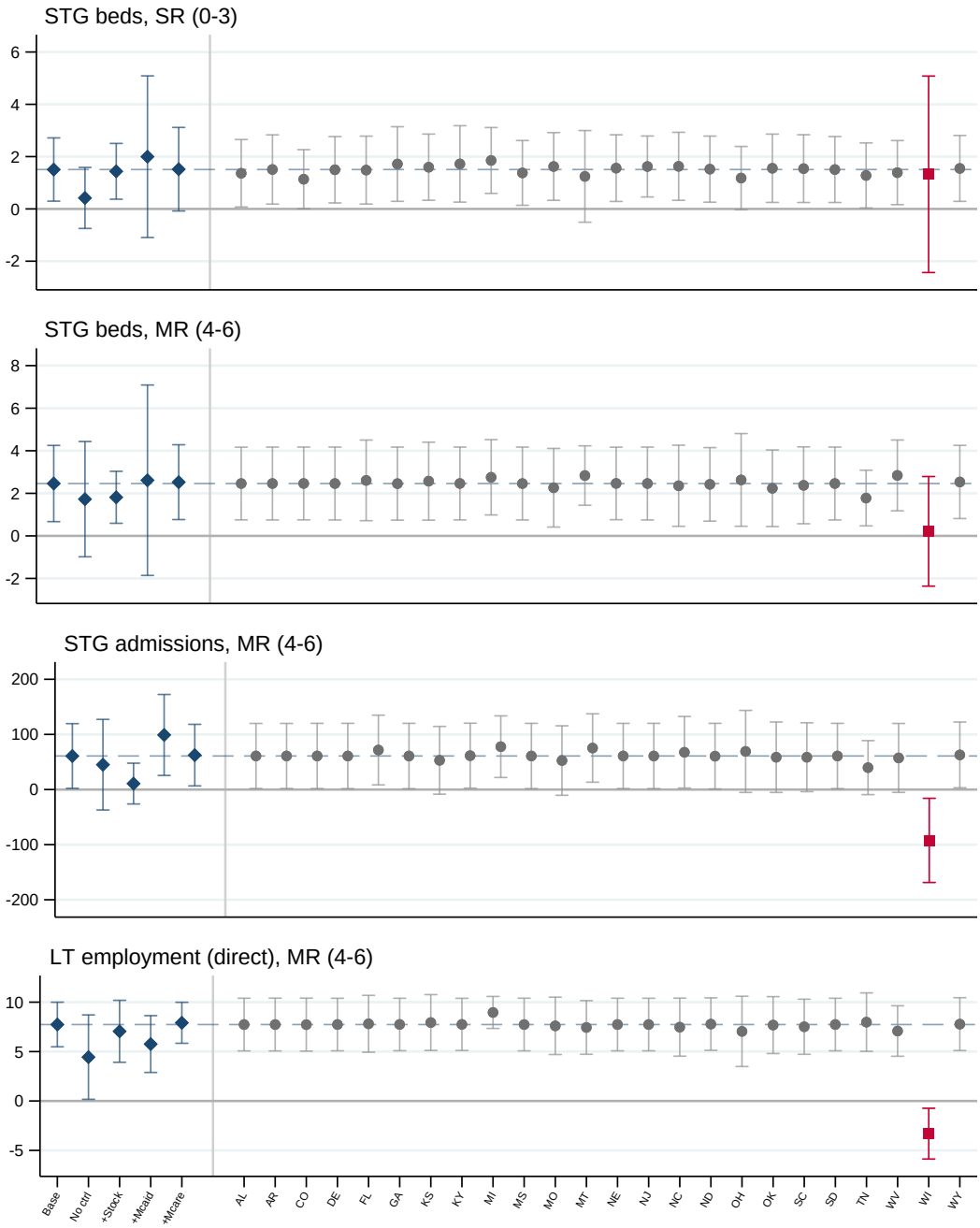
Notes: This figure shows the proportion of treated units in the analysis sample observed at each event time, for the annual (panel a) and AMD (panel b) samples. The event-study window displayed throughout the analysis (-5 to $+5$) is shown in blue shading.



Diamonds: control specifications (wboot CIs). Circles: leave-one-state-out (analytic CIs); WI drop in red.
Dashed line = baseline point; solid grey = zero. Rows 1-2 ladder order: Base / No ctrl / +Stock / +Mcaid / +Mcare.
Rows 3-4: wave clock; Base = no covariate, so the No-ctrl rung repeats it by construction; +Stock = LS share. Row 2: analytic CIs.

Figure A15: Robustness of the Physician Labor Market Estimates

Notes: Each panel reports the estimate for one outcome under two robustness exercises plotted on a common axis. Diamonds trace the control ladder: the baseline specification conditioning on the 1964 FMG-among-trainees ratio, then each additional pre-treatment control in turn. Circles show the leave-one-state-out distribution, re-estimating the specification while dropping one treated state at a time; the Wisconsin exclusion is highlighted in red. Outcomes are FMG first-time licenses, licensure stock, mean experience, and exam pass rate.



Diamonds: control specifications (wboot CIs). Circles: leave-one-state-out (analytic CIs); WI drop in red.
Dashed line = baseline point; solid grey = zero.

Figure A16: Robustness of the Hospital Capacity Estimates

Notes: Each panel reports the estimate for one outcome under two robustness exercises plotted on a common axis. Diamonds trace the control ladder: the baseline specification conditioning on the 1964 FMG-among-trainees ratio, then each additional pre-treatment control in turn. Circles show the leave-one-state-out distribution, re-estimating the specification while dropping one treated state at a time; the Wisconsin exclusion is highlighted in red. Outcomes are short-term general beds in the short and medium run, short-term general admissions in the medium run, and long-term employment in the medium run.

