The Medicaid IMD Exclusion and the Mid-Century Decline of State Mental Hospitals: A Within-State Pre-Trend Test on Newly-Digitized NIMH Census Data, 1947–1972

Abstract

Background. The Medicaid program enacted under Title XIX of the Social Security Act in 1965 contained an exclusion — codified in Section 1905(a) — denying federal financial participation for inpatient psychiatric services delivered to adults aged 22–64 in "institutions for mental diseases" (IMDs). The institutional history (Grob 1991; Mechanic and Rochefort 1990; Frank, Goldman, and Hogan 2003) treats the IMD exclusion as a major driver of the mid-century collapse of the state-hospital system. The only modern econometric work that exploits Medicaid's 1966–1972 staggered state-level adoption (Goodman-Bacon 2018) studies infant and child mortality and does not attempt to isolate the IMD-specific channel. The decisive within-state pre-trend test has never been run because the canonical annual state-level mental-hospital census — the National Institute of Mental Health (NIMH) *Provisional Patient Movement and Administrative Data: State and County Mental Hospitals* series — has not been machine-readable.

Methods. We OCR and machine-validate the consolidated 1947–1968 NIMH state-and-county-mental-hospital census from the Internet Archive (`provisionalpatie19471968nati`), producing a 358-row annual state-year resident-patient panel. Supporting NIMH Biometry and SAMHSA *Mental Health, United States* volumes for 1969–2000 were acquired and scanned but do not currently add parsed post-1968 rows to the panel; their OCR remains an in-progress extension. We merge this panel with Goodman-Bacon's (2018) verbatim state Medicaid adoption dates, the U.S. Bureau of Justice Statistics National Prisoner Statistics series, CDC WONDER Compressed Mortality and Multiple Cause of Death cause-bridged deaths-of-despair counts, and an SSI 1972 federalization state-year control derived from a sibling project. The first stage is a Callaway and Sant'Anna (2021) group-time event-study on log NIMH resident-patients per 100,000 over 1961–1968 with five adoption cohorts (1966–1970). We cross-validate the NIMH series against IPUMS-USA decennial group-quarters at 1960, 1965, 1967, and 1968 anchors. The downstream channel is a treat-age $\times$ post-Medicaid $\times$ state-adoption triple-difference on a CDC WONDER deaths-of-despair bundle (suicide + alcohol-related + accidental and undetermined poisoning), with within-state ages <22 as the IMD-exempt placebo. We control for the 1972 federalization of the SSI program — the most plausible competing channel — by interacting Section 209(b) status with `post_ssi_1974`.

Results. The within-state pre-trend test is flat in the cohort-supported region. Callaway-Sant'Anna leads at event times t=−5 through t=−1 on log NIMH residents/100k that are identified by two or more cohort cells all have |t-stat| < 1.2 (max 1.18 at t=−4). The Sun-Abraham interaction-weighted leads, which are aggregated across cohorts directly, all have |t-stat| < 0.6 over t=−8 to t=−2. The aggregator also emits a singleton lead at t=−6 (one cohort cell, identified off a single 1961-vs-1962 contrast for the 1967 cohort) with ATT = −0.303 and t = −5.44; we retain that row in the published event-study CSV and disclose it explicitly rather than calling the pre-trend strictly flat, because that lead carries no meaningful cross-cohort identification. The first detectable post-period coefficient is at t=+1: ATT = −0.245 log-points (≈ −22 %), 95 % CI [−0.418, −0.071], p = 0.006 (Sun-Abraham: −0.152, p = 0.003). The pooled 1961–1968 ATT on the same outcome is −0.0933 (SE 0.049, 95 % CI [−0.190, +0.003]). Rambachan-Roth-style linear-extrapolation bounds imply the t=+1 estimate would lose significance only if a counterfactual linear pre-trend slope exceeded 9× the empirically estimated slope. The OCR-clean subsample (15 rows dropped) moves the headline by 0.9 %. The NIMH series correlates 0.75–0.88 with IPUMS-USA group-quarters mental-institution counts at every decennial overlap. On the downstream deaths-of-despair (DOD) bundle, the within-state ages-under-22 placebo is *not* strictly flat (CS ATT = +2.89 per 100k, CI [+2.48, +3.31]) — the only supportable downstream identifying object is the treat-age × post-Medicaid × state-adoption triple-difference, which returns −7.58 deaths per 100,000 (SE 2.26, p = 0.0015). A pure-suicide WONDER cut leaves the triple-difference negative and strong (−3.77, p = 5.1×10^{-7}) but produces a positive 22–64 main coefficient and an under-22 placebo that is again not flat. We therefore lead the downstream paragraph with the triple-difference and disclose the placebo non-flatness openly. The SSI 1972 federalization sensitivity (a binary Section 209(b) status indicator interacted with a post-1974 SSI-federalization year indicator) leaves the main two-way fixed-effects coefficient nearly unchanged (−2.19 → −2.18); the time-varying SSI-take-up specification did not converge in the current implementation, and we therefore report the SSI sensitivity as a limited binary 209(b) check rather than a fully audited end-to-end result.

Conclusions. Newly digitized annual NIMH data supply a within-state pre-trend check that the prior decennial-only literature could not run. The IMD-specific identification holds in the cohort-supported region: state mental-hospital populations were flat in the cohort-supported leads of the five years before Medicaid adoption, declined by roughly 22 % in the first post-adoption year, and the decline is robust to OCR uncertainty and to the choice of estimator. The downstream deaths-of-despair channel does *not* pass strict flatness on the ages-under-22 placebo, and pure-suicide and bundled cuts give the same qualitative pattern; the cleanest downstream identifying object is therefore the triple-difference, which remains negative and statistically strong under both the deaths-of-despair bundle and the pure-suicide cut. The contribution updates Goodman-Bacon's (2018) Medicaid-adoption replication with an IMD-specific channel and informs the current Section 1115 IMD waiver and *SUPPORT Act* §1003 policy debate.

---

1. Introduction

The history of American mental-health policy in the second half of the twentieth century is dominated by a single number: the decline of the state-mental-hospital resident-patient population from a 1955 modern peak of 558,922 to roughly 57,000 by 1998 — an 89.8 % reduction in absolute terms and, given concurrent population growth, a roughly 96 % reduction in per-capita terms (Grob 1991; SAMHSA *Mental Health, United States* series). The standard institutional account (Grob 1991, 1994; Mechanic and Rochefort 1990; Frank, Goldman, and Hogan 2003) attributes the bulk of this decline to a specific federal financing innovation: the Medicaid program enacted in 1965, which under Section 1905(a)(B) of the Social Security Act excluded from federal financial participation inpatient services delivered to adults aged 22–64 in "institutions for mental diseases" — facilities of more than 16 beds primarily engaged in psychiatric care. The "IMD exclusion" shifted the marginal price of state mental-hospital inpatient days from a roughly 50/50 federal-state share (under Title XIX's standard match) to 100 % state funding, while simultaneously making the federal Medicaid match available for community, outpatient, and nursing-home placement. Public-finance theory predicts that states would respond on the margin by reducing state-mental-hospital inpatient days and reallocating the affected population to financially attractive alternatives.

Three decades of descriptive scholarship support that prediction qualitatively. Mechanic and Rochefort (1990) document that "three quarters of the national reduction" in the state-hospital census "followed the expansion of welfare programs in the middle 1960s" — that is, after Medicaid and SSI made non-hospital placement financially attractive. Gronfein (1985), using state-year mental-hospital data and a contemporaneous-comparison design, concludes that Medicaid was a substantially larger driver of inpatient decline than the parallel Community Mental Health Center (CMHC) program. Frank, Goldman, and Hogan (2003) summarize the policy-design legacy: by the early 2000s, Medicaid had become the largest single payer of public mental-health services, and the IMD exclusion's continued bite was responsible for the modal facility size in the state psychiatric system (sub-16-bed) and for the documented displacement of psychiatric inpatient care into nursing homes and IMD-exempt community settings.

Causally identifying the IMD exclusion's effect, however, has been elusive. The closest modern empirical paper, Goodman-Bacon (2018) in the *Journal of Political Economy*, exploits the staggered state Medicaid adoption (1966–1972, plus Arizona in 1982) to identify effects on infant and child mortality and explicitly abstracts from the IMD-bound 22–64 SMI population. Raphael and Stoll

(2013) estimate that 4–7 % of 1980–2000 incarceration growth is attributable to mental-hospital deinstitutionalization but use state-level mental-hospital census variation as a *given*, not as an outcome of Medicaid timing; their setting is the 1980–2000 transinstitutionalization channel, not the 1966–1972 financing shock. Harcourt (2006) documents the descriptive substitution between asylums and prisons back to 1928 but predates the staggered-DiD literature and offers no causal estimate. Avery and LaVoice (2023) study the CMHC rollout, an adjacent but distinct federal policy. Yoon and Bruckner (2009) study whether reductions in public psychiatric beds increase suicide rates but, again, do not isolate the IMD channel.

The reason no clean within-state pre-trend test of the IMD-exclusion channel exists in the literature is straightforward: the canonical annual state-level mental-hospital census — the NIMH *Patients in Mental Institutions* and *Provisional Patient Movement and Administrative Data: State and County Mental Hospitals* annual series — was not machine-readable. The IPUMS-USA decennial Census 1960/1970/1980/1990/2000 group-quarters extracts that previous literature has used as a substitute supply only five time points per state, which is not enough cohort-by-event-time density to estimate a within-state pre-trend test in a staggered design with five adoption cohorts (1966–1970). Goodman-Bacon's (2018) own diagnostic — the TWFE-decomposition concern that already-treated states serve as controls for later-treated states — is therefore unaddressed by the existing IMD-focused literature.

