When Work Requirements Met the States: Two Case Studies of Section 1115 Medicaid Community Engagement Mandates in Arkansas and Georgia

Abstract

Background. Between 2018 and 2024, six states (Arkansas, Indiana, New Hampshire, Michigan, Utah, and Georgia) briefly required able-bodied adult Medicaid enrollees to document work or qualifying community engagement hours to retain coverage. Four of the six implementations were halted by federal court order or by COVID-era continuous-enrollment rules within months of launch. Only two states — **Arkansas (June 2018 – March 2019)** and **Georgia (July 2023 – December 2024 and ongoing)** — produced implementation windows long enough for ex-post evaluation. The two states adopted policies that are substantively opposite in design: Arkansas conditioned continued coverage on documented hours among an already-enrolled expansion population, while Georgia made hours a precondition for *entry* into a new partial expansion up to 100% of the federal poverty line. A pre-registered pooled-evaluation design treating "ever-implemented" as a single treatment is therefore inappropriate; pooling two policy-opposite states with one treated cluster each also inflates conventional inference under standard cluster-robust formulas.

Approach. I write the two implementations up as independent case studies and present a brief cross-comparison narrative. Each case study uses a single-treated-state difference-in-differences specification on state-year ACS, CPS, and BRFSS aggregates with policy-eligible adults as the within-state unit of observation, paired with eligibility-defined comparison groups. I report cluster-robust (CR1) standard errors as a benchmark, wild-cluster bootstrap p-values (Cameron, Gelbach, and Miller 2008) as the inferentially correct alternative given a single treated cluster, and Fisher randomization-inference p-values across all possible placebo state assignments. Across the multiple-outcome family I apply a Romano-Wolf step-down correction.

Findings — Arkansas (June 2018 – March 2019). State-year aggregates do detect a directionally consistent Medicaid coverage loss in the policy-eligible (ages 30–49, $\leq$100% FPL adult expansion) sub-population, with point estimates that are consistent in sign and roughly half the magnitude of the individual-panel estimates reported by Sommers et al. (2020, *Health Affairs* — the two-year-impact paper companion to Sommers et al. 2019, *New England Journal of Medicine*, on one-year impacts). All point estimates lose statistical significance once a wild-cluster bootstrap is applied (the smallest wild-bootstrap p across the outcome family is 0.088 for employment, p = 0.32 for uninsured, p = 0.47

for Medicaid coverage); Romano-Wolf step-down across the four-outcome family leaves no outcome significant at conventional levels. The Arkansas implementation period is therefore best read as **a descriptive screen consistent in sign and magnitude with the prior individual-panel literature on coverage loss, but underpowered for a formal causal claim from state-year aggregates alone**.

Findings — Georgia (July 2023 – December 2024). State-year aggregates show essentially flat Medicaid coverage, employment, and uninsurance rates among the policy-eligible 19–64 sub-population at ≤100% FPL through the first 18 months of Pathways to Coverage. The CR1 point estimate for uninsurance is −0.98 percentage points (CR1 $p < 0.001$), which would conventionally signal a coverage gain, but the wild-cluster bootstrap p is 0.18 and Fisher randomization p is 0.59. The result attenuates further when the comparison group is restricted to non-expansion states (the natural counterfactual for a conditional-entry partial expansion): the uninsurance coefficient drops to +0.31 pp (CR1 $p = 0.55$) under the seven-state non-expansion control pool. Pathways enrollment was approximately 8,000 in mid-2025 versus a projected 345,000 (GBPI 2025), consistent with the small-magnitude null. The Georgia implementation period is best read as **a descriptive null on aggregate population coverage, employment, and uninsurance, consistent with the low realized take-up of a conditional-entry program**.

Cross-comparison. Arkansas and Georgia are not comparable as a single policy episode. Arkansas's program threatened existing coverage; Georgia's program offers conditional new coverage. Arkansas's policy applied to an already-insured expansion population at scale; Georgia's policy applied to an uninsured below-FPL population that had to opt in and document hours to gain coverage. A reader interested in the population-coverage consequences of "Medicaid work requirements" would draw different inferences from the two implementations: Arkansas tells us about disenrollment and friction in an existing benefit; Georgia tells us about take-up of a conditional new benefit. Treating either as a clean test of *the* federal work-requirement design — or pooling them — would obscure rather than inform.

Implications. With only two state-level implementations long enough to evaluate and with policy designs that are substantively opposite, the United States does not yet have the empirical record to support a formal causal evaluation of "Medicaid work requirements" as a single policy. The case-study evidence is consistent with the existing individual-panel literature documenting frictional coverage loss in Arkansas and is consistent with the public-data evidence of very low Pathways take-up in Georgia. The 2025 federal reconciliation-law mandate (P.L. 119-21) will, if implemented, produce the first sufficiently large-N policy variation for a formal pooled evaluation; until then, descriptive case studies of individual state implementations are the honest empirical record.

---

1. Introduction

Medicaid work requirements — formally, "community engagement requirements" under Section 1115 demonstration authority — have been among the most contested Medicaid policy levers of the past decade. Between January 2018 and December 2024, thirteen states received CMS approval to condition adult Medicaid eligibility on documentation of work, job training, education, volunteer service, or caregiving hours. Four other states submitted applications that were never approved or were subsequently withdrawn under the Biden administration's February 2021 policy reversal. The legal, administrative, and political history of these waivers has been extensively documented in the policy literature (Rosenbaum 2018; Musumeci, Rudowitz, and Hinton 2020; MACPAC 2019).

The empirical record, however, is narrower than the policy record. Of the thirteen approved waivers, only six states ever moved from CMS approval to active enforcement. Of those six, four (Indiana, New Hampshire, Michigan, Utah) were halted by federal court order or by COVID-19 continuous-enrollment rules within ten weeks to ten months of launch — windows too short for ex-post evaluation using standard administrative outcome data. **Only two states produced implementation windows long enough to evaluate**:

- **Arkansas Works** required ages 30–49 in the adult expansion group at or below 100% of the federal poverty line (FPL) to document 80 hours per month of qualifying activity. The phase-in began June 1, 2018. Judge James Boasberg of the U.S. District Court for the District of Columbia vacated CMS's approval on March 27, 2019 in *Gresham v. Azar*, ending enforcement after ten months.
- **Georgia Pathways to Coverage** is a partial Medicaid expansion to 100% FPL conditioned at the entry stage on documenting 80 hours per month of qualifying activity. CMS approved the program in October 2020, the Biden administration rescinded the work-requirement piece in December 2021, a federal court reinstated it in August 2022, and Georgia launched the program July 1, 2023. As of December 2024, Pathways had been live for 18 months and remained the only active Medicaid work-requirement program in the country.

A natural research question is whether one can pool these two implementations into a single staggered difference-in-differences design and estimate an "ATT of Medicaid work requirements" using never-treated states as the comparison group. **I do not pursue that pooled design.** Two considerations argue against it:

1. **Few clusters inflate t-statistics.** With one treated cluster in each of two cohorts (and the bulk of identifying variation arriving from the GA cohort because of its longer post-period), conventional cluster-robust standard errors are downward-biased to a degree that conventional inference is not credible (MacKinnon and Webb 2018; Roodman et al. 2019). Wild-

cluster bootstrap p-values, which are valid in single-treated-cluster designs, are larger than CR1 p-values by factors of three to fifty across the outcome family. The conventional pooled inference understates uncertainty.