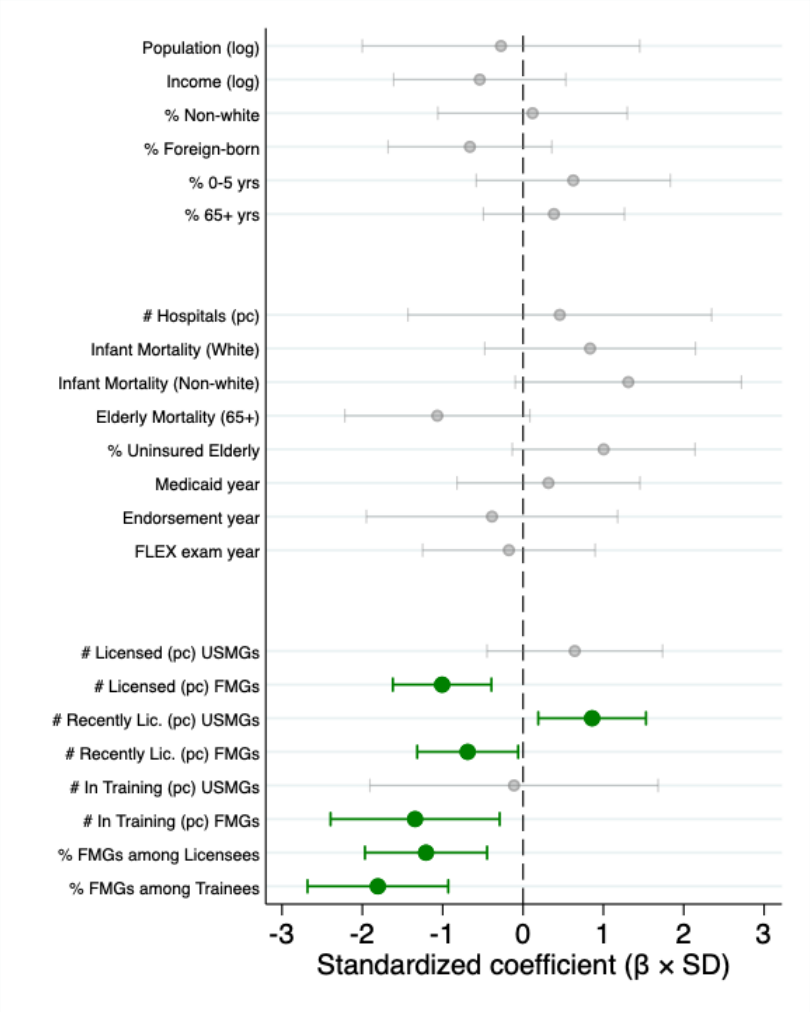


Figure A17: Regressions of Removal Timing on Each Variable

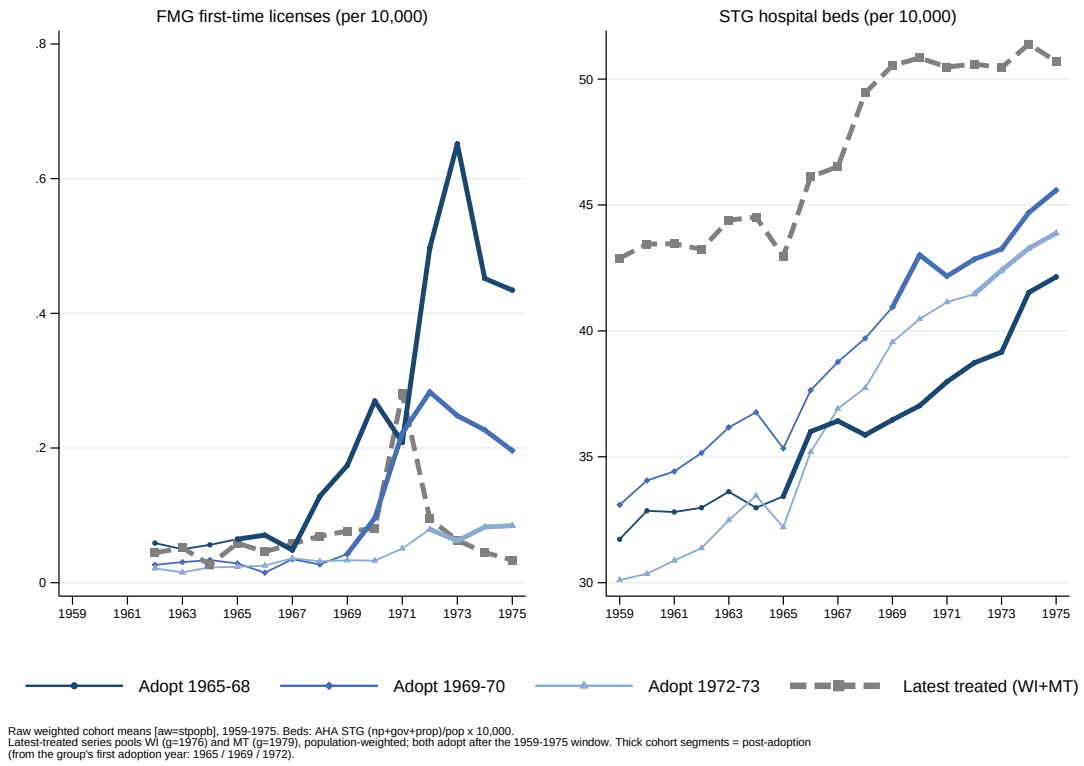
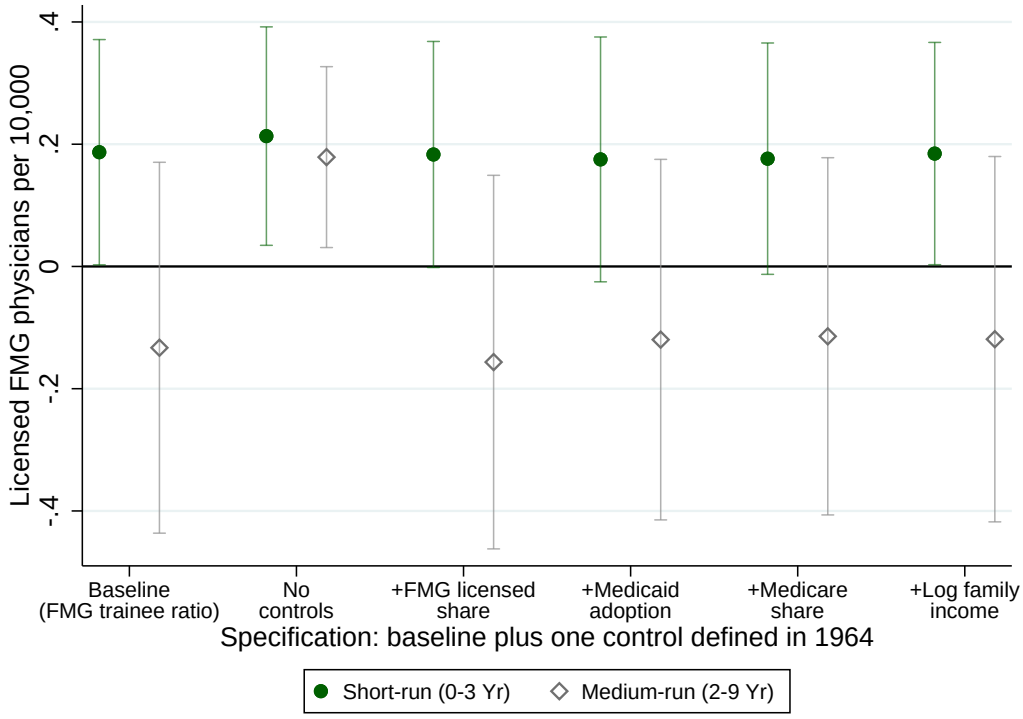
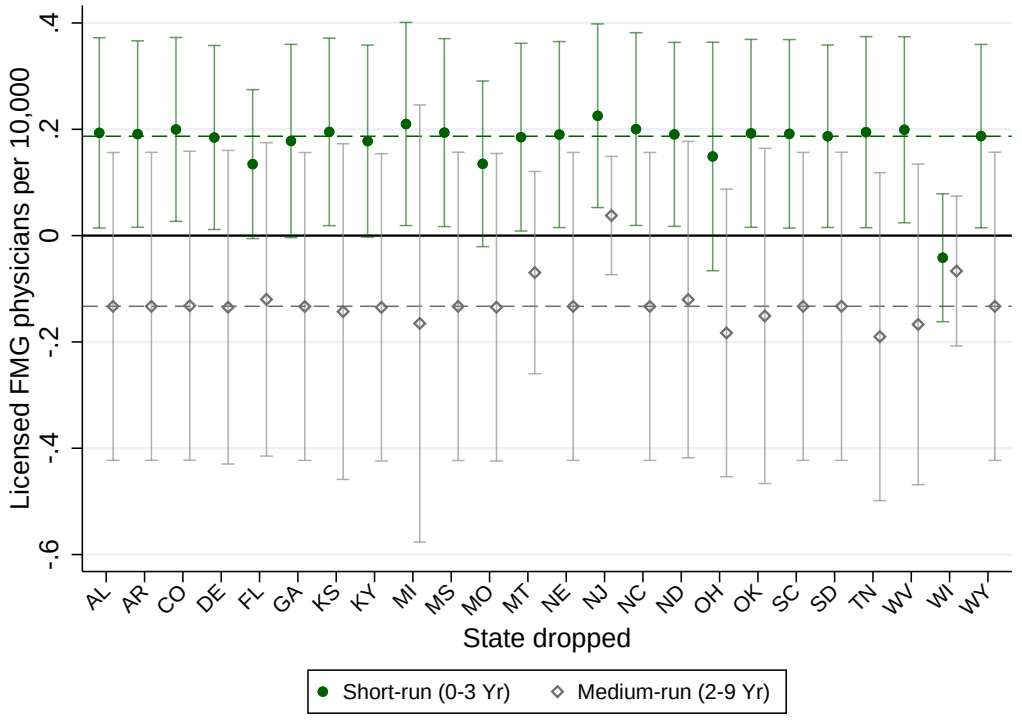


Figure A18: Raw Trends in FMG Licensing Flow and Hospital Beds by Adoption Cohort

Notes: Raw population-weighted outcomes by adoption cohort, with thicker segments marking years after each cohort group's first adoption. The upper panel plots FMG licensing flows; the lower panel plots short-term general hospital beds. The latest-treated comparison group comprises Wisconsin and Montana.