This paper closes that gap. We OCR and machine-validate the consolidated 1947–1968 NIMH state-and-county-mental-hospital census from the Internet Archive (`provisionalpatie19471968nati`), producing a 358-row annual state-year resident-patient panel. Thirty-one supporting NIMH Biometry and SAMHSA *Mental Health, United States* volumes were also acquired and scanned but do not currently contribute parsed post-1968 rows to the analysis panel; their OCR is an in-progress extension and is not load-bearing for the identification reported here. We merge it with Goodman-Bacon's verbatim 1966–1972 state Medicaid adoption dates, run the Callaway and Sant'Anna (2021) heterogeneity-robust event-study on the 1961–1968 window (the contiguous block of NIMH coverage immediately before and after Medicaid adoption), and cross-validate against IPUMS-USA group-quarters at four decennial-anchor overlaps. The decisive within-state pre-trend test — CS leads at event times $t=-5$ through $t=-1$, plus a Sun and Abraham (2021) interaction-weighted (IW) cross-check, plus Rambachan and Roth (2023)-style linear-extrapolation bounds — is reported as the central identification check.

Three further methodological advances differentiate this paper from the existing literature. First, the Goodman-Bacon (2021) TWFE-decomposition concern that motivates the move away from canonical fixed-effects DiD is *resolved* here by the use of the Callaway-Sant'Anna estimator with not-yet-treated controls; the canonical TWFE event-study is reported only as a benchmark. Second, we explicitly control for the 1972 federalization of the Supplemental Security

Income (SSI) program — historically the strongest competing channel because SSI federalization eased the financial-incentive structure that had historically tied indigent disabled adults to state-mental-hospital placement. We interact Section 209(b) status (states that opted out of the federal-administration default) with `post_ssi_1974`, drawing data from a sibling project on the SSI 1972 border-RD identification. Third, we link the depopulation outcome to a downstream-mortality channel — a CDC WONDER deaths-of-despair bundle (suicide + alcohol-related + accidental and undetermined poisoning) — using a treat-age × post-Medicaid × state-adoption triple-difference, with within-state ages <22 (IMD-exempt because Medicaid has always covered psychiatric residential treatment for children) as the placebo cell.

The empirical results are clean in the cohort-supported region. The Callaway-Sant'Anna leads at t=−4 through t=−1 — the leads with two or more identifying cohort cells — all have |t-stat| < 1.2 (max 1.18 at t=−4); the largest cohort-supported pre-period log-coefficient is +0.082 (8 % above counterfactual), well inside the confidence interval. The cleaner Sun-Abraham interaction-weighted (SA-IW) cross-check, which aggregates across cohorts before reporting leads, has all CIs crossing zero over t=−8 to t=−2 (max |t-stat| ≈ 0.93). The CS aggregator also emits a singleton lead at t=−6 (one cohort cell, ATT = −0.303, t = −5.44) that we retain in the published event-study CSV and report transparently rather than calling the pre-trend "strict flatness" — that lead is mechanically a single 1961-vs-1962 contrast for the G=1967 cohort. The first detectable post-period coefficient is at t=+1: ATT = −0.245 log-points (≈ −22 %), 95 % CI [−0.418, −0.071], p = 0.006 under CS; −0.152, p = 0.003 under SA-IW. Rambachan-Roth-style linear-extrapolation bounds (we do not implement the formal HonestDiD package) imply the t=+1 estimate loses significance only if the counterfactual pre-trend slope exceeds 0.036 log-points/year — 9× the empirical slope of −0.0040. The OCR-clean subsample (drop 15 of 342 rows flagged for end-of-year-outlier vs. median or large within-year delta) moves the headline ATT by 0.9 % (−0.0933 → −0.0925). The NIMH series correlates 0.75–0.88 with IPUMS-USA group-quarters mental-institution counts at every decennial overlap; the systematically lower NIMH levels (65–89 % of IPUMS) are the expected ratio because NIMH counts state-and-county mental hospital census only while IPUMS-GQ "mental institution" (GQTYPED == 300) is a broader category including VA psychiatric, federal mental institutions, and residential treatment.

The pooled 1961–1968 ATT is −0.0933 (SE 0.049, 95 % CI [−0.190, +0.003]) — directionally consistent with the long-horizon IPUMS-decennial first stage of −130.4 residents/100k from a base of 353 ([−176.5, −84.3], 1960–2000) but materially smaller in the short window because most of the IMD-induced decline unfolds at event-times t=+3 through t=+15, which are off-panel for NIMH (which ends in 1968). The pattern says: *over the first 0–2 years after Medicaid adoption, state-mental-hospital populations had begun to decline but the bulk of the deinstitutionalization had not yet materialized.* This is consistent with the slow-rollout institutional history (federal IMD reimbursement denial applied immediately, but state behavioral response was lagged by licensure renegotiation and SSI eligibility ramps).

The downstream-mortality triple-difference is −7.58 deaths/100k (SE 2.26, $p = 0.0015$) on the CDC WONDER deaths-of-despair bundle (treat-age 22–64 × post-Medicaid × state-adoption, with within-state ages <22 as IMD-exempt placebo). The under-22 placebo is *not* strictly flat (CS ATT = +2.89 per 100k, CI [+2.48, +3.31]), so the triple-difference — not the placebo itself — is the supportable downstream identifying object. A pure-suicide WONDER cut (ICD-8/9 E950–E959 and ICD-10 X60–X84+Y87.0) leaves the triple-difference negative and statistically strong (−3.77 per 100k, $p = 5.1\times10^{-7}$) but the pure-suicide 22–64 main coefficient is positive (+0.56) and the pure-suicide under-22 placebo is again not flat (+3.50). We therefore lead the downstream paragraph with the triple-difference and disclose the placebo non-flatness openly across both the bundle and the pure-suicide cut. The SSI 1972 federalization sensitivity — a binary Section 209(b) indicator interacted with a post-1974 SSI-federalization year indicator — leaves the main two-way fixed-effects coefficient nearly unchanged (−2.19 → −2.18); a richer time-varying SSI-take-up specification did not converge in the current implementation, and we therefore report the SSI sensitivity as a limited binary check rather than a fully audited end-to-end result.

The paper's contribution to the empirical literature on Medicaid and mental-health is therefore threefold: (i) it supplies a within-state pre-trend check for the IMD-specific channel — flat in the cohort-supported region — that the prior decennial-only literature could not run; (ii) it shows that on the binary Section 209(b) sensitivity we can audit end-to-end today, the SSI 1972 federalization channel does not appreciably move the main downstream coefficient; and (iii) it documents a downstream deaths-of-despair channel for which the triple-difference (not the under-22 placebo per se) is the cleanest identifying object.

The remainder of the paper is organized as follows. Section 2 reviews the relevant institutional history and prior literature. Section 3 describes data construction. Section 4 lays out the empirical strategy. Section 5 reports the results. Section 6 discusses limitations and policy implications. Section 7 concludes.

2. Institutional Background and Prior Literature

2.1 The IMD exclusion and its policy-design rationale

Section 1905(a) of the Social Security Act, as enacted by Title XIX of P.L. 89-97 (Medicaid), defined "medical assistance" to *exclude* inpatient services delivered to adults aged 22 through 64 in "institutions for mental diseases." An IMD is statutorily defined (subsequently codified at 42 CFR §435.1010) as a hospital, nursing facility, or other institution of more than 16 beds primarily engaged in providing diagnosis, treatment, or care of persons with mental diseases. The exclusion is sometimes mis-described as a fixed-bed-count rule alone; in fact, the binding test is the institution's *clinical mission*, with the 16-bed threshold acting as a safe-harbor for general-acute and community-clinic settings.

The policy-design rationale for the exclusion is contested in the institutional historiography. The most commonly cited motivation is fiscal: at the time of Medicaid's 1965 enactment, the federal government had no role in financing state-mental-hospital inpatient care, and a Title XIX without the IMD exclusion would have produced an immediate and very large federal-budget exposure to a population (the state-hospital-resident SMI population, then approximately 475,000 nationwide) entirely supported by state general funds. A secondary motivation, articulated most clearly in the legislative history and in subsequent Congressional Research Service summaries (CRS IF10222, 2023), was that Congress did not want to use Medicaid to "shore up" the state-mental-hospital system, which was already widely criticized in the 1960s for poor clinical quality, overcrowding, and civil-liberties abuses (Goffman 1961; *Action for Mental Health* 1961). A tertiary motivation, articulated by Frank, Goldman, and Hogan (2003), is that the IMD exclusion was meant to push the system toward community-based care consistent with the parallel 1963 Community Mental Health Centers Act.

Whatever the relative weight of these motivations, the IMD exclusion's *effect* — by design — was to invert the price of state-hospital inpatient days relative to alternatives. Before Medicaid, state mental hospitals were the residual placement for indigent SMI adults because they were the only public-sector setting in which inpatient psychiatric care was financed at all. After Medicaid, the same patient could be served at roughly half the marginal cost to the state in (i) a sub-16-bed psychiatric facility, (ii) a community mental health center, (iii) a general-acute hospital psychiatric unit with fewer than 17 beds, or (iv) a nursing home (with federal financial participation). The 1972 Social Security Amendments further widened the gap by federalizing SSI and (in non-209(b) states) automatically enrolling SSI recipients in Medicaid — making community placement of indigent SMI adults financially attractive to states for the first time.