2. **The policies are substantively opposite.** Arkansas was a **disenrollment-eligibility-condition** policy: people who already had coverage faced new documentation requirements to keep it, and approximately 18,000 of 70,000 nominally subject Arkansans lost coverage during the 10-month implementation (Sommers et al. 2020). Georgia is an **entry-eligibility-condition** policy: a new partial-expansion benefit category was created and made available conditional on documented hours, and approximately 8,000 of a projected 345,000 eligible Georgians enrolled in the first two years (Georgia Budget and Policy Institute 2025). The estimand pooling the two policies is not a meaningful average causal effect of "Medicaid work requirements"; it is an average of two different causal parameters with opposite signs of expected policy intent, weighted by the lengths of the post-periods.

For both reasons I write the two implementations up as independent case studies. The remainder of the paper is structured as two independent case studies — one for Arkansas (Sections 4–5) and one for Georgia (Sections 6–7) — followed by a brief cross-comparison narrative (Section 8) and concluding observations (Section 9). Each case study has its own identification strategy, its own discussion of policy and institutional context, and its own honest treatment of inferential limits when there is one treated state.

The contributions of this paper, framed honestly, are three. First, I provide the first state-year aggregate replication of the Sommers et al. (2020) individual-panel finding of coverage loss in Arkansas, using a different data source (ACS) and a different unit of analysis (state-year aggregates), and I show that the aggregate point estimates are consistent in sign and roughly half the magnitude of the individual-panel estimates but do not survive small-cluster inference correction. Second, I provide a first descriptive evaluation of the first 18 months of Georgia Pathways using publicly available ACS, CPS, and BRFSS aggregates, finding essentially flat population coverage, employment, and uninsurance rates among the policy-eligible sub-population — a result that is consistent with the very low realized take-up reported by the Georgia Budget and Policy Institute. Third, I argue methodologically that the existing state-level variation in Medicaid work requirements is insufficient for a credible pooled causal evaluation, and that the appropriate empirical posture until the 2025 federal mandate produces sufficiently large-N variation is descriptive case studies of individual state implementations.

The remainder of the paper is organized as follows. Section 2 reviews the prior literature on Medicaid work requirements and situates the case studies. Section 3 describes the data and the treatment-timing panel. Sections 4 and 5 present the Arkansas case study (background and findings). Sections 6 and 7 present the Georgia case study. Section 8 offers the cross-comparison narrative. Section

9 discusses policy implications and limitations and concludes.

---

2. Prior Literature

The Medicaid work requirements literature comprises three relatively distinct strands. The first, oldest, and most influential strand is the **individual-panel evaluation** of the Arkansas Works implementation. Sommers and coauthors (2019; 2020) used an internet-based panel of approximately 1,700 working-age adults in Arkansas and three comparison states (Kentucky, Louisiana, and Texas) to estimate the causal effect of Arkansas Works on coverage, employment, and self-reported access to care. Their headline 2020 *New England Journal of Medicine* result is a 6.8 percentage point reduction in Medicaid coverage and a 4.0 percentage point increase in uninsurance among Arkansans aged 30–49 during the implementation period, with no detected increase in employment. The Sommers et al. estimates are causally credible because (a) they exploit a clean pre-vs-post implementation contrast within a panel design, (b) they have a within-Arkansas placebo group (ages 19–29 and 50–64, not yet phased in), and (c) the implementation window is well-defined and the policy was salient. Their finding that the disenrollment was concentrated among people who already met the work requirement but failed to document it became the empirical anchor for the policy debate over whether work requirements raise employment (they do not appear to) versus generate paperwork-induced coverage loss (they do).

A closely related individual-panel literature is the **federal-court factual record**. In *Gresham v. Azar*, 363 F. Supp. 3d 165 (D.D.C. 2019), Judge Boasberg vacated CMS's approval of Arkansas Works on Administrative Procedure Act grounds, relying heavily on the disenrollment record (18,164 disenrolled by December 2018) and on the absence of any evidence in the administrative record that CMS had considered the coverage-loss consequences of the policy. The same judge subsequently vacated approvals for Indiana, New Hampshire, Michigan, and Kentucky on similar grounds. The court record, which is in the public domain, is a useful source of denominator and event counts that supplement the published academic literature.

The second strand is the **administrative-claims and survey-aggregate literature** on community engagement requirements before the courts halted them. Gangopadhyaya and Waxman (2025) use a triple-difference design across all 50 states and Washington, DC, applying state-by-age phase-in variation as additional identification leverage. Their results parallel Sommers et al. directionally: coverage loss among the policy-eligible age band, no detected employment effect, and modest evidence of cross-state spillovers in the form of increased uninsurance. Allen, Sommers, and Cole (2024) examine the New Hampshire two-month window using BRFSS aggregates and find no measurable effect on employment or coverage, which is unsurprising given the brevity of the implementation. Hinton et al. (2019) and the Kaiser Family Foundation's Medicaid Waiver Tracker

provide the descriptive policy literature underlying these designs.

The third and newest strand is the **early Georgia Pathways descriptive literature**. The Georgia Budget and Policy Institute has published periodic updates documenting Pathways enrollment, qualifying-hours documentation rates, and budget impact. As of June 2025 (their most recent published update), Pathways enrollment was approximately 8,000, compared to a state projection of 345,000 eligible Georgians and an academic projection from the Center on Budget and Policy Priorities of approximately 100,000 first-year enrollees. The Urban Institute (2024) and the Center for Children and Families (Georgetown University, 2024) have published descriptive analyses of Pathways' first-year design and operations but have not published causal evaluations of population-level coverage outcomes.

Beyond these three direct strands, the broader literature on the political economy of Medicaid work requirements (Rosenbaum 2018; Rosenbaum and Schmidt 2019; Garfield et al. 2018) and on conditional welfare programs more generally (Moffitt 2003; Cancian, Han, and Noyes 2014) provides important context but is not directly probative of the causal questions this paper addresses.

What is missing from the literature, and what this paper provides, is a unified empirical treatment of the two only-actually-implemented state policies using the same data sources and the same inferential approach, with explicit attention to the limits imposed by the small number of treated clusters and by the substantive opposition of the two policy designs.

---

3. Data and Treatment Panel

3.1 Treatment classification

I constructed a state-month treatment panel covering 51 jurisdictions (50 states + DC) from January 2016 through December 2024. The panel records six mutually exclusive status categories at each state-month: `not_approved`, `applied_pending`, `approved_not_implemented`, `implemented`, `enjoined`, and `rescinded`. The primary treatment variable is `is_active_treatment = 1 iff status == implemented`. Construction details, source URLs, and ambiguous-classification decisions are documented in `data/raw/treatment_panel_notes.md` and reconciled against the KFF Medicaid Waiver Tracker, the NASHP state tracker, individual CMS approval letters, and the relevant federal-court orders.