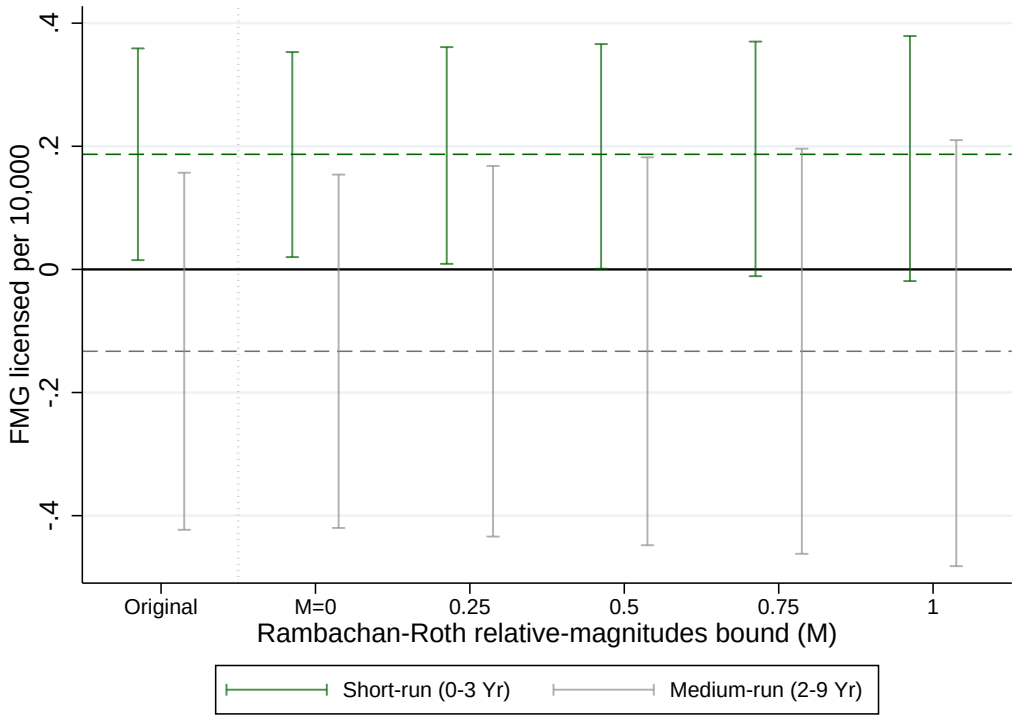


(a) Alternative Controls



(b) Leave-One-State-Out

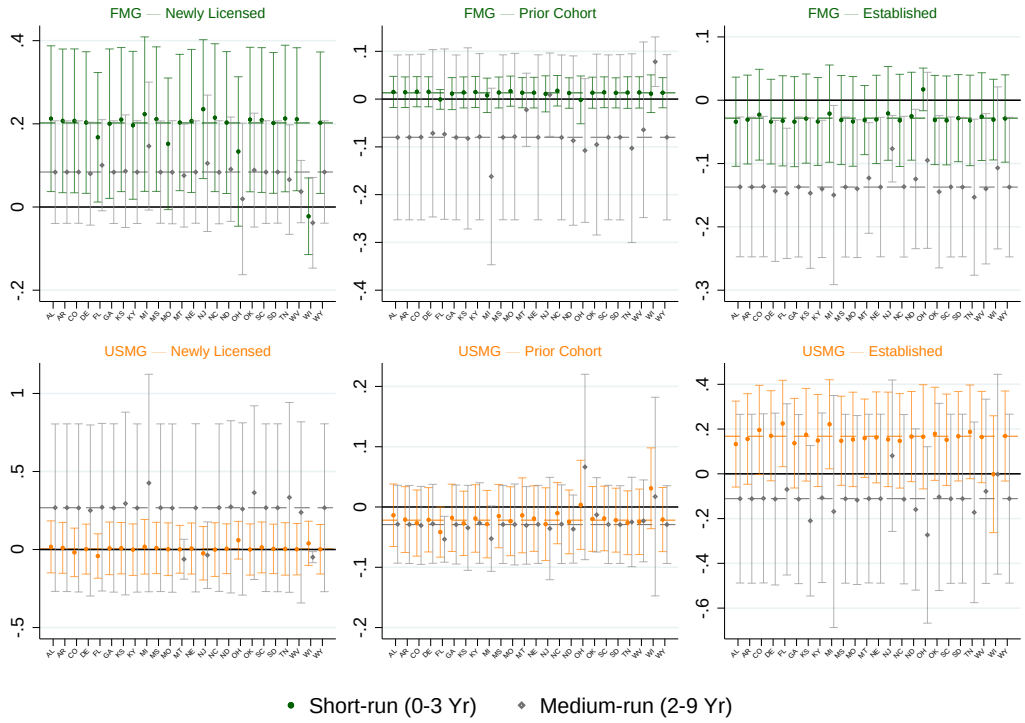
Figure A19: Robustness of the Licensed FMG Stock Estimates



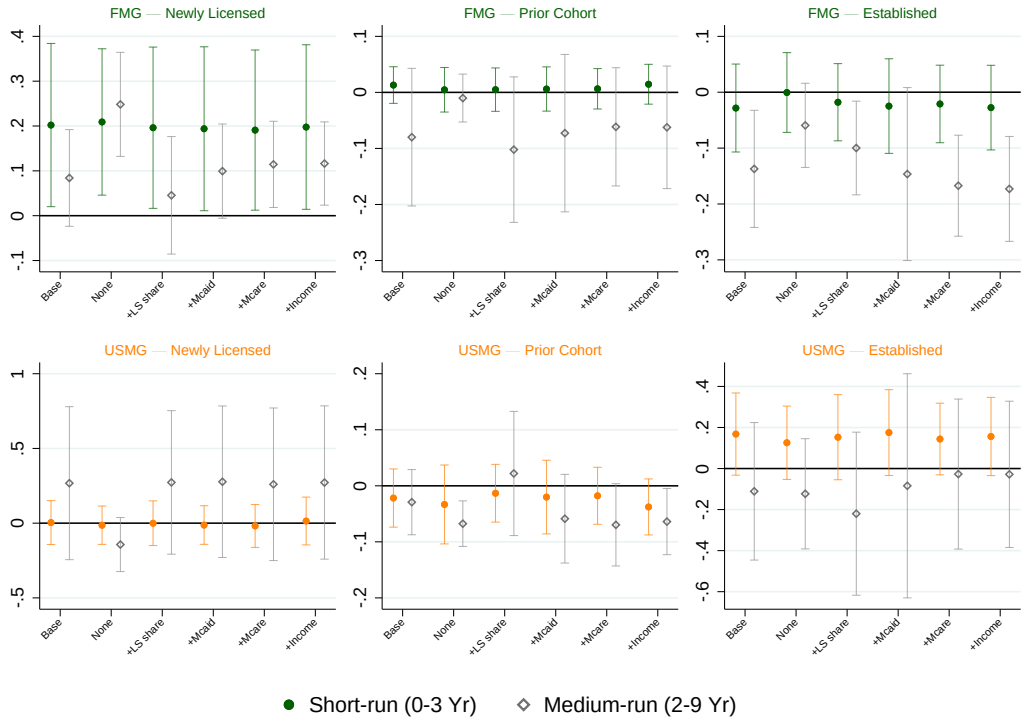
(c) Rambachan–Roth Sensitivity

Figure A19: Robustness of the Licensed FMG Stock Estimates (continued)

Notes: Robustness checks for the main estimates in Figure 4a. Panel (a) varies the baseline covariate; panel (b) drops one treated state at a time; panel (c) reports robust confidence sets allowing the maximal post-treatment parallel-trends violation to be $\overline{M}$ times the maximal pre-treatment violation.