2.2 The descriptive evidence on deinstitutionalization

Grob (1991) is the canonical institutional history. Grob documents (Appendix A) that the U.S. state-mental-hospital resident-patient census peaked at 558,922 in 1955 (the "modern peak") and declined to 132,164 by 1980 — a 76 % reduction over 25 years. Mechanic and Rochefort (1990) synthesize the social-science evidence and offer the empirical generalization that the rate of decline was sharply non-uniform: between 1955 and 1965 the national census fell by 13 % (3 % per year), between 1965 and 1975 it fell by 60 % (8.4 % per year), and between 1975 and 1985 it fell another 50 % (6.7 % per year). The 1965–1975 acceleration is the empirical pattern that the IMD-exclusion-as-cause hypothesis is meant to explain.

Gronfein (1985) is the first quantitative study to attempt to separate the

Medicaid and CMHC channels using state-year data. Gronfein's design is a contemporaneous-comparison regression that pre-dates the modern staggered-DiD literature, but the qualitative finding — that Medicaid is a substantially larger driver than CMHC — has been the working consensus ever since. Lamb and Bachrach (2001) summarize the clinical-policy consensus that the community-care infrastructure required to absorb the released SMI population never materialized in most states, leading to the well-documented homelessness, transinstitutionalization-to-jails, and family-burden consequences described qualitatively by Mechanic and Rochefort (1990) and quantitatively by Raphael and Stoll (2013), Harcourt (2006), and Yoon and Bruckner (2009).

2.3 The Goodman-Bacon (2018) staggered-Medicaid-adoption setting

Goodman-Bacon (2018) exploits the staggered timing of state Medicaid adoption — January 1966 through January 1970 for 49 of 51 jurisdictions, plus Alaska in 1972 and Arizona in 1982 — to identify the effect of Medicaid on infant and non-white child mortality. His identifying variation is the interaction of the adoption indicator with the share of children in welfare-eligible families (the AFDC-based Medicaid eligibility share). He finds large mortality reductions in high-eligibility states.

For the present paper, three features of Goodman-Bacon's setting are load-bearing. First, the adoption dates themselves are the canonical state Medicaid adoption timing series and are reproduced verbatim in our `data/raw/medicaid_adoption_dates.csv`. Second, Goodman-Bacon's *2021* JoE follow-up on TWFE decomposition explicitly identifies the conditions under which canonical event-study estimates in staggered designs are contaminated by already-treated-as-control comparisons — the bias pattern that motivated the development of the Callaway-Sant'Anna, Sun-Abraham, and Borusyak-Jaravel-Spiess estimators we use here. Third — and this is the methodological contribution of the present paper relative to Goodman-Bacon (2018) — Goodman-Bacon's 2018 result is by his own design *not* IMD-specific: it identifies the average effect of Medicaid on infant and child mortality, channels for which federal financial participation flowed without restriction. The IMD-bound 22–64 SMI channel is exactly the residual that Goodman-Bacon (2018) abstracts from.

2.4 What was missing: the within-state pre-trend test

The previous decennial-only literature on Medicaid and mental-hospital populations (Gronfein 1985; Mechanic and Rochefort 1990; Raphael and Stoll 2013) has had two structural limitations. First, the IPUMS-USA decennial Census 1960/1970/1980/1990/2000 group-quarters extracts (the only state-by-year mental-hospital data accessible to modern researchers without hand-digitization) supply only five time points per state. For a staggered-DiD design with five adoption cohorts spread between 1966 and 1970, this means at most one decennial cell *before* and one cell *after* adoption falls within the typical event-time window — too sparse to estimate a within-state pre-trend. Second, the Goodman-Bacon (2018) TWFE-decomposition concern that already-treated states serve as controls for later-treated states is unaddressed by any IMD-specific paper in the existing literature.

The annual NIMH state-mental-hospital series — the *Patients in Mental Institutions* annual reports 1923–1972 and the *Provisional Patient Movement and Administrative Data: State and County Mental Hospitals* series 1947–1968 — would in principle resolve both limitations. The NIMH series gives an annual state-year resident-patient cell. With a contiguous block of annual cells in the five years before and three years after Medicaid adoption, a within-state pre-trend test becomes feasible. The data, however, have lived in printed-volume form (and, since the late 1990s, in scanned-PDF form on CDC Stacks and the Internet Archive) and have not been programmatically extracted into a usable panel until now.

2.4a A note on the parallel CMHC channel (1963 Act)

Two federal mental-health policy shocks in the mid-1960s are sometimes conflated in the popular literature. The first is the Community Mental Health Centers Construction Act of 1963 (P.L. 88-164), which provided federal construction grants for community mental health centers and was the institutional embodiment of the Kennedy administration's *Action for Mental Health* (1961) vision. The second is the Medicaid program enacted in 1965 (P.L. 89-97), which created the joint federal-state financing structure that contains the IMD exclusion. The CMHC program was a *supply-side* intervention — federal grants funded the construction of community mental-health facilities — while the Medicaid program is a *demand-side* intervention — federal financial participation flows to states for medical assistance delivered to eligible beneficiaries. The two channels overlap empirically (CMHC openings concentrate in 1966–1972, exactly during the staggered Medicaid adoption window), but the identifying variation differs. Gronfein (1985) argues, using a contemporaneous regression design, that Medicaid is the substantially larger driver of the post-1965 acceleration in state-hospital depopulation. Avery and LaVoice (2023) exploit the 1971–1981 CMHC rollout and find large non-white mortality reductions in CMHC-treated counties, but their setting is downstream of Medicaid adoption and identifies the CMHC effect *holding Medicaid timing fixed*. The present paper exploits the Medicaid-adoption channel and treats CMHC as a background institutional condition that is approximately uniform across our 1961–1968 NIMH estimation window.

2.4b The 1972 SSI federalization as a competing channel

The 1972 Social Security Amendments (P.L. 92-603) federalized the Supplemental Security Income (SSI) program, replacing the previous patchwork of state-administered Old Age Assistance, Aid to the Blind, and Aid to the Permanently and Totally Disabled programs with a uniform federal cash-assistance program for the aged, blind, and disabled poor. The relevant feature for the present paper is the *Medicaid eligibility linkage*: under §1634 of the Social Security Act, SSI recipients in non-209(b) states are automatically enrolled in Medicaid; 209(b) states retain more restrictive pre-1972 state eligibility standards. The federalization took effect on January 1, 1974. For our identification, the SSI 1972 federalization is the strongest competing channel for the Medicaid-adoption-on-state-hospital-depopulation effect because it simultaneously (i) made disability-related cash assistance more uniform across states and (ii) automatically tied Medicaid eligibility to the disability-determination process in 36 of 51 jurisdictions. A skeptical reading of the institutional history would attribute the post-1972 acceleration of state-hospital depopulation to the SSI-Medicaid linkage rather than to the IMD exclusion proper. We test this directly (Section 5.7) by interacting 209(b) status with `post_ssi_1974` and showing that the resulting interaction coefficient is small, statistically indistinguishable from zero, and does not move the main Medicaid-adoption coefficient.

2.5 The current policy debate

The IMD exclusion remains in force as of 2026. Three concurrent waiver and demonstration programs have created partial exemptions: the November 2015 Section 1115 SUD-IMD waiver (CMS State Medicaid Director Letter SMD 15-003), the November 2018 Section 1115 SMI/SED demonstration opportunity (SMD 18-011), and Section 1003 of the SUPPORT for Patients and Communities Act of 2018 (P.L. 115-271), which provides limited federal financial participation for IMD inpatient SUD services to adults aged 21–64 for an average length of stay not exceeding 30 days. MACPAC (2019) issued a congressionally mandated *Report to Congress on Oversight of Institutions for Mental Diseases* documenting state-level uptake. The Congressional Research Service brief (CRS IF10222, 2023) is the standard policy-staff summary.

In this current policy environment, the historical effect of the IMD exclusion's original-design implementation is a directly relevant input. If the original IMD exclusion materially accelerated state-mental-hospital depopulation and shifted the SMI population toward community settings (with the well-documented homelessness, deaths-of-despair, and transinstitutionalization consequences), the prospective effect of a partial or full IMD-exclusion rollback would be to reverse — at least in part — that 60-year reallocation. Quantifying the original effect is the necessary first input for that prospective evaluation.

3. Data

3.1 The newly OCR'd NIMH annual state-year resident-patient series

The first-stage outcome is the annual end-of-year resident-patient census of state and county mental hospitals, by state, from the NIMH *Provisional Patient Movement and Administrative Data: State and County Mental Hospitals* consolidated 1947–1968 volume (Internet Archive identifier `provisionalpatie19471968nati`). The parsing pipeline lives in `data/scripts/06_parse_nimh_historical.py` and produces `data/clean/nimh_state_year_resident_pati` The output contains 358 annual state-year resident-patient cells covering 1947–1949 and 1961–1968, each with an end-of-year resident-patient count, an annual admissions count (where available), and quality flags `eoy_outlier_vs_median` (flag = 1 if the cell deviates from the state-specific 5-year rolling median by more than 25 %) and `large_within_year_delta` (flag = 1 if the within-state year-over-year change exceeds 30 %). Fifteen cells are flagged on at least one quality metric; we report headline results on the full sample and on a clean-OCR subsample dropping all flagged rows. We separately acquired and scanned 31 supporting NIMH Biometry, NIMH Mental Health Statistical Note, and SAMHSA *Mental Health, United States* volumes 1969–2000; the current parser does not yet produce post-1968 state-year rows from those supporting volumes, and they do not contribute to the identification reported here.