An initial cleaning pass flagged Arkansas, Georgia, Indiana, Michigan, New Hampshire, and Utah as ever-treated. A subsequent treatment-classification review restricted the ever-treated set to Arkansas and Georgia on the grounds that the other four states' implementations were too brief (two to ten weeks of active enforcement, immediately halted by federal court order or by FFCRA continuous-enrollment rules) to support causal evaluation. Indiana's ten-month

implementation involved a graduated phase-in with effectively zero hours required through September 2019, and the state administratively suspended reporting in October 2019 pending *Rose v. Azar*. Michigan and Utah implemented for at most three months before COVID-related continuous-enrollment rules suspended all redeterminations nationwide. New Hampshire's implementation lasted approximately two months and was halted by *Philbrick v. Azar* before any sanctioning of beneficiaries.

The result is a treated set of two: Arkansas (cohort 2018, ten months of implementation) and Georgia (cohort 2023, 18 months and ongoing). Indiana, Michigan, New Hampshire, and Utah remain in the analytic sample but are coded as never-treated for main-spec purposes. As a robustness check, the analysis scripts also produce specifications that drop these four states entirely; results are substantively unchanged.

3.2 Outcome data

I draw outcome measures from three population surveys:

1. **American Community Survey (ACS)**: 1% public-use microdata samples from IPUMS USA covering 2016–2023. The ACS provides Medicaid coverage (`HCOVANY` and `HINSCAID`), employment (`EMPSTAT`), and uninsurance (constructed from `HCOVANY` and `HCOVPRIV`). I construct state-year aggregates restricted to the policy-eligible sub-population in each case study (defined in Sections 4 and 6 below). The ACS is the workhorse of the state-year analyses because of its large sample sizes and consistent measurement across years.
2. **Current Population Survey Annual Social and Economic Supplement (CPS-ASEC)**: 2017–2024 waves from IPUMS CPS, covering reference-year 2016–2023 outcomes. CPS provides Medicaid-last-year (`HIMCAID`) as a cross-check on ACS Medicaid coverage. Sample sizes are smaller and the question reference period is "any time during the prior calendar year," which dilutes within-year treatment timing.
3. **Behavioral Risk Factor Surveillance System (BRFSS)**: 2016–2023 waves from CDC, covering self-rated health, cost-related care barriers, and chronic-disease indicators. BRFSS is used as a secondary battery to test for downstream health spillovers; the primary outcome focus is coverage and employment.

Sample restrictions, variable construction, and outlier-handling rules are documented in `data/scripts/01_clean_acs.py`, `02_clean_cps.py`, and `03_clean_brfss.py`. Cleaning scripts log row counts before and after each restriction.

3.3 Analytic panel

The main analytic file is `data/clean/analytic_panel_state_year.parquet`, a state-year balanced panel of 51 jurisdictions $\times$ 9 years (2016–2024) = 459

state-year observations. Each row carries the policy-eligible-population outcome aggregates for both Arkansas-style eligibility (ages 30–49, ≤100% FPL adult expansion adults) and Georgia-style eligibility (ages 19–64, ≤100% FPL adults in non-expansion states or non-expansion-relevant population), along with state-by-year fixed-effect dummies, an `is_active_treatment` indicator, and `event_time` indicators.

3.4 Why ACS state-year aggregates instead of individual-panel data

Sommers et al. (2020) used an individual-panel survey, which is the right design for direct causal evaluation of disenrollment effects because it allows within-person pre-vs-post comparisons. I do not replicate their individual-panel design because (a) the underlying panel is proprietary, and (b) ACS state-year aggregates are sufficient for a descriptive screen of the implementation episode. The cost is precision: a state-year aggregate restricted to the policy-eligible subpopulation, with only one treated state, is much less powered than an individual-panel design.

I am explicit throughout the paper that this is a descriptive screen, not a re-estimation of the Sommers et al. parameter. The expected magnitude of the diluted Sommers et al. effect, propagated through to ACS state-year aggregates, is documented in `analysis/tables/sommers_reconciliation.csv` and discussed in Section 5.3.

---

4. Arkansas Case Study: Policy Context

4.1 The Arkansas Works community engagement requirement

Arkansas's Section 1115 work-requirement amendment, formally titled the Arkansas Works community engagement requirement, was approved by CMS on March 5, 2018 in a letter from then-Administrator Seema Verma to Arkansas Department of Human Services Director Cindy Gillespie. The amendment required Medicaid expansion enrollees aged 30–49 at or below 100% of the federal poverty line to document 80 hours per month of work, job training, education, volunteer service, or caregiving activity, with reporting to occur via an online portal. Failure to report in any month was a "strike"; three strikes in a calendar year resulted in disenrollment for the remainder of that calendar year. The age 30–49 group was the first phased in (June 1, 2018), with younger age bands (19–29) scheduled for phase-in beginning January 2019.

The administrative implementation was notably aggressive. Arkansas elected to enforce reporting through an online-only portal (`Access.Arkansas.gov`) with limited telephonic assistance, no late-reporting grace period, and a default-to-disenrollment posture. The state did not match enrollees against existing administrative datasets (state unemployment-insurance wage records, SNAP work-registration records) that would have allowed automatic credit for documented

employment. Approximately 70 percent of the nominally subject population already met the qualifying-hours threshold through employment recorded in other administrative datasets; the reporting requirement therefore functioned primarily as a paperwork burden rather than as a substantive activity requirement (Sommers et al. 2020).

By December 2018, Arkansas reported 18,164 cumulative disenrollments from Arkansas Works attributable to the community engagement requirement. The disenrollment count was the central factual basis of Judge Boasberg's *Gresham v. Azar* opinion vacating CMS approval on March 27, 2019. The state suspended enforcement effective April 1, 2019; the Biden administration's CMS formally withdrew the waiver on March 17, 2021.

The implementation window is therefore well-defined: **June 1, 2018 through March 31, 2019** (ten months of active enforcement; four state-quarters). The post-implementation comparison period extends from April 2019 forward but is contaminated by COVID-related continuous-enrollment rules from March 2020 onward, so I focus the case study on calendar year 2018 vs. calendar year 2019 outcomes and report sensitivity to extending the post-window through 2024.

4.2 Identification strategy

The Arkansas case study uses a single-treated-state difference-in-differences specification:

$$Y_{st} = \alpha_s + \gamma_t + \beta \cdot \text{Treated}_s \cdot \text{Post}_t + \varepsilon_{st}$$

where Y_{st} is the state-year aggregate outcome (Medicaid coverage, employment, uninsurance, or CPS Medicaid last year) for the policy-eligible sub-population (ages 30–49, $\leq$100% FPL adult-expansion-eligible adults), α_s are state fixed effects, γ_t are year fixed effects, and Treated $\cdot$ Post is the active-implementation indicator for Arkansas in 2018 and 2019. The comparison group is the 46 never-treated states (excluding Arkansas, Georgia, and the four "approved-but-halted" states whose policy environments may be contaminated by anticipation effects).

This specification is **not a credible causal design** in the conventional sense. With one treated state, conventional cluster-robust inference is biased; with one treated unit, there is no identification of the treatment effect separately from any Arkansas-specific time-varying confounder during 2018–2019. I report this specification as a descriptive screen — a question of whether the Arkansas-specific 2018–2019 deviation from never-treated states in the policy-eligible sub-population is in the direction the prior individual-panel literature would predict.

To stress-test the inference, I report three complementary p-values:

- **CR1 cluster-robust**: the conventional Stata-default cluster-robust standard error clustered at the state level. Known to be downward-biased with one treated cluster (MacKinnon and Webb 2018).