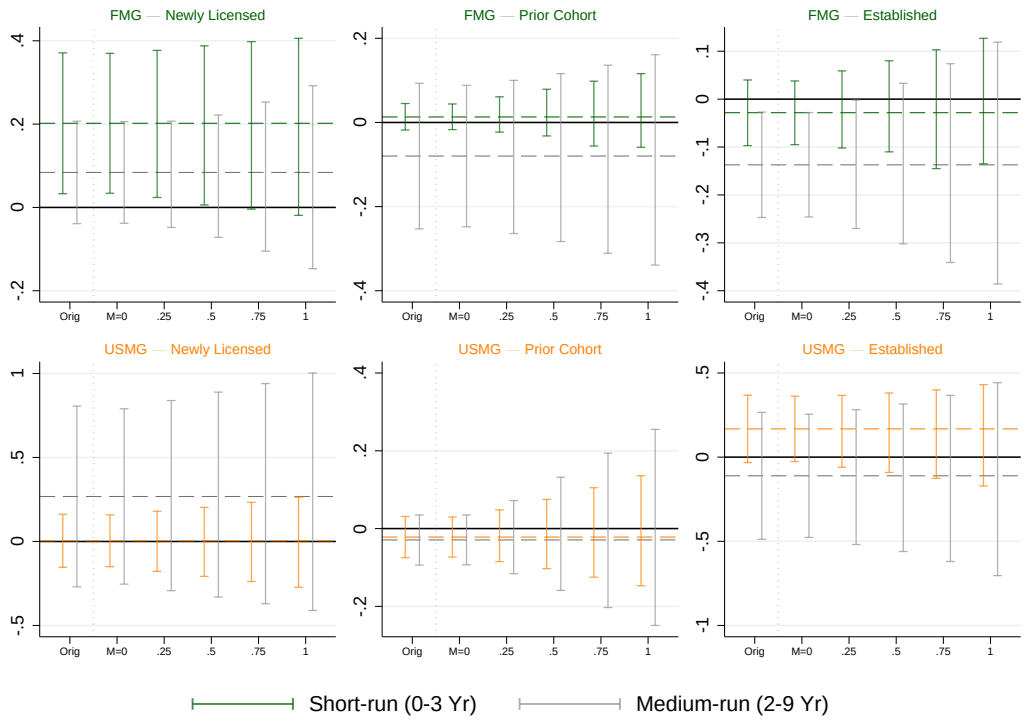


(a) Leave-One-State-Out



(b) Alternative Controls

Figure A20: Robustness of the Stock Decomposition by Recency of Licensure



(c) Rambachan–Roth Sensitivity

Figure A20: Robustness of the Stock Decomposition by Recency of Licensure (continued)

Notes: Robustness checks for the estimates in Figure A4.

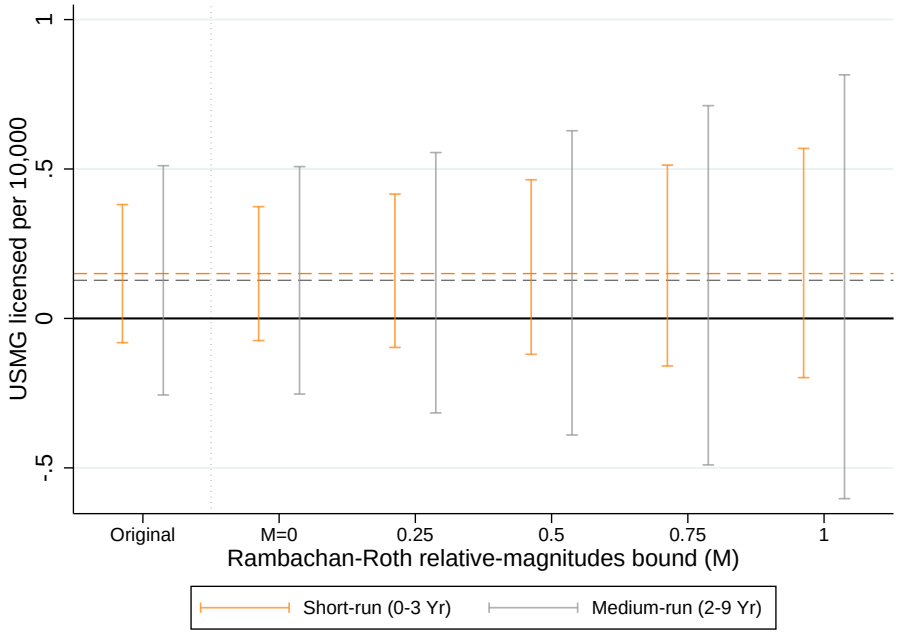


Figure A21: Rambachan–Roth Sensitivity: Licensed USMG Stock

Notes: Robust confidence sets for $\overline{M} \in \{0.25, 0.5, 0.75, 1\}$, where $\overline{M}$ bounds the maximal post-treatment violation of parallel trends relative to the maximal pre-treatment violation.

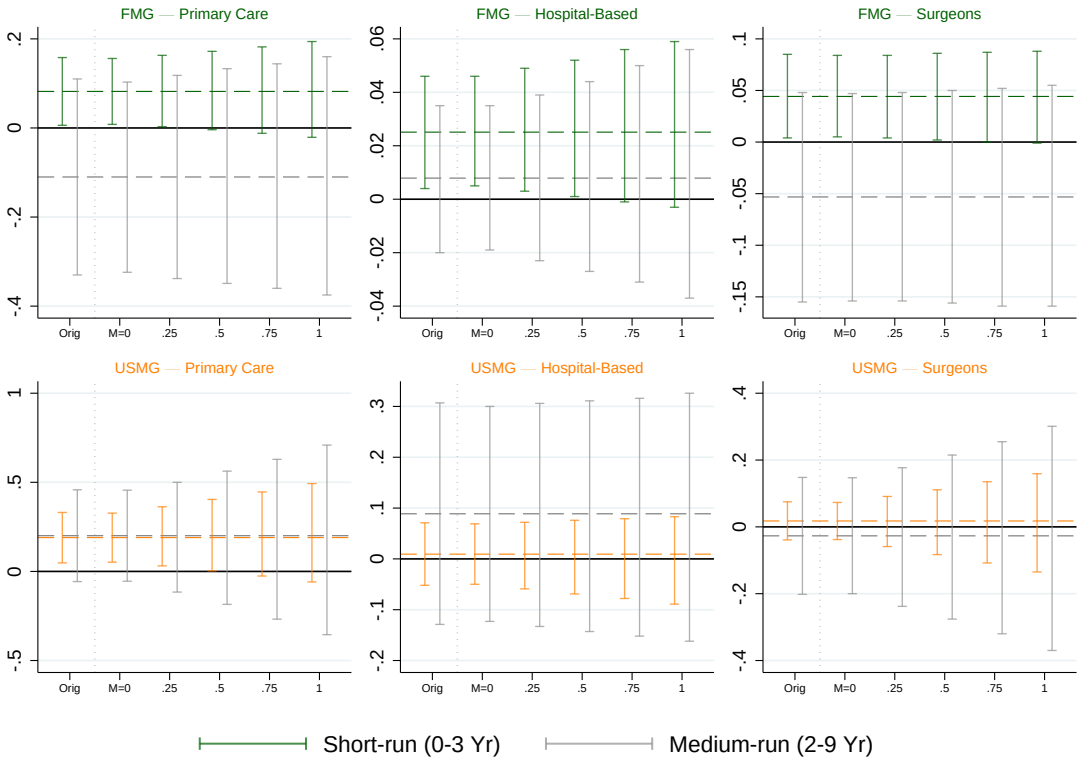
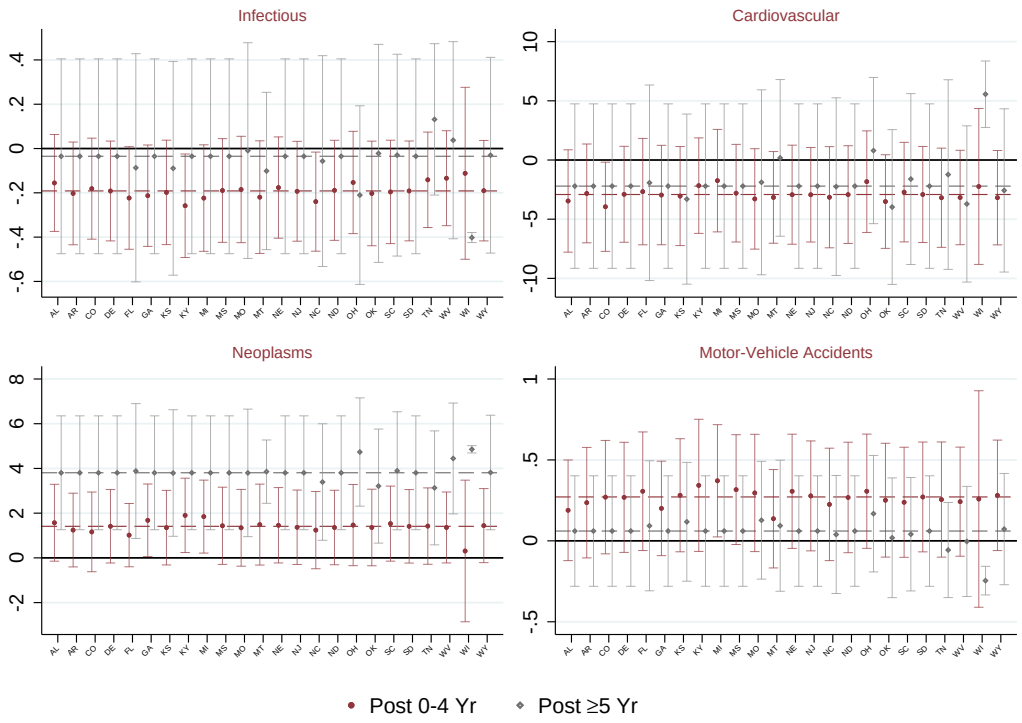
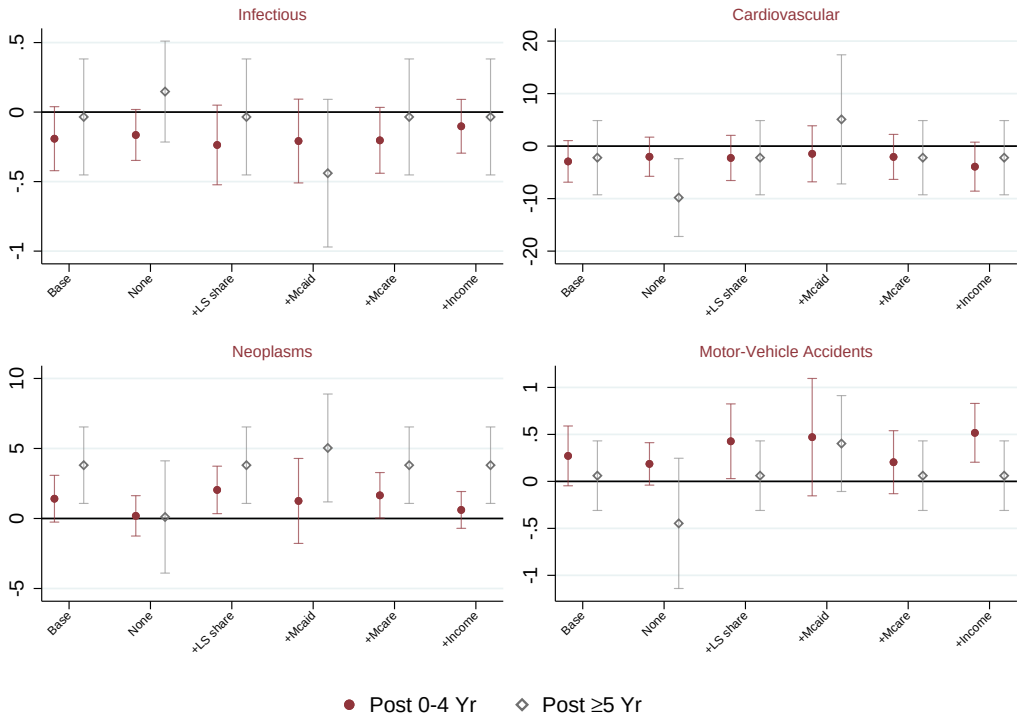