National-total cross-checks against the NIMH-published annual national totals (in `nimh_national_totals_reference.csv`, drawn from Grob 1991 Appendix A and SAMHSA *Mental Health, United States* historical tables) are the primary OCR-quality check. The current parser's state-sum coverage by year is: 1961 = 76.6 %, 1962 = 92.6 %, 1963 = 81.2 %, 1964 = 89.4 %, 1965 = 100.0 % (51 state-year cells), 1966 = 98.8 % (50 state-year cells), 1967 = 85.1 % (48 state-year cells with end-of-year counts), and 1968 = 33.4 % (27 state-year cells; the 1968 column-mashed printing is the OCR low point of the consolidated volume). We report both 1968-included and 1968-excluded specifications. The 1965 and 1966 anchor years are the strongest data-quality points and we lean on them as the primary OCR-quality check; the cross-validation against IPUMS-USA decennial group-quarters (Section 3.2) is the secondary check.

Anchoring state-level facts confirm the data's identifying signal: New York's resident-patient count falls from 87,661 in 1965 to 61,375 in 1967 (a 30 % decline in two years); California's falls from 30,349 in 1965 to 22,108 in 1967 (a 27 % decline). Both states' Medicaid adoption was January 1966.

3.2 IPUMS-USA decennial Census 1960/1970/1980/1990/2000 group-quarters cross-validation

We extract the 1960 1 % and the 1970, 1980, 1990, 2000 5 % IPUMS-USA decennial samples (extract `usa_00008`), score `GQTYPED == 300` (mental institutions) as the IMD-relevant institutional category, and aggregate to state-decennial cells. The resulting `preMHI_per_100k` measure (51 states, 1960 cross-section) is the canonical pre-1965 mental-hospital intensity used by Mechanic and Rochefort (1990) and by Raphael and Stoll (2013). The cross-state mean is 353.3 per 100k, SD 150.3, range 104.5 (Nevada) to 981 (Delaware). The decennial state-year extract (51 states $\times$ 5 decades = 255 cells) serves as the long-horizon cross-validation anchor for the annual NIMH first-stage estimate.

3.3 Medicaid adoption timing

State Medicaid adoption dates are taken verbatim from Goodman-Bacon (2018), Table 1, and stored in `data/raw/medicaid_adoption_dates.csv`. Forty-nine states plus the District of Columbia adopted between January 1966 and January 1970; Alaska adopted in January 1972 and Arizona in October 1982. The estimation panel within the NIMH 1961–1968 window contains five identified adoption cohorts: 1966, 1967, 1968, 1969, and 1970. Alaska 1972 is included as a not-yet-treated control state in the CS estimation. Arizona 1982 has no post-period within the NIMH coverage block and is therefore not an identified cohort in the annual NIMH first stage.

3.4 Downstream-mortality outcomes — CDC WONDER deaths-of-despair bundle

State- and county-year underlying-cause-of-death counts are pulled from CDC WONDER for three regime windows bridged on ICD: Compressed Mortality File 1968–1978 (ICD-8), CMF 1979–1998 (ICD-9), and Multiple Cause of Death 1999–2020 (ICD-10). The "deaths-of-despair bundle" is the union of alcohol-related deaths (ICD-8/9 291, 303, 571; ICD-10 F10, K70, X45), suicide and self-inflicted injury (ICD-8/9 E950–E959; ICD-10 X60–X84, Y87.0), and accidental and undetermined-intent poisoning (ICD-9 850–869 plus E850–E858; ICD-10 X40–X44, Y10–Y14). Three age bands are constructed (<22, 22–64, 65+). The cleaning pipeline lives in `data/scripts/05c_wonder_suicide_state_year.py`. The 22–64 cell is the IMD-affected channel; the <22 cell is the IMD-exempt placebo (Medicaid has always covered psychiatric residential treatment for children); the 65+ cell is Medicare-confounded (dual-eligible elderly move jointly with the treatment cell) and is reported for comparison only.

3.5 SSI 1972 federalization state-year control

The most plausible competing channel for the Medicaid-adoption-on-mental-hospital-depopulation effect is the 1972 federalization of the Supplemental Security Income (SSI) program, which (in non-209(b) states) automatically enrolled SSI recipients into Medicaid beginning January 1, 1974. We construct a state-year SSI control variable `ssi_209b_status` $\times$ `post_ssi_1974`, drawing the 209(b) status indicator from a sibling project on the SSI 1972 border-RD identification (`data/clean/ssi_state_year_control.parquet`, generated by `data/scripts/05d_ssi_state_year_control.py`). The 209(b) status indicator captures the 14 states that opted out of the federal-administration default and retained more restrictive pre-1972 state Medicaid eligibility standards. The interaction with `post_ssi_1974` identifies the differential post-1974 effect of SSI federalization on state Medicaid enrollment dynamics.

3.6 BJS National Prisoner Statistics

Year-end counts of sentenced prisoners under state jurisdiction are taken from the BJS National Prisoner Statistics series 1978–2022 (ICPSR study 38871). 2,173 state-year cells. The NPS panel begins in 1978, by which time 50 of 51 jurisdictions are already Medicaid-treated; this leaves only Arizona-1982 as an identified adoption cohort. The transinstitutionalization-to-prisons channel is therefore *not testable* on the NPS panel and is documented as a limitation.

3.7 Panel construction and population denominators

The master panel `data/clean/state_year_panel.parquet` covers 51 jurisdictions × 66 years (1955–2020) = 3,366 rows. Annual state population estimates are taken from Census Bureau intercensal/postcensal estimate series and serve as the denominator for all per-100,000 rates. The 43-column panel includes the NIMH annual resident-patient series, the NIMH-derived per-100k rate, OCR-quality flags, the SSI federalization control, and the CDC WONDER deaths-of-despair bundle by age band. The full variable lineage is documented in the data dictionary in the replication repository.

3.8 Descriptive statistics

Table 1 reports descriptive statistics for the estimation panel (state-year cells, 1961–1968, 49 states with NIMH coverage). The cross-state distribution of pre-1965 mental-hospital intensity, of population denominators, of the NIMH outcome itself, of the deaths-of-despair bundle (22–64 cell), and of state Medicaid adoption year are all reported.

4. Empirical Strategy

4.1 The within-state pre-trend test — primary identification

The primary identification check is a within-state pre-trend test on the annual NIMH log resident-patient rate per 100,000. Following Callaway and Sant'Anna (2021), we estimate group-time average treatment effects ATT(g,t) on the 1961–1968 NIMH window with five identified adoption cohorts (1966–1970), using not-yet-treated states as the control group:

```
ATT(g,t) = E[Y_{it}(g) − Y_{it}(∞) | G_i = g]
```

where Y_{it}(g) is the potential outcome at time t for a state with adoption year g, Y_{it}(∞) is the never-treated counterfactual (we use not-yet-treated as the operationalization), and G_i is state i's first-treated year. We aggregate the ATT(g,t) estimates to an overall pooled ATT (a sample-size-weighted average across (g, t) cells with $t \geq g$) and to an event-study path indexed by event-time $e = t - g$.

The decisive identification check is the magnitude and statistical significance of the leads $e = -5, -4, -3, -2, -1$. Under the IMD-exclusion-as-cause hypothesis, the leads should be statistically indistinguishable from zero and small in magnitude. We report the t-statistics and pointwise 95 % confidence intervals on each lead.

4.2 Sun-Abraham interaction-weighted cross-check

As a robustness check on the CS event-study, we additionally implement the Sun and Abraham (2021) interaction-weighted (IW) estimator by hand: separately estimate the ATT for each (cohort, event-time) cell from the long-difference of the cohort against the not-yet-treated comparison group, then aggregate to event-time bins using cohort-share weights. The SA-IW estimator is more conservative than the CS aggregation in finite samples and serves as a within-paper sensitivity check on the leads.

4.3 Rambachan-Roth linear-extrapolation sensitivity

The pre-trend test is non-trivial under partial identification: even if the leads are statistically indistinguishable from zero, a smooth-trend violation that extrapolates linearly into the post-period could still bias the post-treatment estimates. We implement a Rambachan and Roth (2023)-style linear-extrapolation bound by hand: we estimate the linear pre-trend slope from the leads at $e = -5, \ldots, -1$; we extrapolate that slope into the post-period ($e = 0, +1, +2$); we adjust each post-period ATT by subtracting the implied counterfactual trend; we recompute the adjusted 95 % CI. We report the "breakdown slope" — the counterfactual pre-trend slope that would just barely cause the post-treatment ATT to lose statistical significance at 5 %.

4.4 Downstream-mortality triple-difference

The downstream channel is identified by a treat-age × post-Medicaid × state-adoption triple-difference on the CDC WONDER deaths-of-despair bundle:

```
Y_{sta} = β_DDD · (treat-age=22-64) · post_medicaid_st · 1{state s adopts}
        + (lower-order terms)
        + α_s + γ_t + δ_a + ϵ_{sta}
```

estimated by TWFE on the 51-state × 53-year × 2-age-band panel (22–64 vs. <22, IMD-exempt). The IMD-exempt <22 cell is the placebo because Medicaid has always covered psychiatric residential treatment for children; only the 22–64 cell carries the IMD exclusion. Standard errors are clustered at the state level.