- **Wild-cluster bootstrap**: 1,999 Rademacher draws over state-level residuals (Cameron, Gelbach, and Miller 2008; Webb 2014). Valid in single-treated-cluster designs but conservative.
- **Fisher randomization inference**: rotates the treatment label across all 47 candidate states (Arkansas plus 46 never-treated) and computes the share of placebo coefficients larger in absolute value than the actual Arkansas estimate. Distribution-free but requires the treatment effect to be exchangeable across candidate placebo states.

I also report a Romano-Wolf step-down multiple-testing correction (Romano and Wolf 2005, 2016) across the four-outcome family (ACS Medicaid coverage, ACS employed, ACS uninsured, CPS Medicaid last year). The Romano-Wolf correction is applied to the wild-cluster bootstrap p-values, with 1,999 bootstrap draws and the family ordered by absolute t-statistic.

4.3 Sensitivity analyses

For robustness I also report:

- **HonestDiD bounds** (Rambachan and Roth 2023): allows the post-treatment violation of parallel trends to be bounded by a multiple M of the maximum pre-period violation, with the headline robust confidence set returned for $M \in \{0.5, 1.0, 2.0\}$.
- **Cross-language implementation comparison**: a separate R-based wild-cluster bootstrap (`fwildclusterboot` package) and an R-based HonestDiD implementation (`HonestDiD` package) verify that the Python implementations are not subject to language-specific numerical or default-parameter artifacts. The comparison tables are in `analysis/tables/python_r_bootstrap_comparison.csv` and `analysis/tables/python_r_honestdid_comparison.csv`.
- **Sommers reconciliation**: I compute the implied state-year ACS aggregate effect under the assumption that the Sommers et al. (2020) individual-panel estimate of −6.8 pp Medicaid coverage and +4.0 pp uninsurance applies uniformly across the policy-eligible population, and I check whether the state-year aggregate point estimates from the case-study specification are consistent in sign and magnitude with that diluted Sommers benchmark.

The Sommers reconciliation is the inferentially load-bearing piece of the Arkansas case study: it allows the reader to interpret the case-study point estimates as either "consistent with the prior individual-panel evidence" or "inconsistent" regardless of whether the case-study standard errors reject zero.

---

5. Arkansas Case Study: Findings

5.1 Headline state-year point estimates

[Exhibit 1: AR case-study state-year DiD estimates — see `analysis/tables/two_case_study_summary.csv`, rows where `case == "AR"` and `spec == "state_year_TWFE"`]

The headline ACS state-year point estimates for Arkansas during the implementation period are:

Outcome	Coefficient	CR1 SE	CR1 p	Wild bootstrap p	Fisher RI p
ACS Medicaid coverage (eligible)	+0.0015 (+0.15 pp)	0.0021	0.476	0.474	0.913
ACS Employed (eligible)	+0.0145 (+1.45 pp)	0.0015	<0.001	0.088	0.174
ACS Uninsured (eligible)	+0.0047 (+0.46 pp)	0.0018	0.014	0.325	0.696
CPS Medicaid last year (eligible)	+0.0061 (+0.61 pp)	0.0036	0.097	0.461	0.761

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 Romano-Wolf step-down corrected p-values across the four-outcome family are: ACS Medicaid 0.494, ACS Employed 0.098, ACS Uninsured 0.321, CPS Medicaid 0.458. **No outcome survives multiple-testing correction at conventional levels.**

The CR1-versus-wild divergence on the employment outcome is particularly informative. The conventional CR1 t-statistic is 9.4 (a six-sigma rejection), but the wild-cluster bootstrap p is 0.088. The ratio of wild-bootstrap SE to CR1 SE on this outcome is approximately 4.8, reflecting the severe downward bias of CR1 with one treated cluster on an outcome with concentrated within-Arkansas time-series variance. This is the canonical few-clusters inference inflation pattern documented in MacKinnon and Webb (2018, Section 4) and is one reason a pooled two-state design would not be credible here.

5.2 Direction and magnitude versus Sommers et al. (2020)

[Exhibit 2: Sommers reconciliation table — see `analysis/tables/sommers_reconciliation.csv`]

The Sommers et al. (2020) individual-panel estimates apply directly to the policy-eligible Arkansas population (ages 30–49, ≤100% FPL, adult expansion).

They are: Medicaid coverage −6.8 pp; uninsurance +4.0 pp; employment effectively zero. Under the assumption that the policy-eligible sub-population represents roughly 13.5% of the broader ACS aged-30–49 sample in Arkansas (the in-data eligible-share weight), the implied state-year ACS aggregate effect would be: Medicaid coverage −0.92 pp, uninsurance +0.54 pp, employment ≈ 0.

The case-study point estimates above show: Medicaid coverage **+0.15 pp** (wrong sign relative to Sommers; not significant; magnitude smaller than implied), uninsurance **+0.46 pp** (right sign; magnitude consistent with implied Sommers; not significant under wild bootstrap), employment **+1.45 pp** (positive sign inconsistent with Sommers's null but not significant under wild bootstrap).

The ACS Medicaid coverage result is the puzzle. A state-year aggregate of the same policy-eligible population should show a directional decline, not a small directional increase. The most likely explanation is that the post-period (2019) Arkansas Works disenrollees did not disappear from ACS coverage rates in the aggregate because (a) the Sommers et al. effect operated primarily within a narrowly defined panel of working-age adults who had been continuously enrolled in 2017, whereas the ACS aggregate captures cross-sectional churn that absorbs the disenrollment, and (b) the state-year aggregate definition of "policy-eligible" includes adults who were not yet eligible for Medicaid but became eligible as their incomes changed, partially offsetting the disenrollment in the aggregate. The HonestDiD bounds (`analysis/tables/honestdid_case_studies.csv`) confirm that the Medicaid coverage result is fragile to even small violations of parallel trends in the pre-period; at $M = 0.5$ the robust confidence interval already includes zero.

I read this triangulation as: **the state-year ACS aggregate evidence is consistent with the Sommers et al. individual-panel evidence on uninsurance (right sign, plausible diluted magnitude, not statistically significant in this specification) and inconsistent or null on Medicaid coverage (wrong sign in the point estimate, attributable to aggregation and churn effects)**. The case study does not contradict the prior individual-panel literature; it simply does not provide an independent causal estimate.

5.3 Sensitivity

The HonestDiD sensitivity analysis (`analysis/tables/honestdid_case_studies.csv`; $k = 0$ and $k = 1$ post-period horizons, $M \in \{0.5, 1.0, 2.0\}$, CR1 and wild-rescaled standard errors) confirms that the Arkansas case study's point estimates are fragile. The Medicaid coverage and uninsurance robust confidence intervals at $M = 0.5$ include zero under the wild-rescaled SE; the employment robust confidence interval at $M = 0.5$ and $k = 0$ is approximately $[-0.019, +0.037]$, i.e., the positive employment point estimate is consistent with

anything from a 1.9 percentage point employment decline to a 3.7 percentage point gain once realistic pre-trend violation bounds are allowed.

The cross-language Python-vs-R comparison (`analysis/tables/python_r_bootstrap_comparison.csv` and `python_r_honestdid_comparison.csv`) shows agreement to within rounding on all reported quantities.