Figure A22: Rambachan–Roth Sensitivity: Licensed Stock by Specialty Group

Notes: Robust confidence sets by specialty group, reported separately for FMGs and USMGs.

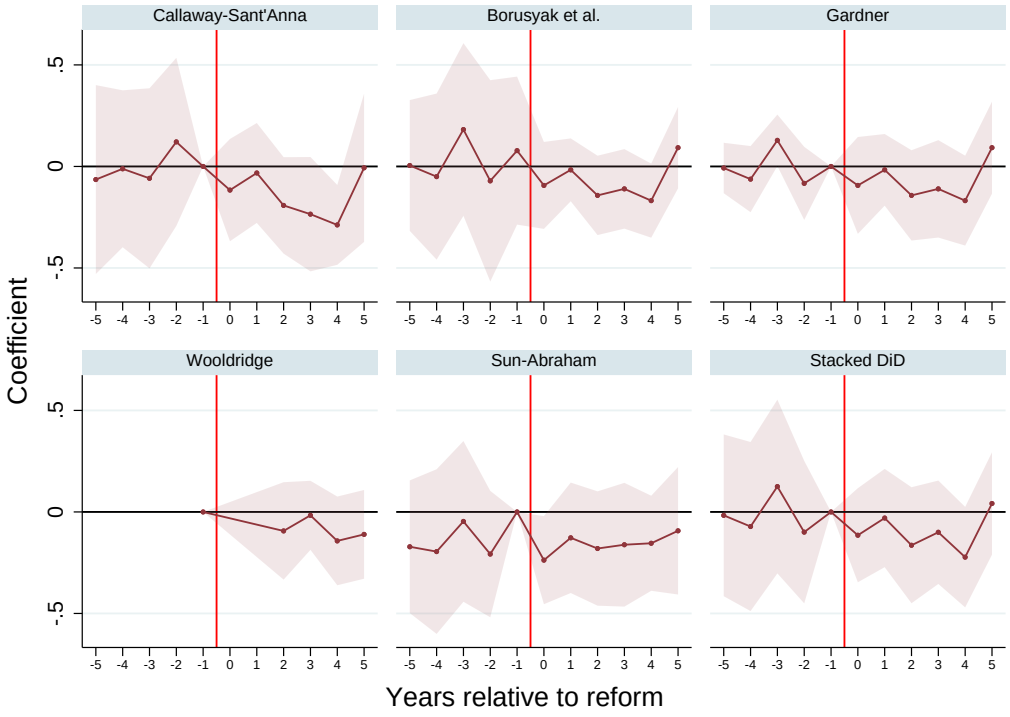


(a) Leave-One-State-Out



(b) Alternative Controls

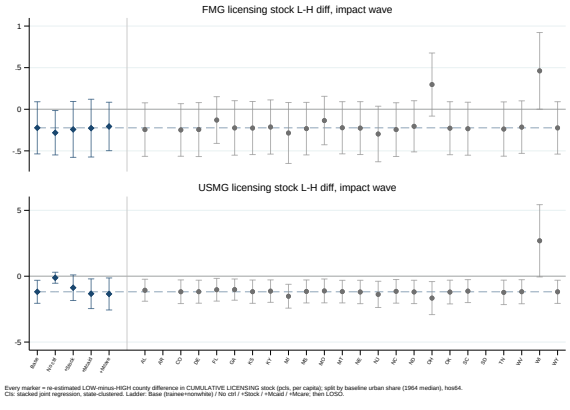
Figure A23: Robustness of Age-Standardized Mortality Estimates, Ages 65 and Over



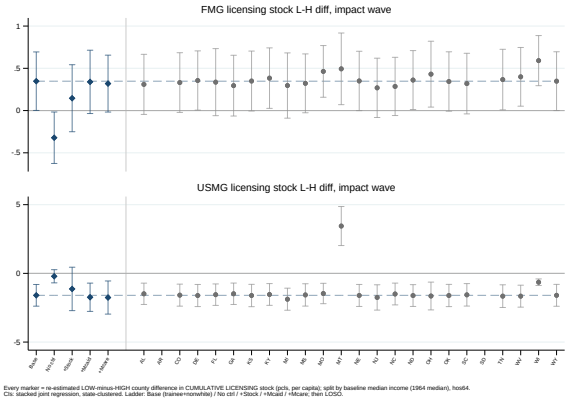
(c) Alternative Estimators, Infectious Causes

Figure A23: Robustness of Age-Standardized Mortality Estimates, Ages 65 and Over (continued)

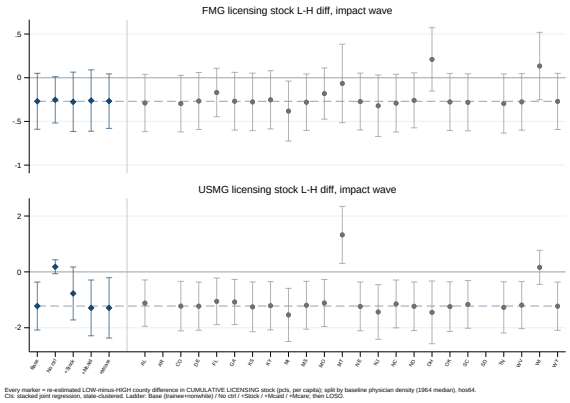
Notes: Panel (a) drops one treated state at a time; panel (b) varies the baseline covariate; panel (c) compares heterogeneity-robust staggered difference-in-differences estimators for infectious-cause mortality.