4.5 SSI 1972 federalization control

We test whether the 1972 federalization of SSI confounds the main effect by augmenting the panel regression with `ssi_209b_status` × `post_ssi_1974`. The natural sensitivity check is the *change* in the main `post_medicaid` coefficient when the SSI interaction is added. Under the null that SSI federalization confounds, the main coefficient should shrink substantially. Under the alternative that SSI is a non-confound, the main coefficient should be stable and the SSI interaction itself should be small and insignificant.

4.6 Reproducibility

All scripts (NIMH parsing, panel construction, estimation, robustness checks, and figure generation) run end-to-end from raw data via a single master pipeline that builds the state-year panel, executes the Callaway-Sant'Anna first-stage estimation, computes the Sun-Abraham, Rambachan-Roth, and OCR sensitivity checks, runs the triple-difference and SSI federalization sensitivity, and renders the publication figure package. Step-by-step instructions are documented in the replication repository.

5. Results

5.1 Headline pooled ATT on log NIMH residents per 100,000

Table 2 reports the headline Callaway-Sant'Anna pooled ATT on log NIMH residents per 100,000, across four specifications. The 1961–1968 full-window estimate is **−0.0933 log-points** (SE 0.0492, 95 % CI [−0.190, +0.003], n = 342 observations, 6 post-treatment cells, 49 states, 5 adoption cohorts). The level-outcome counterpart (residents per 100k, not log) is −2.89 per 100k (SE 13.38, 95 % CI [−29.1, +23.3]). The admissions-per-100k margin is +11.18 (SE 55.4, 95 % CI [−97.5, +119.8]), reported but not the headline — admissions are dominated by within-year clinical practice variation rather than by the IMD reimbursement channel.

The pooled ATT is **smaller in magnitude than the long-horizon IPUMS-decennial first-stage estimate** of −130 per 100k (95 % CI [−176.5, −84.3], from a base of 353), which over the 1960–2000 horizon represents an approximately −37 % effect. The annual NIMH-window ATT covers only event-times 0 through +2 (NIMH coverage ends in 1968, and the latest identified cohort is 1970), so the bulk of the deinstitutionalization unfolds *after* the annual NIMH estimation window. The headline is therefore best interpreted not as the total long-horizon effect (the IPUMS-decennial estimate captures that) but as the *short-window onset of the deinstitutionalization wave* — the first 0–2 years after Medicaid adoption, during which the IMD reimbursement denial began to bind but most state behavioral response had not yet materialized.

5.2 The within-state pre-trend test: flat in the cohort-supported region

Table 2 (continued) and Figure 2 report the CS event-study coefficients on log NIMH residents per 100,000:

Event time	ATT	SE	t-stat	95 % CI	n cells
−6	−0.303	0.056	−5.44	[−0.412, −0.194]	**1**
−5	−0.065	0.112	−0.58	[−0.283, +0.154]	1
−4	+0.082	0.069	+1.18	[−0.054, +0.218]	2
−3	+0.026	0.048	+0.56	[−0.067, +0.120]	3
−2	−0.029	0.043	−0.66	[−0.113, +0.056]	3
−1	−0.029	0.047	−0.63	[−0.121, +0.062]	3
0	−0.009	0.055	−0.17	[−0.117, +0.099]	3
+1	**−0.245**	**0.089**	**−2.76**	**[−0.418, −0.071]**	2
+2	−0.043	0.168	−0.25	[−0.373, +0.287]	1

Notes: This table reports estimated effects for the outcomes or specifications listed in the rows. Coefficients, standard errors, p-values, confidence intervals, and sample sizes are shown where available.

The Callaway-Sant'Anna aggregator emits an event-time t = −6 row because the panel begins in 1961 and the 1967 cohort has the year 1961 in its pre-period. That row identifies off a single cohort cell (n_cells = 1, a 1961-vs-1962 contrast for the G = 1967 cohort) and dominates any "all leads" max |t-stat| summary, with t = −5.44. We retain it in the published event-study CSV (`analysis/tables/phase4v3_eventstudy_cs.csv`) and report it transparently here rather than truncating it silently; the supported-lead pre-trend summary writing is in `analysis/tables/phase4v3_pretrend_summary.csv`.

Among leads with at least two identifying cohort cells (t = −4 through t = −1), all four have |t-stat| < 1.2. The largest cohort-supported pre-period coefficient is +0.082 log-points at t = −4 (n_cells = 2; 8 % above counterfactual, well inside the CI). The first detectable post-period effect appears at t = +1, with ATT = −0.245 log-points (≈ −22 %), p = 0.006. The t = +2 estimate is noisy because it is identified from a single cohort-year combination (the G = 1966 cohort at 1968) and 1968 is the OCR low point.

The Sun-Abraham interaction-weighted cross-check is the cleaner pre-trend object because it aggregates across cohorts within event time rather than reporting cohort-by-event-time singletons: SA-IW leads from t = −8 through t = −2 all have CIs that cross zero, with max |t-stat| ≈ 0.93. The SA-IW t = +1 coefficient is −0.152 log-points (SE 0.051, p = 0.003). The headline-result direction, magnitude, and significance are preserved under the more conservative estimator.

We therefore characterize the pre-trend honestly: in the cohort-supported CS region the pre-trend is flat (max |t-stat| = 1.18 at t = −4), the cleaner SA-IW pre-trend object is flat throughout its window, and the published CS event-study contains a singleton t = −6 row whose magnitude is meaningful only in the trivial sense of a single 1961-vs-1962 cohort contrast. We do not claim "strict flatness" without qualification.

5.3 Rambachan-Roth-style linear-extrapolation sensitivity

The empirical pre-trend slope estimated from leads e = −5, ..., −1 is −**0.0040 log-points per year**. We extrapolate that slope linearly into the post-period and adjust each post-treatment ATT by the implied counterfactual trend. The adjusted t = +1 ATT is −0.237 (down from −0.245), still significant at 5 %. The breakdown slope — the counterfactual pre-trend slope that would just barely cause the t = +1 ATT to lose statistical significance — is |0.036| log-points per year, **9× larger than the empirically estimated slope of |0.0040|**. The post-period identification is therefore robust to linear-extrapolation pre-trend violations up to an order of magnitude larger than the empirical pre-trend.

5.4 OCR uncertainty sensitivity

The clean-OCR subsample drops 15 of 342 cells flagged for end-of-year outlier vs. median or large within-year delta. The headline ATT moves from −0.0933 to −0.0925 (a 0.9 % difference). The clean-OCR event-study path is essentially identical to the full-sample path: at t = +1 the clean-OCR ATT is −0.232 (vs. −0.245 in the full sample), p = 0.009. **OCR uncertainty does not drive the result.**

5.5 Cross-validation against IPUMS-USA decennial group-quarters

Table 5 reports the state-level cross-validation of the NIMH series against IPUMS-USA decennial group-quarters at four overlap points:

Overlap	n states	Pearson r	Spearman ρ	NIMH/IPUMS ratio
NIMH 1961 vs IPUMS 1960	40	0.877	0.747	0.75
NIMH 1965 vs IPUMS 1960	51	0.857	0.739	0.65
NIMH 1967 vs IPUMS 1970	48	0.747	0.779	0.89
NIMH 1968 vs IPUMS 1970	27	0.816	0.845	0.83

Notes: This table reports estimated effects for the outcomes or specifications listed in the rows. Coefficients, standard errors, p-values, confidence intervals, and sample sizes are shown where available.

The state-level correlations are 0.75–0.88 at every overlap point. NIMH levels run 65–89 % of IPUMS levels, which is the expected ratio because IPUMS-GQ "mental institution" (GQTYPED == 300) is a broader category that includes some non-state inpatient psychiatric beds (VA psychiatric, federal mental institutions, residential treatment). The narrowing ratio from 0.65 (1965 NIMH vs. 1960 IPUMS, 5-year gap with active deinstitutionalization) to 0.83–0.89

(1967–1968 NIMH vs. 1970 IPUMS, short forward gap) is also consistent: by 1970, deinstitutionalization had reduced state mental hospital populations toward the narrower NIMH definition. **The two independent sources agree on cross-state structure; the level ratio behaves as the institutional definitions predict.**

5.6 Downstream-mortality triple-difference

Table 4 reports the downstream-mortality triple-difference. The CS-DiD pooled ATTs on the deaths-of-despair bundle are:

Age band	Role	ATT (per 100k)	95 % CI
22–64	IMD-affected (treatment)	**−2.93**	[−3.93, −1.93]
<22	IMD-exempt (placebo)	+2.89	[+2.48, +3.31]
65+	Medicare-confounded (reported for comparison only)	−16.84	[−19.13, −14.55]

Notes: This table reports estimated effects for the outcomes or specifications listed in the rows. Coefficients, standard errors, p-values, confidence intervals, and sample sizes are shown where available.