5.4 Race-and-ethnicity stratification

Stratifying the Arkansas policy-eligible subpopulation (ages 30–49, ≤100% FPL) by race and ethnicity sharpens the headline. Between the 2016–17 pre-period and the 2018–19 implementation period, ACS uninsurance among policy-eligible Black (non-Hispanic) Arkansans rose 7.6 percentage points (from 12.6% to 20.2%); the corresponding rise among policy-eligible White (non-Hispanic) Arkansans was 1.3 percentage points (from 18.2% to 19.6%). Hispanic and Asian/Other strata have small unweighted cell counts ($n < 200$ in either period) and should be read with that caveat; the directional pattern in those strata is small and noisy. The within-Arkansas Black-vs-White divergence on uninsurance is the cleanest within-state subgroup signal in the case study and is consistent with the subgroup pattern reported by Sommers et al. (2020), who found that documented disenrollment in Arkansas Works concentrated among Black enrollees. Full subgroup table is in `analysis/tables/ha_ar_race_subgroup.csv` and `ha_ar_race_subgroup_diffs.csv`.

5.5 What the Arkansas case study supports

The honest reading of the Arkansas case study is:

1. **Directionally consistent with the prior individual-panel literature on uninsurance**: the state-year ACS aggregate uninsurance coefficient is +0.46 pp, right sign, and roughly consistent with the diluted Sommers et al. (2020) +4.0 pp individual-panel effect after population dilution.
2. **Not independently identifying a causal effect**: with one treated state and a 10-month post-window, no outcome survives wild-cluster bootstrap inference at conventional levels, and HonestDiD bounds at $M = 0.5$ already include zero for the load-bearing outcomes.
3. **Useful as a screening exercise**: the case study confirms that the implementation period is detectable in publicly available state-year aggregates, and the direction is consistent with the prior literature. A researcher with access to T-MSIS Analytic File administrative claims data could perform a more powered evaluation; this paper does not have access.

I therefore frame the Arkansas case study as a **descriptive screen of the implementation episode** rather than as a re-estimation of the Sommers et al. causal parameter.

---

6. Georgia Case Study: Policy Context

6.1 The Georgia Pathways to Coverage program

Georgia Pathways to Coverage is structurally different from every other Section 1115 work-requirement waiver. Where Arkansas, Indiana, New Hampshire, Michigan, and Utah added a community engagement requirement to an existing adult expansion population, Georgia created a **new** partial expansion category — up to 100% FPL — and made enrollment in that new category conditional on documenting 80 hours per month of qualifying activity. The qualifying-activity definition is similar to the other states' (work, job training, education, volunteer service, caregiving). The reporting infrastructure is also similar (online portal, monthly attestation). What is different is the policy lever: Pathways gates *entry* into a new benefit rather than *continuation* of an existing benefit.

CMS approved Pathways on October 15, 2020 in the final months of the first Trump administration. The Biden administration's CMS partially rescinded the approval on December 23, 2021, withdrawing the qualifying-hours and premium provisions but leaving the partial-expansion authority in place. Georgia and the Biden administration's CMS then litigated the rescission. On August 19, 2022, U.S. District Judge Lisa Godbey Wood (S.D. Ga.) set aside the rescission on Administrative Procedure Act grounds, reinstating the qualifying-hours and premium provisions. Georgia announced a phased-implementation plan in early 2023 and the program went live on **July 1, 2023**.

The implementation window for the Georgia case study is therefore **July 1, 2023 through December 31, 2024** — 18 months and ongoing. Pathways has not been judicially halted, and the Biden administration declined to attempt a second rescission after the August 2022 ruling. The Trump administration that took office in January 2025 is supportive of Pathways and the broader 2025 federal mandate.

6.2 The take-up problem

Pathways was projected by Georgia's Department of Community Health (DCH) to enroll approximately 345,000 adults in its first year. Independent academic projections (Center on Budget and Policy Priorities 2023; Urban Institute 2023) were lower — approximately 100,000 in year one — primarily because of the documented friction of the reporting requirement. Actual realized enrollment, per the Georgia Budget and Policy Institute's June 2025 update, was approximately **8,000** as of mid-2025: **less than 3 percent of the state's own projection and roughly 8 percent of the academic projection**.

The low realized take-up has direct implications for the empirical analysis. A program whose realized enrolled population is approximately 0.08% of the

state's working-age adult population cannot produce a population-aggregate effect detectable in state-level ACS data on the order of magnitude that Arkansas-style disenrollment did. The expected ACS uninsurance effect under realized 8,000 enrollment is approximately −0.16 percentage points if all 8,000 enrollees were previously uninsured (which is the upper-bound assumption); this is well below the cluster-robust minimum detectable effect of the state-year case-study design.

This is not a statistical argument; it is a substantive one. The Georgia case study cannot detect a Pathways effect on state-population coverage because Pathways has not (yet) covered enough people to move state-population coverage rates. The case study can, however, ask whether *any* aggregate signal is detectable in the first 18 months, and the result is informative about both the policy's reach and the appropriateness of pooling-based designs that implicitly weight Georgia's small post-period by its full state-year observation count.

6.3 Identification strategy

The Georgia case study uses the same single-treated-state DiD specification as Arkansas:

$$Y_{st} = \alpha_s + \gamma_t + \beta \cdot \text{Treated}_s \cdot \text{Post}_t + \varepsilon_{st}$$

where the treated period is 2023 and 2024 and the policy-eligible sub-population is adults aged 19–64 at ≤100% FPL who would have been newly eligible under Pathways (i.e., the partial-expansion-eligible population that did not previously have Medicaid coverage). The comparison group is the 46 never-treated states (excluding Arkansas and the four approved-but-halted states).

Comparison-group choice is more consequential for Georgia than for Arkansas. Georgia is a non-expansion state (it did not adopt the ACA Medicaid expansion to 138% FPL). The natural counterfactual for a partial-expansion conditional-entry program is the other non-expansion states (currently AL, FL, KS, MS, SC, TN, TX, WY, and WI through 2023 — though WI is unusual because it covers up to 100% FPL without ACA expansion authority). I therefore report the headline DiD specification with the full 46-state comparison group and a robustness specification restricted to a "non-expansion-only" comparison pool. The non-expansion-only specification is the cleaner counterfactual; the full-pool specification is comparable to the Arkansas analysis.

I report the same three p-values (CR1, wild-cluster bootstrap, Fisher randomization) and Romano-Wolf step-down across the same four-outcome family as in the Arkansas case study.

6.4 Outcome definitions

The Georgia eligibility definition is broader than Arkansas's. Arkansas Works applied to ages 30–49 at ≤100% FPL; Pathways applies to ages 19–64 at ≤100% FPL. The Georgia case-study eligibility filter is therefore "adults aged 19–64 with family income ≤100% FPL" in the ACS state-year aggregate. The CPS Medicaid-last-year outcome uses the same eligibility filter.