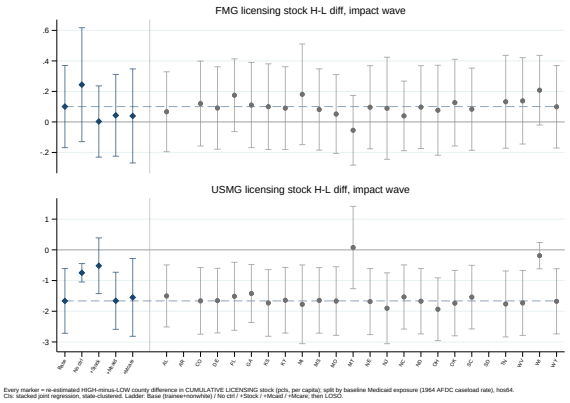
(a) Urban Share



(b) Income



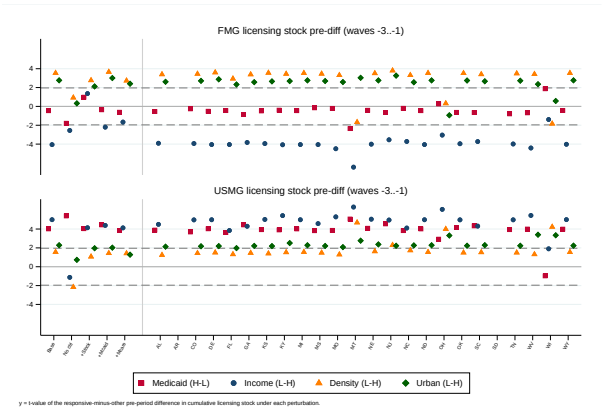
(c) Physician Density



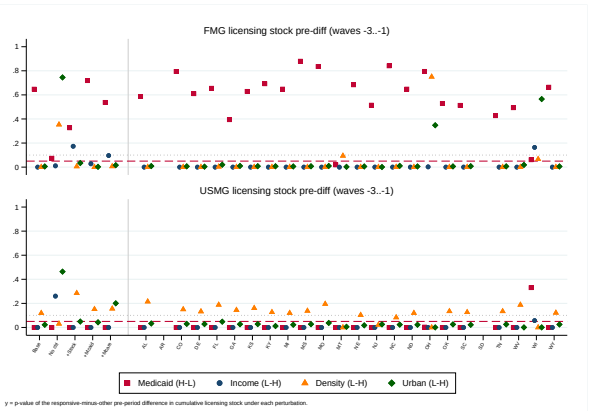
(d) Expected Medicaid Growth

Figure A24: Robustness of the County Divergence in Licensed Physician Stock

Notes: Each panel plots the high-minus-low divergence estimate for one county split across specifications. Arkansas is omitted because county-level characteristics are not available in the 1973 AMD year.



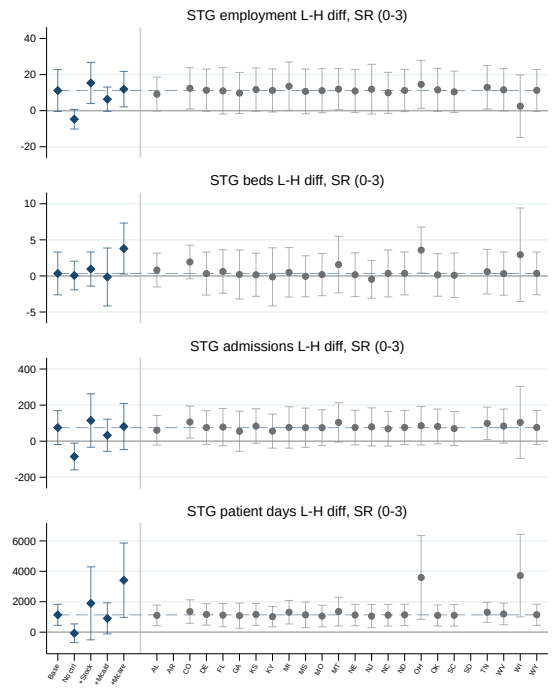
(a) Pre-Treatment Divergence



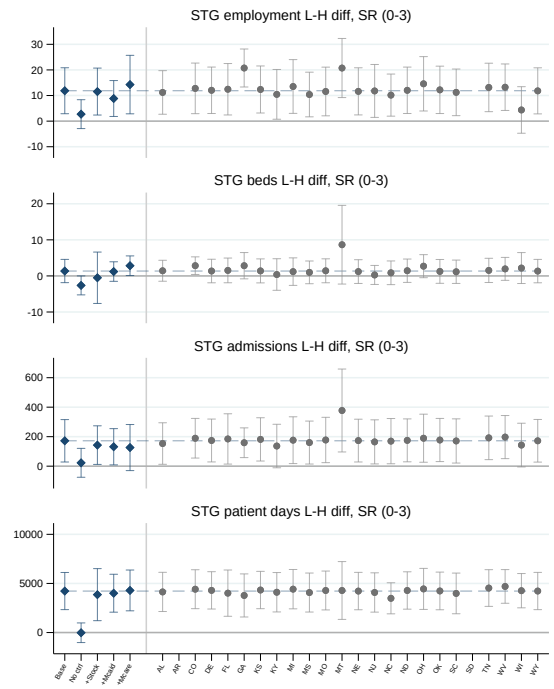
(b) Wald Test p-values

Figure A25: Pre-Treatment Divergence in Licensed Physician Stock by County Split

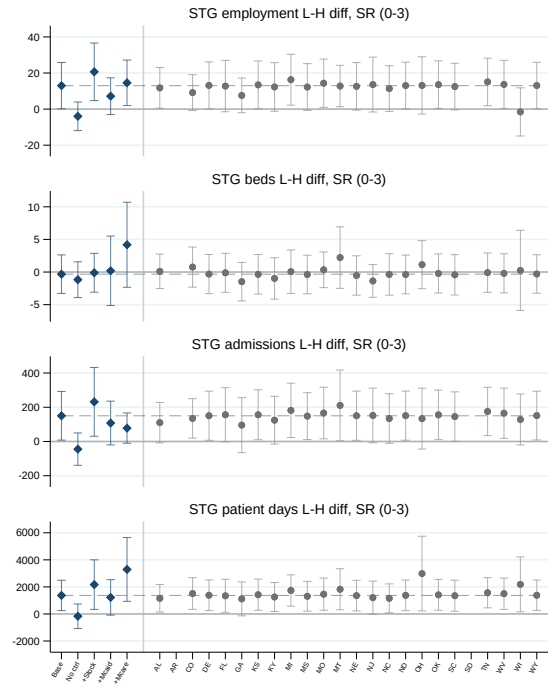
Notes: Panel (a) plots the pre-treatment divergence between county groups; panel (b) reports p-values from the Wald test of equality across partitions, using a state-clustered covariance matrix that preserves covariance across subgroups.



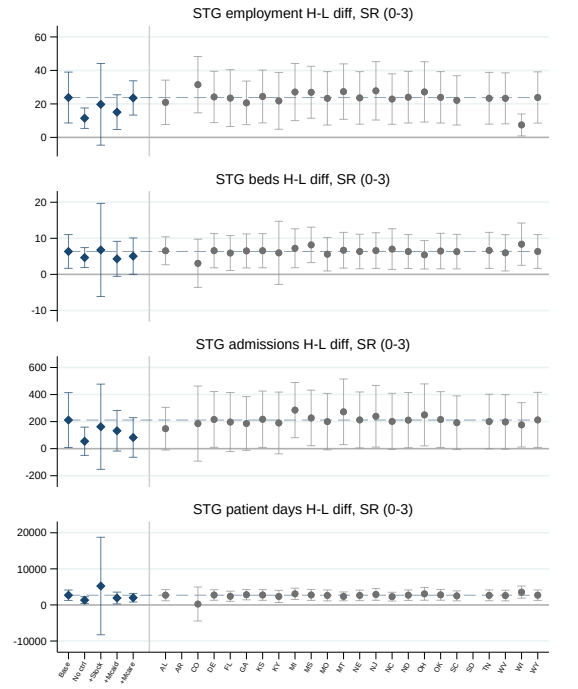
(a) Urban Share



(b) Income



(c) Physician Density



(d) Expected Medicaid Growth

Figure A26: Robustness of the County Divergence in Hospital Input and Output

Notes: Each panel plots the high-minus-low divergence estimate for one county split across specifications, for short-run hospital outcomes. Arkansas is omitted because county-level characteristics are not available in the 1973 AMD year.