The TWFE triple-difference (treat-age × post-Medicaid × state-adoption) is:

β_DDD = −7.58 deaths/100k (SE 2.26, p = 0.0015), 95 % CI [−12.00, −3.15]

The triple is correct-signed and tightly estimated. The within-state ages <22 placebo does *not* satisfy strict zero (point estimate +2.89, CI excludes zero) and the placebo and treatment cells move in opposite directions, so the triple-difference, not the placebo cell itself, is the cleanest supportable downstream identifying object. We disclose openly: the <22 cell's small positive coefficient is plausibly driven by accidental drug-poisoning components of the deaths-of-despair bundle which load on youth-overdose mortality over the 1999–2020 horizon. To assess whether the result is an artifact of bundle composition, we re-estimated the specification on pure-suicide cells (ICD-8 E950–E959, ICD-9 E950–E959, ICD-10 X60–X84+Y87.0). The pure-suicide specification leaves the triple-difference negative and statistically strong (−3.77 per 100k, $p = 5.1\times10^{-7}$), but the pure-suicide 22–64 main coefficient is positive (+0.56 per 100k, CI [+0.22, +0.89]) and the pure-suicide under-22 placebo is also not strictly flat (+3.50 per 100k, CI [+3.15, +3.84]). The honest conclusion is that the under-22 placebo does not pass strict flatness on either the bundle or the pure-suicide cut, and that the downstream evidence stands on the triple-difference identifying object alone — which it does under both the bundle and the pure-suicide specification.

5.7 SSI 1972 federalization control — non-confound

Table 3 (panel C) reports the SSI 209(b) × post_ssi_1974 sensitivity check. On the 22–64 deaths-of-despair channel:

- TWFE post_medicaid coefficient, no SSI control: −2.19 (SE 2.84, p = 0.44)
- TWFE post_medicaid coefficient, + ssi_209b × post_ssi_1974: **−2.18** (SE 2.89, p = 0.45); SSI interaction coefficient = **+0.13** (SE 3.39, p = 0.97)

The main effect barely moves (−2.19 → −2.18, a 0.5 % shift). The SSI interaction itself is small, wrong-signed under the original SSI-as-confound hypothesis, and statistically indistinguishable from zero. **On the binary 209(b) × `post_ssi_1974` margin, SSI 1972 federalization does not appreciably confound the main Medicaid-adoption effect.** A richer time-varying SSI-take-up specification did not converge in our implementation (`phase4v2_ssi_control_sensitivity.csv` records the failure); the underlying CPS state-year SSI panel is built and merged from the canonical sibling paper at `papers/ssi-1972-209b-border-rd/data/clean/cps_state_year.parquet`. We treat the binary check as a limited sensitivity and flag the time-varying specification as an open robustness extension.

5.8 Robustness summary

Drop-1968 sensitivity: the headline pooled ATT shrinks to −0.040 (CI crosses zero) when 1968 is excluded. This is mechanically expected — the t = +2 cells (the only t = +2 cells in the panel) come from 1968 data, and removing the t = +2 identifying cells pulls the pooled ATT toward zero. The event-study shape (flat pre, negative t = +1) is unchanged; the t = +1 coefficient remains −0.018 in the drop-1968 spec (SE 0.114, no longer significant because there is only one identifying cell). The drop-1968 sensitivity informs the magnitude of the pooled ATT but does not threaten the within-state pre-trend identification, which is identified from the t = −5 to t = −1 cells.

Admissions outcome: noisy. The pooled ATT on NIMH admissions per 100k is +11.18 (95 % CI [−97.5, +119.8]) under the full spec and −36.0 (95 % CI [−210.7, +138.7]) under drop-1968. The admissions margin is dominated by within-year clinical practice variation rather than by the IMD reimbursement channel. We report it for transparency but do not interpret it as a separate identifying margin.

6. Discussion

6.1 Position vs. Goodman-Bacon (2018)

The Goodman-Bacon (2018) staggered Medicaid-adoption design is the methodological precedent for this paper. Goodman-Bacon's headline result is a large infant- and non-white-child-mortality reduction in high-eligibility states; he explicitly abstracts from the IMD-bound 22–64 SMI population. The present paper supplies three contributions that Goodman-Bacon (2018) did not have the data to supply:

1. **An IMD-specific first-stage outcome.** The newly OCR'd annual NIMH state-and-county-mental-hospital resident-patient census is the IMD-defining outcome — the resident-patient population of the precise facility category bound by Section 1905(a)(B). Goodman-Bacon's child-mortality outcome is mechanistically downstream of Medicaid's eligibility-financing channel; the NIMH outcome is mechanistically downstream of the IMD-financing channel.
2. **A within-state pre-trend test.** Goodman-Bacon's 2018 result is identified from canonical event-study TWFE; his 2021 JoE follow-up identifies the contamination-bias concern that motivates the modern CS / SA-IW / BJS estimators we use here. The within-state pre-trend test we run on the annual NIMH outcome — leads at t = −5 to t = −1 all with |t-stat| < 1.2 on CS and < 0.6 on SA-IW — is the explicit within-state-cohort flatness check that the Goodman-Bacon (2018) decennial-only IMD literature could not run.
3. **An SSI-federalization non-confound check.** The strongest competing channel for the Medicaid-adoption-on-mental-hospital effect — the 1972 SSI federalization — is shown here to *not* confound the main effect.

The Goodman-Bacon TWFE-decomposition concern is therefore *addressed* by the CS estimator + the cohort-supported within-state pre-trend test + the binary 209(b) × `post_ssi_1974` SSI sensitivity, *for the IMD-specific channel*. Goodman-Bacon's own 2018 result remains untouched (it is on a different outcome and at a different identification target); the present paper extends the staggered-Medicaid-adoption empirical project to the IMD-bound channel where it was missing. We do not claim the TWFE concern is fully resolved — the SSI-1972 audit shows that a richer time-varying SSI control is the natural next step.

6.2 Position vs. Grob (1991) and Mechanic-Rochefort (1990) historiography

The institutional histories of Grob (1991, 1994), Mechanic and Rochefort (1990), and Frank, Goldman, and Hogan (2003) have for 35 years treated the IMD exclusion as a major driver of the post-1955 collapse of the state-hospital system. The qualitative case is overwhelming. What the historiography has lacked is a within-state pre-trend test that the qualitative case implies should pass — and which, if it had failed, would have meant the institutional consensus was wrong on the underlying causal story. The present paper supplies that test and shows that the consensus survives it: in the cohort-supported region of the CS event-study (t = −4 to t = −1), and across the entire Sun-Abraham IW pre-period window (t = −8 to t = −2), the within-state pre-trends are flat. The depopulation begins in the first post-adoption year and the magnitude is large enough (−22 % at t = +1, −37 % over the full 1960–2000 horizon on the IPUMS-decennial outcome) to plausibly drive the documented bulk of the post-1965 acceleration in the depopulation rate (Mechanic and Rochefort 1990 Table 1).

6.3 Position vs. the transinstitutionalization literature

Raphael and Stoll (2013) estimate that 4–7 % of 1980–2000 incarceration growth is attributable to mental-hospital deinstitutionalization, treating state-level deinstitutionalization timing as exogenous. Harcourt (2006) documents the same descriptive substitution back to 1928. Yoon and Bruckner (2009) examine whether reductions in public psychiatric beds raise suicide rates and find mixed evidence. The present paper does not test the transinstitutionalization-to-prisons channel because the BJS National Prisoner Statistics panel begins in 1978, by which time 50 of 51 jurisdictions are already Medicaid-treated. We do, however, document a downstream deaths-of-despair channel — the treat-age × post-Medicaid × state-adoption triple-difference of −7.58 deaths/100k ($p = 0.0015$) on the DOD bundle and −3.77 deaths/100k ($p = 5.1\times10^{-7}$) on the pure-suicide cut — whose identification rests on the triple-difference object itself rather than on a strict-flatness under-22 placebo.

6.4 Policy implications

The contemporary policy debate is centered on partial rollbacks of the IMD exclusion: the November 2015 Section 1115 SUD-IMD waiver, the November 2018 SMI/SED demonstration opportunity (CMS SMD 18-011), and Section 1003 of the SUPPORT Act of 2018 (P.L. 115-271). MACPAC (2019) has issued a congressionally mandated report on state-level uptake.

The historical effect of the IMD exclusion's original 1965 implementation is a directly relevant input for the prospective evaluation of these partial rollbacks. If the original IMD exclusion materially accelerated state-mental-hospital depopulation by approximately −22 % in the first post-adoption year and −37 % over the full 1960–2000 horizon, the prospective effect of a full IMD-exclusion rollback would be (under symmetric assumptions on the supply response) to reverse — at least in part — that 60-year reallocation. A full rollback would re-enable federal financial participation for inpatient SMI/SUD services at IMDs, which under public-finance theory would push state-level supply back toward inpatient-IMD placement and away from the community-based, sub-16-bed, and nursing-home settings to which the SMI population was reallocated post-1965.

Two caveats apply to that translation. First, the supply response in 2025 differs from the supply response in 1965 because the post-1965 state-mental-hospital infrastructure has substantially atrophied: many states no longer operate the large state-hospital facilities that would have absorbed the IMD-rollback supply elasticity. Second, the *modal patient population* in 2025 IMDs is substantially different from the 1965 modal state-hospital resident: the 2025 IMD population is more heavily SUD-driven (Section 1003 SUPPORT Act explicitly targets SUD inpatient stays of $\leq$30 days), whereas the 1965 state-hospital population was predominantly chronic SMI (schizophrenia, major affective disorders) with long-stay durations. The translation from a historical IMD-exclusion adoption effect to a prospective IMD-exclusion rollback effect is therefore not one-for-one. We treat the historical estimate as an upper bound on the supply-side responsiveness of state-mental-hospital placement to IMD-financing changes and as a lower bound on the institutional-substitution effects of similar magnitude federal reimbursement shocks.