---

7. Georgia Case Study: Findings

7.1 Headline state-year point estimates (full 46-state comparison)

[Exhibit 3: GA case-study state-year DiD estimates — see `analysis/tables/two_case_study_summary.csv`, rows where `case == "GA"` and `spec == "state_year_TWFE"`]

The headline ACS state-year point estimates for Georgia during the Pathways implementation period are:

Outcome	Coefficient	CR1 SE	CR1 p	Wild bootstrap p	Fisher RI p
ACS Medicaid coverage (eligible)	+0.0021 (+0.21 pp)	0.0022	0.358	0.489	0.891
ACS Employed (eligible)	+0.0003 (+0.03 pp)	0.0017	0.882	0.876	0.978
ACS Uninsured (eligible)	−0.0098 (−0.98 pp)	0.0024	<0.001	0.180	0.587
CPS Medicaid last year (eligible)	+0.0006 (+0.06 pp)	0.0028	0.842	0.827	1.000

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 Romano-Wolf step-down corrected p-values are: ACS Medicaid 0.552, ACS Employed 0.975, ACS Uninsured 0.194, CPS Medicaid 0.975. **No outcome survives multiple-testing correction at conventional levels.**

The uninsurance coefficient is the most interesting. The CR1 t-statistic is approximately 4.2 (CR1 $p < 0.001$), but the wild-cluster bootstrap p is 0.18 and the Fisher RI p is 0.59. The ratio of wild-bootstrap SE to CR1 SE is approximately 2.2 for this outcome — a smaller inflation than the Arkansas employment outcome but still substantial. The conventional inference would call this a 0.98

percentage point reduction in uninsurance with high confidence; the wild-cluster inference correctly recognizes that with one treated state, that point estimate is consistent with the noise distribution.

7.2 Non-expansion-only comparison group

[Exhibit 4: GA non-expansion-only sensitivity — see `analysis/tables/ga_non_expansion_only_summary.cs`

Restricting the comparison group to the seven strictly-non-expansion states (AL, FL, MS, SC, TN, TX, WY) flips the uninsurance coefficient to +0.31 pp (CR1 p = 0.55, wild p = 0.48). Under a broader non-expansion definition that includes states without full ACA expansion authority (adding KS and WI), the uninsurance coefficient is +0.14 pp (CR1 p = 0.74, wild p = 0.70). The Medicaid coverage coefficient is +0.57 pp (CR1 p = 0.09, wild p = 0.40) under strict non-expansion comparison and +0.27 pp (CR1 p = 0.35, wild p = 0.44) under broad non-expansion comparison.

Under the substantively correct non-expansion-only comparison, the Georgia case study shows essentially flat aggregate outcomes through 18 months of Pathways. The full-pool specification's apparent uninsurance reduction is driven by the inclusion of ACA-expansion states whose 2023–2024 uninsurance trajectories diverged from Georgia's for reasons unrelated to Pathways (notably, the differential timing of the post-COVID Medicaid unwinding).

7.3 Direction and magnitude versus Pathways enrollment

The Pathways realized enrollment of approximately 8,000 in the first 18 months represents approximately 0.08 percent of Georgia's working-age adult population and approximately 2 percent of the policy-eligible (≤100% FPL, 19–64) subpopulation. The expected aggregate uninsurance reduction under the strongest assumption (all 8,000 enrollees previously uninsured, all enroll in the policy-eligible aggregate) is approximately −0.16 percentage points. The strict non-expansion comparison-group case-study point estimate (+0.31 pp) is the opposite sign; the full-pool point estimate (−0.98 pp) is the right sign but five times the expected magnitude, suggesting that the full-pool result is driven by the unwinding-period divergence between expansion and non-expansion states rather than by Pathways enrollment.

I therefore read the Georgia case study as: **a descriptive null on aggregate population coverage, employment, and uninsurance, consistent with the very low realized take-up of Pathways through 2024**. The implementation is detectable in administrative enrollment data (because the program exists and 8,000 people are enrolled), but it is not detectable in state-year ACS aggregates because 8,000 enrollees is too small a population to move state-level rates.

7.4 Sensitivity

HonestDiD bounds (`analysis/tables/honestdid_case_studies.csv`, GA rows) confirm that all four outcome coefficients are consistent with zero at $M = 0.5$. The Python-R cross-language comparison (`analysis/tables/python_r_bootstrap_comparison.csv` GA rows) confirms agreement.

7.5 What the Georgia case study supports

The honest reading of the Georgia case study is:

1. **A descriptive null on aggregate state-level outcomes through the first 18 months of Pathways**, consistent with the low realized take-up reported by the Georgia Budget and Policy Institute.
2. **Not contradicting any prior individual-level or administrative-claims literature**, because there is essentially no individual-level or administrative-claims literature on Pathways yet.
3. **Not interpretable as evidence that work requirements have no effect on entry into Medicaid more broadly**: the case study shows only that *this* program, under *this* implementation, with *this* level of state outreach and administrative infrastructure, has not yet produced an aggregate signal. The result is more about realized program reach than about the underlying policy mechanism.

---

8. Cross-Comparison: Arkansas Versus Georgia

[Exhibit 5: Side-by-side summary of Arkansas and Georgia case studies]

A reader who has worked through Sections 4–7 will recognize that Arkansas and Georgia are not measuring the same thing. The two implementations differ along three substantively important dimensions:

Policy mechanism. Arkansas Works was a *disenrollment-eligibility condition*: people who already had Medicaid coverage faced new documentation requirements to keep it. Georgia Pathways is a *new-enrollment eligibility condition*: people who do not currently have Medicaid coverage face documentation requirements to obtain it under a new partial-expansion category. These are policy opposites in terms of who bears the documentation burden, what the default-outcome is, and what kind of coverage-rate effect would constitute "the policy working."

Population scale. Arkansas Works applied to approximately 70,000 adults aged 30–49 at $\leq$100% FPL who were already enrolled in Arkansas Works. By December 2018, approximately 18,000 had been disenrolled — a 26% disenrollment rate. Georgia Pathways had a projected eligible population of approximately 345,000 and realized enrollment of approximately 8,000 — a 2.3% take-up rate.

The denominators and the policy levers are sufficiently different that a single "ATT" parameter applied to both is not meaningful.

Identification context. Arkansas Works ran for 10 months in 2018–2019, in a stable pre-COVID policy environment. Georgia Pathways has run for 18 months in 2023–2024, in the immediate aftermath of the post-COVID Medicaid unwinding (April 2023 – June 2024), during which all states experienced large coverage churn. Anything an analyst attributes to Pathways during 2023–2024 must be netted against the unwinding-induced churn affecting all states, especially non-expansion states like Georgia where the unwinding's effects on coverage transitions were more concentrated.

The case-study point estimates and inferential conclusions support reading the two implementations separately:

Dimension	Arkansas Works (June 2018 – March 2019)	Georgia Pathways (July 2023 – December 2024)
Treated population	~70,000 adults aged 30–49 at ≤100% FPL, already enrolled	~345,000 projected, ~8,000 realized, 19–64 at ≤100% FPL
Policy lever	Documentation to retain existing coverage	Documentation to gain new partial-expansion coverage
Post-period length	10 months	18 months and ongoing
Case-study Medicaid coverage estimate	+0.15 pp (wild p = 0.47)	+0.21 pp (wild p = 0.49)
Case-study uninsurance estimate	+0.46 pp (wild p = 0.32)	−0.98 pp full pool (wild p = 0.18); +0.31 pp non-expansion (wild p = 0.48)
Case-study employment estimate	+1.45 pp (wild p = 0.09)	+0.03 pp (wild p = 0.88)
Consistency with prior literature	Directionally consistent with Sommers et al. (2020) on uninsurance	No prior causal estimate to compare; consistent with low realized take-up
Inferential status	Descriptive screen, not independently causal	Descriptive null, consistent with low realized enrollment

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 narrative reader should walk away with **two separate findings**:

1. **Arkansas Works generated a frictional coverage loss** that the prior individual-panel literature (Sommers et al. 2020) documents convincingly. The state-year aggregate evidence in this paper is consistent with that finding in sign and rough diluted magnitude but is underpowered for an independent causal claim. The case study supports the prior literature; it does not displace or contradict it.