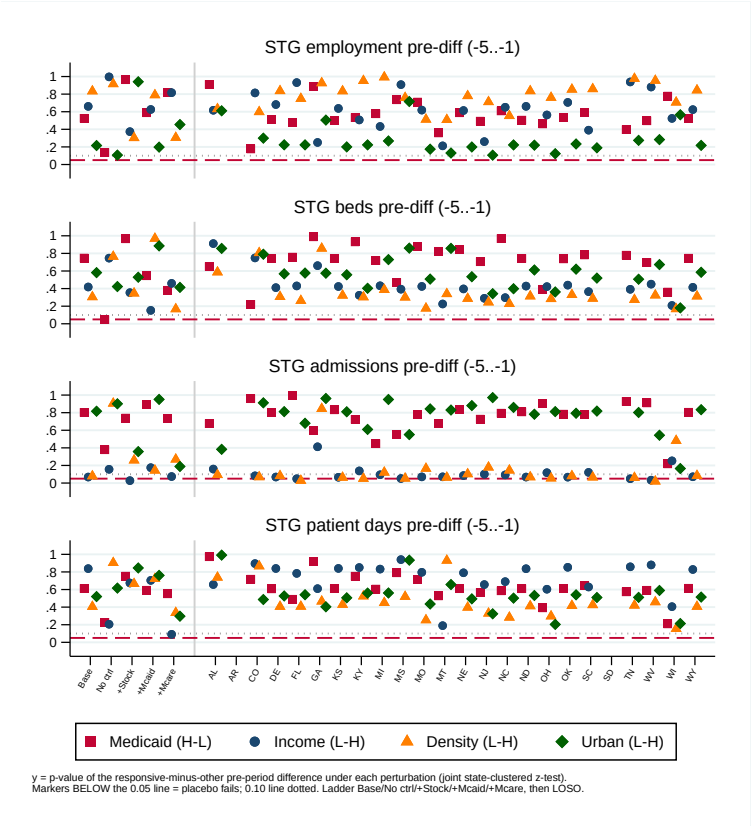
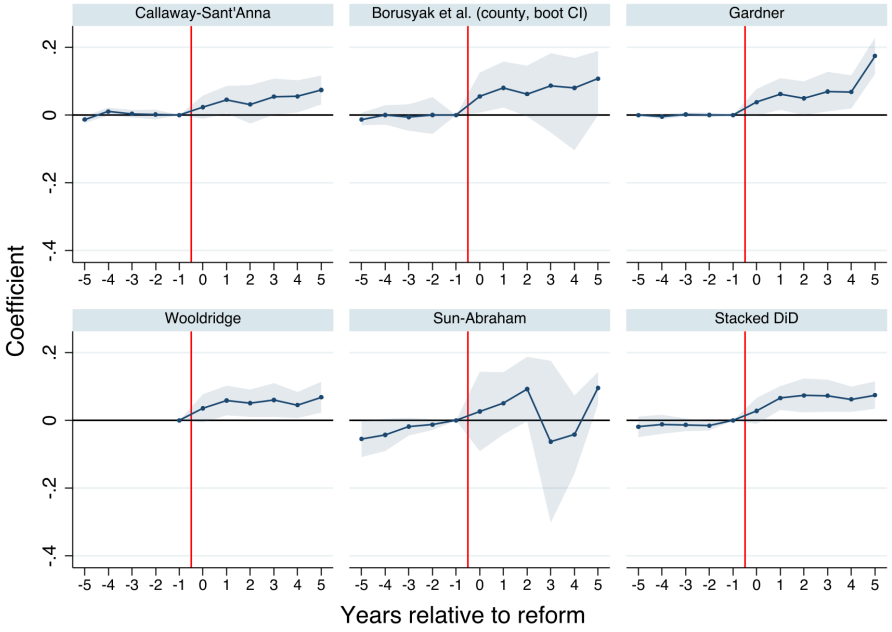
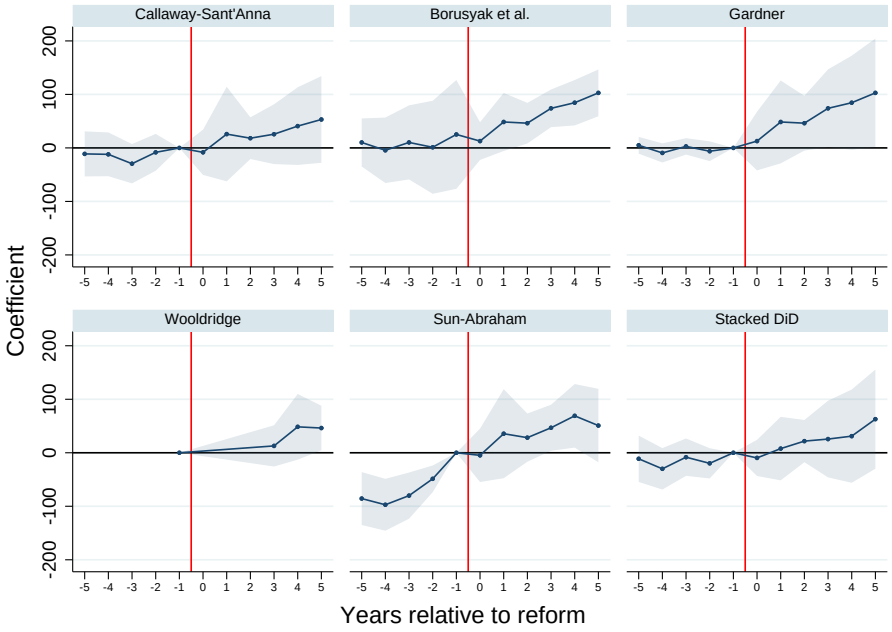


Figure A27: Pre-Treatment Divergence in Hospital Outcomes: Wald Test p-values

Notes: p-values from the Wald test of equality of hospital-outcome effects across county partitions, using a state-clustered covariance matrix that preserves covariance across subgroups.



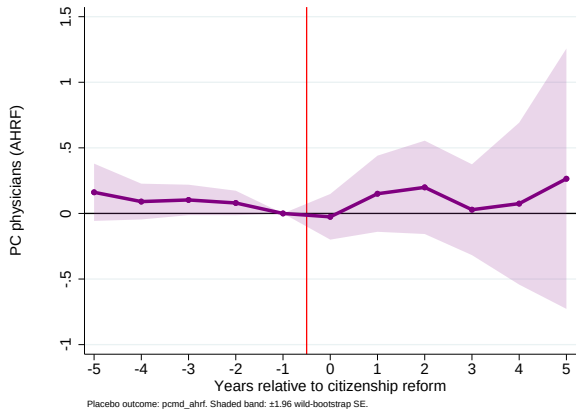
(a) FMGs Newly Licensed in Year t



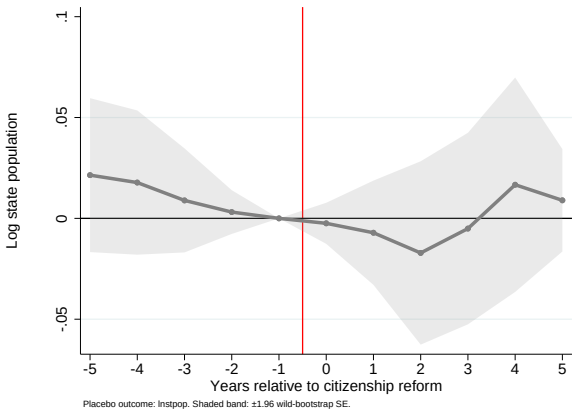
(b) Short-Term General Hospital Admissions