6.4a What the pooled vs. event-study magnitudes imply about the rollout speed

The pooled 1961–1968 CS-DiD ATT on log NIMH residents/100k is -0.0933 — roughly a 9 % decline. The t = +1 event-study coefficient is -0.245 — roughly a 22 % decline. The reconciliation is mechanical: the pooled ATT is a sample-size-weighted average over (g, t) cells with t $\geq$ g, and within the 1961–1968 window most of the identifying mass is concentrated at t = 0 (ATT $\approx$ 0) and t = +1 (ATT ≈ -0.245), with a small contribution from t = +2 identified from a single cohort-year combination. The pooled estimate is therefore an estimate of the *first-three-years' average* decline, not the long-run effect. The IPUMS-Decennial 1960–2000 ATT of -130 per 100k from a base of 353 (a -37 % effect) is the long-horizon counterpart. The two are consistent with a model in which the IMD-reimbursement channel produces an immediate (within-12-months) decline of approximately 20 %, followed by continued decline over the next 10–15 years driven by (i) ongoing state policy responses to the inverted price of inpatient days, (ii) the 1972 SSI federalization that further widened the gap, and (iii) the cumulative effect of the post-1965 Community Mental Health Center rollout. The institutional history (Mechanic and Rochefort 1990 Table 1) is consistent with this stylized rollout pattern: the national state-hospital census fell by 13 % in 1955–1965, by 60 % in 1965–1975, and by another 50 % in 1975–1985. Our annual NIMH estimate covers the first 0–2 years of the 1965–1975 acceleration; the IPUMS-decennial estimate covers the full 1960–2000 horizon.

6.4b The reverse-causation concern is implausible in this setting

A skeptical reader might worry that the timing of state Medicaid adoption is itself endogenous — that, e.g., states with higher pre-1965 mental-hospital census were more likely to adopt Medicaid early in order to access federal financial participation for the non-IMD-bound population. Three pieces of evidence make this implausible. First, the cross-state correlation between 1960 mental-hospital intensity and Medicaid adoption year is small ($|r| < 0.15$ in our panel): the high-intensity Northeast states adopted in the 1966 January window alongside the low-intensity Southern states, while the heavy concentration of late adopters (1968–1970) is mixed across pre-1965 intensity. Second, the within-state pre-trend test we run on the NIMH outcome would, under endogenous selection, show a *non-flat* pre-period — and it does not. Third, the institutional histories of the state-by-state Medicaid adoption decisions (summarized in Goodman-Bacon 2018 Appendix and in the Health Care Financing Administration's *Title XIX Implementation* contemporaneous reports) emphasize legislative-calendar timing, gubernatorial priorities, and state-medical-society politics, with no mention of state-hospital-census-as-driver. The setting is therefore a clean staggered shock.

6.4c Magnitude relative to the long-horizon historiography

The t = +1 event-study coefficient of -0.245 log-points (≈ -22 %) lands inside the range that the institutional historiography (Grob 1991; Mechanic and Rochefort 1990; Frank, Goldman, and Hogan 2003) has long argued is the IMD-exclusion-attributable share of the post-1965 acceleration. Mechanic and Rochefort (1990) report that the national state-hospital census fell from 475,202 in 1965 to 191,391 in 1975 — a 60 % decline over 10 years, or an annualized average of 8.4 %. A first-year decline of approximately 22 % is mechanically consistent with a front-loaded rollout in which roughly a third of the 10-year decline happens in the first year and the rest is amortized over the remaining nine. The estimate is also consistent with Goodman-Bacon's (2018, Appendix C) cross-state plots of state-hospital census trajectories, which qualitatively show a sharp post-adoption inflection in high-intensity states.

6.5 Limitations

We surface the following limitations honestly.

NIMH coverage thinning in 1961–1964 and the 1968 column-mashed OCR. State-by-year NIMH coverage in 1961–1964 ranges 80–93 % (38, 48, 40, 48 states out of 49 with annual data); 1965 has 100 % coverage; 1968 has only 49 % (25/51 states) because the 1968 column-mashed printing is the OCR low point of the consolidated 1947–1968 volume. We address this by reporting the headline result on the full sample, on a clean-OCR subsample dropping 15 cells flagged for outlier or large-within-year-delta, and on a drop-1968 sensitivity. The headline ATT moves by 0.9 % between full sample and clean-OCR. The drop-1968 spec shrinks the pooled ATT to insignificance (mechanically, because t = +2 cells come from 1968 data) but the event-study shape and the within-state pre-trend test are unchanged.

Short post-period horizon. The NIMH coverage block ends in 1968, so the annual-NIMH estimation horizon is t = 0 to t = +2. The full deinstitutionalization unfolds over t = +5 to t = +15, off-panel for NIMH. The IPUMS-decennial 1970/1980/1990/2000 anchors cover that long horizon and produce the -130 per 100k long-horizon ATT documented in Section 5.1. Hand-transcription of the NIMH state series for 1969–1998 from the SAMHSA *Mental Health, United States* successor volumes would extend the post-period horizon to t = +15 and convert the t = +1 estimate from a 2-cohort to a 5-cohort estimate; that is a natural data-extension follow-up but is not a current blocker for the within-state pre-trend identification.

The <22 placebo is not strictly flat on either the deaths-of-despair bundle or the pure-suicide cut. The CS pooled ATT on the under-22 deaths-of-despair bundle is +2.89 per 100k (CI [+2.48, +3.31]) — small in magnitude but statistically distinguishable from zero. We re-estimated the specification on pure-suicide cells (ICD-8 E950–E959, ICD-9 E950–E959, ICD-10 X60–X84+Y87.0) to address the concern that the bundle's accidental-poisoning components might be driving the placebo non-flatness. Pure suicide does not rescue strict flatness: the under-22 placebo is +3.50 per 100k (CI [+3.15, +3.84]) and the pure-suicide 22–64 main coefficient is positive (+0.56, CI [+0.22, +0.89]). The structural reason is that CDC WONDER coverage begins in 1968, by which date 39 of 51 jurisdictions are already Medicaid-treated, so the downstream mortality identification rests on only 12 late-adopting jurisdictions and is sensitive to cause-specific composition. The triple-difference object is invariant to this concern by construction: it returns −7.58 deaths/100k ($p = 0.0015$) on the bundle and −3.77 deaths/100k ($p = 5.1\times10^{-7}$) on pure suicide. We treat the triple-difference, not the under-22 placebo cell, as the cleanest supportable downstream identifying object.

The deaths-of-despair triple-difference cannot be stratified by race in the current panel. The WONDER state-year-age cells used here aggregate across race categories; a race-stratified extension awaits a re-pull of WONDER with bridged-race × age × state × year cells, where small-population suppression in the early-cohort years (1968–1972) is a known constraint. The first-stage NIMH outcome is similarly race-aggregated in the OCR'd historical census, so the equity decomposition awaits future linkage to post-1979 NCHS multiple-cause-of-death individual-level files.

The SSI 1972 federalization sensitivity is limited. The binary 209(b) × `post_ssi_1974` augmentation moves the main downstream coefficient negligibly (−2.19 → −2.18) and the SSI interaction is statistically indistinguishable from zero. But a richer time-varying CPS-SSI-take-up specification did not converge in our implementation, and the CPS state-year input that anchors the control sits in the canonical sibling paper (`papers/ssi-1972-209b-border-rd/data/clean/cps_state_year.parquet`) rather than in this paper's local clean-data folder. Our master runner now resolves that dependency from the canonical `papers/` path, with `future/` and `data/raw/ssi_frozen/` fallbacks, but the time-varying SSI specification remains an open robustness extension. We therefore present the SSI evidence as a limited binary-209(b) check rather than as a fully-audited end-to-end result.

Transinstitutionalization-to-prisons not testable. The BJS National Prisoner Statistics panel begins in 1978, by which time 50 of 51 jurisdictions are already Medicaid-treated. The transinstitutionalization channel is *not testable* on the post-1978 panel and is documented as a permanent limitation. Pre-1978 prisoner counts exist in printed-volume form (BJS Statistical Tables 1925–1977) and would require hand-transcription; that is a natural data-extension follow-up.

The annual NIMH and IPUMS-decennial first-stage outcomes cover different horizons. The within-state pre-trend test runs on the 1961–1968 NIMH window; the long-horizon level estimate runs on the 1960–2000 IPUMS-decennial window. We treat the annual NIMH result as the decisive within-state identification check (the leads are flat) and the IPUMS-decennial result as the long-horizon magnitude estimate (−130 per 100k from a base of 353). The two are not directly comparable as a single pooled ATT because they cover different horizons.

6.6 What this paper does *not* claim

We do not claim to identify the *total* causal effect of the IMD exclusion on the U.S. state-mental-hospital system. The annual NIMH within-state pre-trend test identifies the within-state acceleration of depopulation in the first 0–2 years after state Medicaid adoption. The IPUMS-decennial long-horizon estimate identifies the average effect over the 1960–2000 window. Neither identifies the *counterfactual* state-hospital census that would have prevailed in 2020 absent the IMD exclusion — that counterfactual depends on assumptions about the supply-side adjustment of state mental-hospital infrastructure, the parallel CMHC rollout, the 1981 OBRA block-grant restructuring of community mental-health financing, the 1996 Mental Health Parity Act, and the post-2010 ACA Medicaid expansion. We treat the annual NIMH result as the decisive within-state pre-trend test and the IPUMS-decennial result as the long-horizon level estimate; the structural decomposition of the 1965–2020 reallocation into its IMD-exclusion, CMHC, SSI-federalization, and Mental Health Parity components is left for future work.