2. **Georgia Pathways has not yet produced an aggregate-detectable effect** in its first 18 months. The most likely explanation is the very low realized take-up (8,000 of 345,000 projected). The case study supports

> a "wait and see" empirical posture; it does not yet inform a normative judgment about the policy's effects on coverage, employment, or health.

Treating these two findings as the average of "a policy that visibly took coverage away" and "a policy that has not yet given coverage to anyone" would be substantively misleading. The most defensible empirical posture is the one this paper takes: write each case study separately, present the cross-comparison as a narrative rather than as a pooled estimate, and acknowledge that the existing state-level variation does not support a generalized causal claim about Medicaid work requirements.

---

9. Discussion

9.1 What we learn from the two case studies

The two case studies together teach four substantive lessons.

First, Medicaid work requirements as implemented at the state level between 2018 and 2024 have generated empirical records that do not support generalized causal claims. Only two states produced implementation windows long enough to evaluate. Those two states adopted policies that are substantively opposite in design and produced empirical records that should be read separately.

Second, the Arkansas record is consistent with frictional coverage loss. The prior individual-panel literature (Sommers et al. 2020) is the empirical foundation. This paper's state-year aggregate evidence is directionally consistent on uninsurance and underpowered for an independent causal claim. The most likely policy interpretation is that Arkansas Works' aggressive online-only reporting design generated paperwork-induced disenrollment of people who already met the qualifying-hours threshold but did not document it. This is a finding about program administration as much as about the underlying eligibility condition.

Third, the Georgia record is a descriptive null through 18 months, consistent with very low realized take-up. The case study cannot distinguish between "the policy works exactly as intended but realized enrollment is far below projection" and "the policy does not work as intended." Both readings are consistent with the data. A reader interested in evaluating Pathways' policy effects should wait for additional post-period data and ideally for administrative-claims access (T-MSIS Analytic File) rather than rely on state-year aggregates.

Fourth, the methodological lesson generalizes. Single-treated-cluster designs with conventional cluster-robust inference are not credible for causal claims. Wild-cluster bootstrap and Fisher randomization-inference adjustments are necessary corrections but cannot manufacture statistical power that the design does not have. Multiple-testing corrections across outcome families further

discipline the inference. Researchers evaluating future single-state implementations should report all three p-values and a Romano-Wolf-corrected step-down family p-value as a matter of course.

9.2 What we cannot learn

The case studies in this paper cannot speak to:

- **Federal-mandate effects.** The 2025 federal reconciliation-law mandate (P.L. 119-21), which requires states to impose community engagement requirements on Medicaid expansion adults beginning in 2027, will generate state-by-state variation in implementation timing, administrative aggressiveness, and exemption rules. The empirical record from this paper does not generalize to the federal mandate because (a) the federal mandate operates at a different policy lever (federal floor rather than state ceiling), and (b) the variation across states under the federal mandate will be in administrative design and stringency rather than in the binary "implemented vs. not."
- **Employment effects.** Neither case study has the power to detect an employment effect of the magnitude predicted by the labor-supply literature. The Arkansas case-study employment coefficient is positive (+1.45 pp) but not statistically distinguishable from zero under wild-cluster inference. The Georgia case-study employment coefficient is essentially zero.
- **Health-outcome effects.** I report secondary BRFSS outcome analyses in the analytic panel (results in `analysis/tables/main_did_summary.csv` for additional outcomes), but none of the BRFSS outcomes show effects detectable at the state-year aggregate level for either implementation. This is consistent with the brevity of the implementations and the dilution of any treatment effect through the BRFSS state-level sample.
- **Mechanism.** The case studies cannot decompose the Arkansas effect (such as it is) into mechanisms — paperwork-induced disenrollment versus eligibility-based disenrollment versus employment-related ineligibility — because they operate at the state-year aggregate level. The mechanism literature has to come from the individual-panel literature (Sommers et al. 2020) or from administrative-claims work using T-MSIS Analytic File data.

9.3 Policy implications, narrowly

Given the constrained empirical record, the policy implications are necessarily narrow.

For state Medicaid agencies considering whether and how to implement community engagement requirements under the 2025 federal mandate, the Arkansas record suggests that **administrative design matters more than the underlying eligibility condition**. The Sommers et al. (2020) finding that approximately 70% of Arkansas Works' policy-eligible population already met the

qualifying-hours threshold through employment recorded in other administrative datasets implies that an administrative design that automatically matches enrollees against state UI wage records, SNAP work-registration records, and similar sources would convert paperwork-induced disenrollment into automatic compliance for the vast majority of subject enrollees. Arkansas chose not to do this; future implementations could.

For federal CMS and state policymakers evaluating whether work requirements achieve their stated employment-promotion goals, **neither case study supports a positive employment effect**. The Arkansas case-study employment coefficient is positive but fragile; the Georgia case-study employment coefficient is essentially zero. The Sommers et al. (2020) individual-panel design also found no detected employment effect. The empirical record consistently suggests that the policy lever's primary detectable margin is coverage loss (Arkansas) or non-enrollment (Georgia), not employment promotion.

For congressional oversight of the 2025 federal mandate's implementation, **the realized-take-up gap in Georgia Pathways** is a specific data point worth tracking. Pathways' first-year realized enrollment was approximately 2 percent of the state's own projection. If the federal mandate's implementation produces analogous realized-take-up gaps in additional states, the policy's actual coverage effects may differ substantially from CBO projections that assume full administrative compliance.

9.4 Limitations

Five limitations are explicit.

One-treated-state inferential limits. Each case study has one treated state. Conventional cluster-robust inference is biased; wild-cluster bootstrap inference is the correct correction but is conservative. No outcome in either case study survives Romano-Wolf step-down correction at conventional levels. The case studies are descriptive screens, not causal estimates.

Aggregation dilution. State-year ACS aggregates dilute any within-policy-eligible effect by the size of the non-policy-eligible majority in the state-year aggregate. The Sommers reconciliation in Section 5.3 documents this dilution explicitly for Arkansas; the analogous calculation for Georgia (Section 7.3) shows that the expected dilution renders the case study unable to detect any plausible Pathways effect on state-population aggregates.

Comparison-group sensitivity. The Georgia case study results depend on the choice of comparison group, particularly on whether to include ACA-expansion states whose 2023–2024 unwinding trajectories differ systematically from Georgia's. I report both the full-pool and non-expansion-only specifications; the latter is the substantively cleaner counterfactual.

No administrative-claims data. Neither case study uses T-MSIS Analytic File or state-specific administrative-claims data. The Sommers et al. (2020)

individual-panel record is the closest available administrative substitute for Arkansas; no analogous administrative-claims record is yet available for Georgia. Future revisions of this paper or follow-on work using T-MSIS Analytic File access would supersede the case-study evidence reported here.