Figure A28: Comparison Across Staggered Difference-in-Differences Estimators

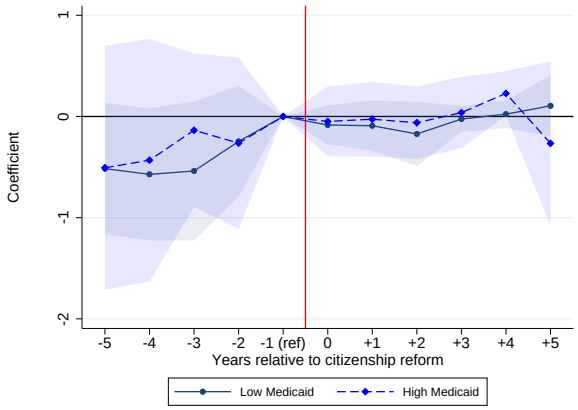
Notes: Each panel reports event-study estimates from a different heterogeneity-robust staggered difference-in-differences estimator. Shaded bands are 95% confidence intervals. *Callaway and Sant'Anna* and *Gardner* use the not-yet-treated states as the comparison group. *Wooldridge* does not report pre-treatment coefficients as when using only not-yet-treated groups, it imposes the parallel-trends assumption rather than testing it with placebo leads. *Sun and Abraham* requires a control cohort, which is specified as the latest-treated states (those reforming after 1974) because no never-treated states exist. *Stacked DiD* builds one sub-experiment per adoption cohort.



(a) Total Physicians per Capita (AHRF)



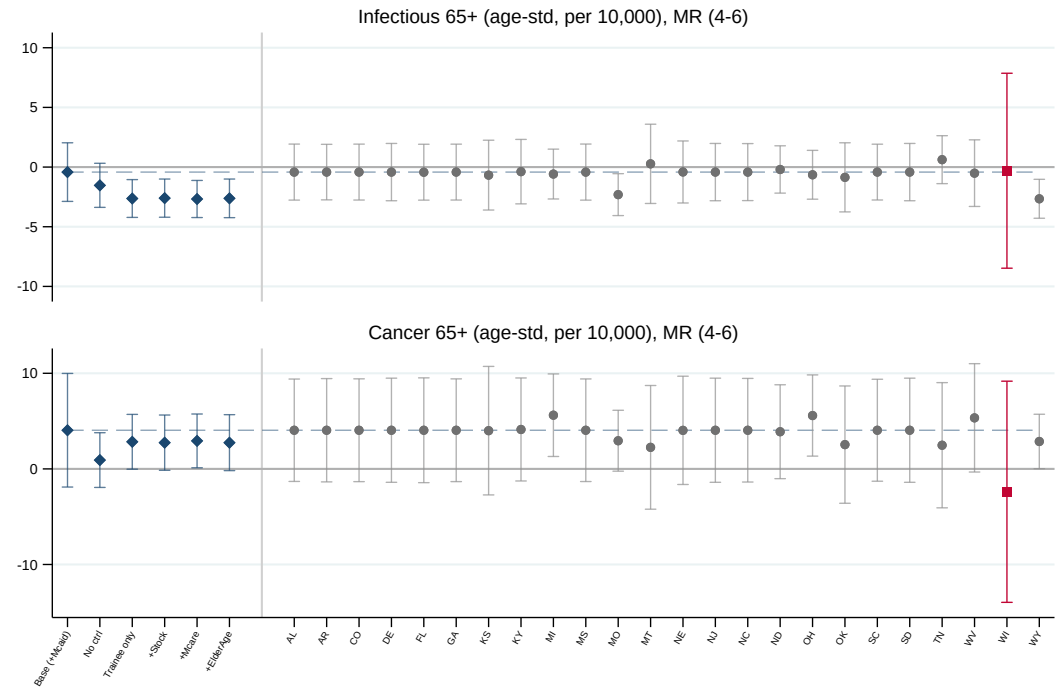
(b) Log State Population



(c) Log Cumulative Hill-Burton Funding (Nonprofit)

Figure A29: Placebo Outcomes: Total Physicians, State Population, and Hill-Burton Funding

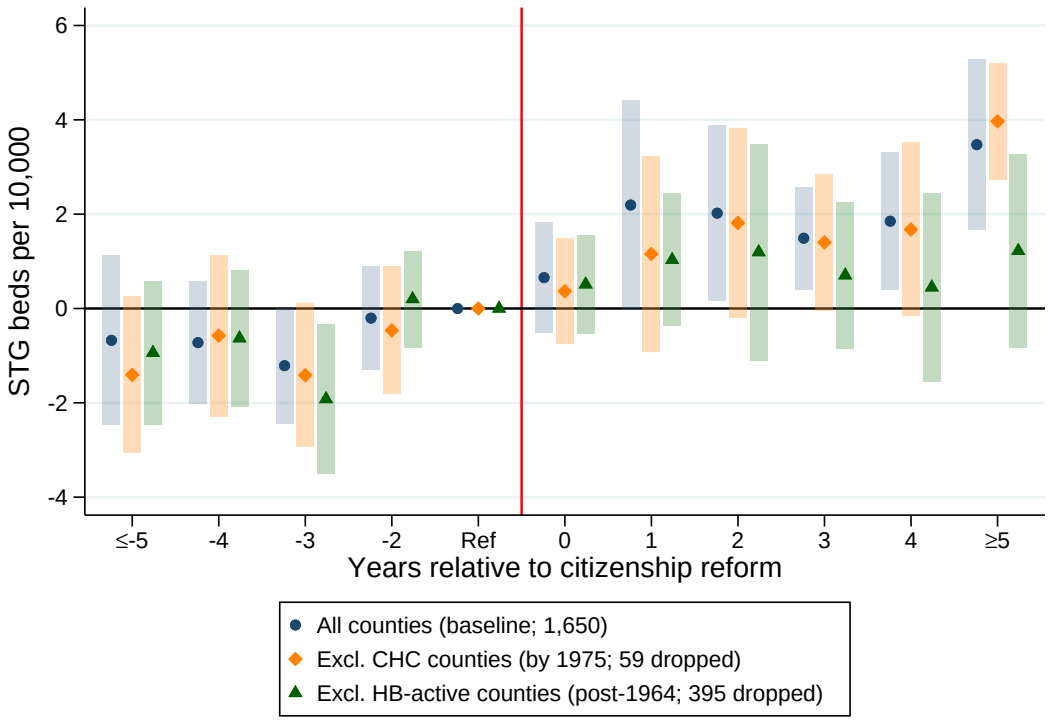
Notes: Panel (a) plots the event-study estimates using per capita total physician count from the Area Health Resource File (AHRF) as a placebo outcome. Panel (b) plots the estimates using log state population as a placebo outcome. Panel (c) plots log cumulative Hill-Burton funding for nonprofit hospitals. None of the series should respond to the removal of the citizenship requirement if the identification is valid. Point estimates near zero with overlapping confidence intervals across all pre- and post-reform periods support the parallel trends assumption and indicate that the estimates are not driven by spurious population dynamics.



Diamonds: control specs (wboot CIs). Circles: LOSO under the +Medicaid spec (analytic CIs); WI in red.
BASELINE = +Medicaid (c.rtfmg_trfb i.mcaid_post); dashed line anchored there. Rungs 3-6 = trainee + listed covariate WITHOUT Medicaid.
Medicaid control zeroes g1970(1973-75)/g1973 cells - estimates lean on 1965-69 cohorts.

Figure A30: Robustness of the Mortality Estimates

Notes: Robustness checks for the by-cause mortality estimates in Figure A8.



State aggregates rebuilt from county AHA data (22-state panel; AR/SD unavailable); rates over included counties' pop. Orange drops the 59 counties with a CHC by 1975 (~24% of pop); green drops the 395 counties with post-1964 Hill-Burton funding (~53% of pop). Base spec (c.rtfmg_trb, [iw=stpopb]), annual clock 1959-75, wboot 95% CI; end bins pool e≤-5 / e≥5.

Figure A31: Short-Term General Hospital Outcomes under CHC and Hill-Burton County Exclusions

Notes: Event-study estimates for short-term general hospital outcomes re-estimated under two sample restrictions: excluding the 40 counties that received a Community Health Center by 1975 (Bailey and Goodman-Bacon, 2015), and excluding the 199 counties with Hill-Burton funding receipt during 1964–1971 (Chung et al., 2017). The baseline estimate on the full sample is shown for comparison.