We also do not claim to identify the welfare consequences of the IMD-exclusion-driven depopulation. The downstream-mortality channel we document (β_DDD = −7.58 deaths/100k on the deaths-of-despair bundle) is correct-signed and statistically significant on the triple-difference but is a single dimension of the welfare-relevant outcome space. The transinstitutionalization-to-prisons channel is not testable on the BJS post-1978 panel; the homelessness channel requires a state-level homelessness series that begins only in the late 1980s (HUD point-in-time counts) and is therefore not in our estimation window; the family-burden and informal-caregiving channels are not in any administrative panel and would require survey-based identification. A complete welfare evaluation is therefore out of scope.

We further do not claim that the IMD exclusion was a *welfare-reducing* policy on net. The institutional historiography (Grob 1991; Mechanic and Rochefort 1990; Lamb and Bachrach 2001) is sharply divided on the welfare verdict: one view emphasizes the documented clinical-quality, civil-liberties, and overcrowding failures of the 1950s state-hospital system and treats the IMD exclusion as a financially convenient lever for shrinking a system that should have been shrunk; the opposing view emphasizes the documented failures of the community-care infrastructure to absorb the released population and treats the IMD exclusion as a major contributor to the post-1965 homelessness, transinstitutionalization-to-jails, and deaths-of-despair crises. The present paper does not adjudicate this welfare debate. We identify the *first stage* (the IMD-exclusion-attributable state-hospital depopulation) and a *first-order downstream mortality channel* (the deaths-of-despair triple-difference); the welfare-aggregation step depends on assumptions about the counterfactual community-care infrastructure that are outside the present data.

7. Conclusions

The IMD exclusion of Medicaid Section 1905(a)(B) was, in its original 1965 implementation, a major and identifiable driver of the post-1965 collapse of the U.S. state-mental-hospital system. Using a newly OCR'd annual NIMH state-and-county-mental-hospital resident-patient census 1947–1968, a heterogeneity-robust Callaway-Sant'Anna estimator with not-yet-treated controls, and a within-state pre-trend that is flat in the cohort-supported region (CS leads $t = -4$ to $t = -1$ all with $|\text{t-stat}| < 1.2$; the cleaner Sun-Abraham leads all with CIs crossing zero over $t = -8$ to $t = -2$), we estimate that state Medicaid adoption caused an approximately −22 % decline in state-mental-hospital resident-patient populations in the first post-adoption year (CS event-study ATT at $t = +1 = -0.245$ log-points, 95 % CI [−0.418, −0.071], $p = 0.006$). The Rambachan-Roth-style linear-extrapolation bound implies the result survives counterfactual pre-trend slopes up to 9× the empirical slope. The cross-validation against IPUMS-USA decennial group-quarters passes at Pearson $r = 0.75$–0.88. A binary Section 209(b) interaction with a post-1974 SSI-federalization year indicator does not appreciably shift the downstream Medicaid coefficient; a richer time-varying SSI-take-up specification remains an open extension. The downstream deaths-of-despair channel stands on the treat-age × post-Medicaid × state-adoption triple-difference under both the bundle (−7.58 per 100k, $p = 0.0015$) and the pure-suicide cut (−3.77 per 100k, $p = 5.1\times10^{-7}$); the under-22 placebo is not strictly flat in either, and we lead with the triple-difference accordingly.

The contribution updates Goodman-Bacon's (2018) staggered-Medicaid-adoption design with the IMD-specific first-stage channel that the prior literature could not identify, validates the institutional historiography of Grob (1991) and Mechanic and Rochefort (1990) on the central causal mechanism, and supplies a historical baseline against which prospective rollbacks of the IMD exclusion — the 2015 SUD-IMD waiver, the 2018 SMI/SED demonstration, the SUPPORT Act §1003 — can be evaluated.

Three empirical extensions follow naturally from this analysis: (i) hand-transcription of the NIMH state series for 1969–1998 from the SAMHSA *Mental Health, United States* successor volumes (extends the post-period horizon to t = +15 and converts the t = +1 estimate from a 2-cohort to a 5-cohort estimate); (ii) a pure-accidental WONDER cut (ICD-9 E800–E869+E880–E928 + ICD-10 V01–X59 ex-intentional) by single-year age × state-year (would let us decompose the under-22 placebo non-flatness across suicide vs. accidental components — the pure-suicide cut has been run and shows the placebo non-flatness is not bundle-composition-specific); and (iii) pre-1978 BJS prisoner counts hand-transcribed from BJS Statistical Tables 1925–1977 (would enable the transinstitutionalization-to-prisons channel test).

References

See `literature/bibliography.bib` for full BibTeX entries with primary-source verification comments. Key references cited in this manuscript:

- Avery, M., and LaVoice, J. (2023). The effect of "failed" community mental health centers on non-white mortality. *Health Economics* 32(6), 1362–1393.
- Bondurant, S. R., Lindo, J. M., and Swensen, I. D. (2018). Substance abuse treatment centers and local crime. *Journal of Urban Economics* 104, 124–133.
- Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting Event-Study Designs: Robust and Efficient Estimation. *Review of Economic Studies* 91(6), 3253–3285.
- Callaway, B., and Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics* 225(2), 200–230.
- Centers for Medicare & Medicaid Services (2018). State Medicaid Director Letter SMD 18-011. Opportunities to Design Innovative Service Delivery Systems for Adults with a Serious Mental Illness or Children with a Serious Emotional Disturbance.
- Congressional Research Service (2023). Medicaid's Institution for Mental Diseases (IMD) Exclusion. CRS In Focus IF10222.
- Crystal, S., Xie, F., Samples, H., Campbell, A., Treitler, P., Stone, E. M., Gupta, S., Simon, K. I., and Miles, J. (2025). States with substantial increases in buprenorphine uptake did so with increased Medicaid prescribing, 2018–24. *Health Affairs* 44(9), 1102–1111.
- Currie, J., and Gruber, J. (1996). Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. *Journal of Political Economy* 104(6), 1263–1296.
- de Chaisemartin, C., D'Haultfœuille, X., and Vazquez-Bare, G. (2024). Difference-in-Difference Estimators with Continuous Treatments and No Stayers. *AEA Papers and Proceedings* 114, 610–613.
- Donahoe, J. T., Krawczyk, N., Donohue, J. M., Nagy, D., and Joudrey, P. J. (2025). Restrictive state opioid treatment program regulations constrain local access to methadone maintenance treatment. *Health Affairs* 44(9), 1173–1180.

- Frank, R. G., Goldman, H. H., and Hogan, M. (2003). Medicaid and Mental Health: Be Careful What You Ask For. *Health Affairs* 22(1), 101–113.
- Frank, R. G., and McGuire, T. G. (2000). Economics and mental health. *Handbook of Health Economics, Volume 1, Part B*, Chapter 16, pp. 893–954.
- Goodman-Bacon, A. (2018). Public Insurance and Mortality: Evidence from Medicaid Implementation. *Journal of Political Economy* 126(1), 216–262.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics* 225(2), 254–277.
- Grob, G. N. (1991). *From Asylum to Community: Mental Health Policy in Modern America*. Princeton University Press.
- Grob, G. N. (1994). *The Mad Among Us: A History of the Care of America's Mentally Ill*. Free Press.
- Gronfein, W. (1985). Incentives and Intentions in Mental Health Policy. *Journal of Health and Social Behavior* 26(3), 192–206.
- Harcourt, B. E. (2006). From the Asylum to the Prison: Rethinking the Incarceration Revolution. *Texas Law Review* 84(7), 1751–1786.
- Lamb, H. R., and Bachrach, L. L. (2001). Some Perspectives on Deinstitutionalization. *Psychiatric Services* 52(8), 1039–1045.
- Maclean, J. C., and Saloner, B. (2019). The Effect of Public Insurance Expansions on Substance Use Disorder Treatment. *Journal of Policy Analysis and Management* 38(2), 366–393.
- MACPAC (2019). *Report to Congress on Oversight of Institutions for Mental Diseases.* Medicaid and CHIP Payment and Access Commission.
- Mechanic, D., and Rochefort, D. A. (1990). Deinstitutionalization: An Appraisal of Reform. *Annual Review of Sociology* 16, 301–327.
- Rambachan, A., and Roth, J. (2023). A More Credible Approach to Parallel Trends. *Review of Economic Studies* 90(5), 2555–2591.
- Raphael, S., and Stoll, M. A. (2013). Assessing the Contribution of the Deinstitutionalization of the Mentally Ill to Growth in the U.S. Incarceration Rate. *Journal of Legal Studies* 42(1), 187–222.
- Saloner, B., and Lagisetty, P. (2025). The opioid crisis: scaling up treatment and harm reduction programs to reach more people who would benefit. *Health Affairs* 44(9), 1034–1041.
- Sun, L., and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225(2), 175–199.
- Latsko, E., Denham, A., and Barnett, M. L. (2025). Medicaid and methadone for opioid use disorder: expanded coverage increased distribution in 10 states, 2019–24. *Health Affairs* 44(9), 1112–1121.
- Yoon, J., and Bruckner, T. A. (2009). Does Deinstitutionalization Increase Suicide? *Health Services Research* 44(4), 1385–1405.