COVID-era confounding. The Arkansas post-implementation period extends into 2019 but the FFCRA continuous-enrollment rule began in March 2020, contaminating any analysis of 2020–2023. The Georgia implementation period (July 2023 – December 2024) overlaps the post-COVID Medicaid unwinding (April 2023 – June 2024), during which all states experienced unusual coverage churn. I treat the unwinding period as a confounder for which the non-expansion-only comparison group provides partial control; the full-pool comparison does not.

9.5 What would change the assessment

Three concrete empirical developments would change the assessment in this paper.

T-MSIS Analytic File access for Arkansas 2017–2019 and Georgia 2022–2025. With administrative-claims data, the policy-eligible population can be defined precisely (down to month-of-coverage and reported activity-hours bins), and the inference can be performed at the individual or county-month level with sufficient cluster counts for standard cluster-robust inference. A T-MSIS-based replication of the Sommers et al. (2020) Arkansas finding and a first administrative-claims evaluation of Georgia Pathways are both feasible given DUA approval; I have not pursued the application in this paper but flag it as the most productive next step.

Additional state implementations. If the 2025 federal reconciliation-law mandate is implemented as written, beginning in 2027 all states will have to design and roll out community engagement requirements for the Medicaid expansion population. The cross-state variation in administrative design, exemption rules, and reporting infrastructure will produce sufficiently large-N policy variation for a credible pooled evaluation by approximately 2029–2030 (allowing 24 months of mature post-mandate ACS/CPS/BRFSS waves).

Georgia Pathways enrollment scale-up. If Georgia's outreach efforts produce a substantial enrollment increase in 2025–2026 (the state has committed to a redesigned outreach campaign and to consolidating its eligibility-determination infrastructure), the case-study evidence would update accordingly. As of the December 2024 panel end, the realized take-up problem was the dominant interpretive fact.

9.6 Concluding observation

The Section 1115 Medicaid work-requirements experiments of 2018–2024 produced an empirical record of two policy episodes that are substantively oppo-

site in design, brief in duration, and underpowered for formal causal evaluation in publicly available data. The honest scholarly response to that record is to write each episode up as a case study, refrain from pooling, and wait for either administrative-claims access or the 2025 federal mandate's larger-N variation to support generalized causal claims. This paper takes that posture.

The substantive policy question — what Medicaid work requirements do to coverage, employment, and health — remains open in the population-aggregate empirical record. The administrative-claims and individual-panel literatures have begun to answer it; the population-aggregate evidence in this paper does not contradict those literatures and is consistent with their main findings where comparison is possible. The empirical record will become substantially richer if the 2025 federal mandate is implemented as written and if researchers acquire T-MSIS Analytic File access for the relevant state-years. Until then, the case-study framing is the honest empirical record.

References

References are listed in Vancouver style in the accompanying `bibliography.bib`. Core citations include:

- Sommers BD, Goldman AL, Blendon RJ, Orav EJ, Epstein AM. Medicaid Work Requirements — Results from the First Year in Arkansas. *N Engl J Med.* 2019;381(11):1073–1082. doi:10.1056/NEJMsr1901772
- Sommers BD, Chen L, Blendon RJ, Orav EJ, Epstein AM. Medicaid Work Requirements in Arkansas: Two-Year Impacts on Coverage, Employment, and Affordability of Care. *Health Aff (Millwood).* 2020;39(9):1522–1530. doi:10.1377/hlthaff.2020.00538
- Gangopadhyaya A, Waxman E. Medicaid Work Requirements: New Evidence on Coverage, Employment, and Cross-State Spillovers. Urban Institute Working Paper, 2025. [URL primary source verification pending]
- Allen H, Sommers BD, Cole MB. New Hampshire's Brief Experience with Medicaid Work Requirements. *Health Aff Scholar.* 2024;2(3):qxae012. doi:10.1093/haschl/qxae012 [primary-source verification pending]
- Gresham v. Azar, 363 F. Supp. 3d 165 (D.D.C. 2019); aff'd Gresham v. Cochran, 950 F.3d 93 (D.C. Cir. 2020).
- Stewart v. Azar, 313 F. Supp. 3d 237 (D.D.C. 2018); Stewart v. Azar, 366 F. Supp. 3d 125 (D.D.C. 2019).
- Cameron AC, Gelbach JB, Miller DL. Bootstrap-Based Improvements for Inference with Clustered Errors. *Rev Econ Stat.* 2008;90(3):414–427.
- MacKinnon JG, Webb MD. The Wild Bootstrap for Few (Treated) Clusters. *Econom J.* 2018;21(2):114–135.
- Rambachan A, Roth J. A More Credible Approach to Parallel Trends. *Rev Econ Stud.* 2023;90(5):2555–2591. doi:10.1093/restud/rdad018
- Romano JP, Wolf M. Stepwise Multiple Testing as Formalized Data Snoop-

ing. *Econometrica.* 2005;73(4):1237–1282.
- Georgia Budget and Policy Institute. Pathways to Coverage: Looking Back Two Years and Into the Future. June 2025. https://gbpi.org/pathways-to-coverage-looking-back-two-years-and-into-the-future/
- Kaiser Family Foundation. Medicaid Waiver Tracker: Approved and Pending Section 1115 Waivers by State. Accessed April 2026. https://www.kff.org/medicaid/medicaid-waiver-tracker-approved-and-pending-section-1115-waivers-by-state/
- Rosenbaum S. Medicaid Work Requirements: A Legal Analysis. *Milbank Q.* 2018;96(4):621–625.

Full bibliography is in `literature/bibliography.bib`.

Online Appendix — Two Case Studies of Section 1115 Medicaid Community Engagement Mandates in Arkansas and Georgia

A0. Headline DiD point-estimate tables (companion to Exhibits 1 and 2)

Arkansas case study — ACS state-year DiD point estimates, full p-value battery

Outcome (policy-eligible adults aged 30 to 49 at ≤100% FPL)	Coefficient	CR1 p	Wild-bootstrap p	Randomization p	Pre-period baseline (2016-17)	Relative change
ACS uninsured	+0.46 pp	0.014	0.325	0.696	20.9%	+2.2%
ACS Medicaid coverage	+0.15 pp	0.476	0.474	0.913	13.5%	+1.1%
ACS employed	+1.45 pp	<0.001	0.088	0.174	76.2%	+1.9%
CPS Medicaid (prior year)	+0.61 pp	0.097	0.461	0.761	—	—

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.

Georgia case study — ACS state-year DiD point estimates, full p-value battery (full 46-state comparison)

Outcome (policy-eligible adults aged 19 to 64 at ≤100% FPL)	Coefficient	CR1 p	Wild-bootstrap p	Randomization p	Pre-period baseline (2016-22)	Relative change
ACS uninsured	−0.98 pp	<0.001	0.180	0.587	24.4%	−4.0%
ACS Medicaid coverage	+0.21 pp	0.358	0.489	0.891	5.0%	+4.2%
ACS employed	+0.03 pp	0.882	0.876	0.978	75.1%	+0.0%
CPS Medicaid (prior year)	+0.06 pp	0.842	0.827	1.000	—	—

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.

Romano-Wolf step-down corrected p-values across the four-outcome family: AR — Medicaid 0.494, Employed 0.098, Uninsured 0.321, CPS Medicaid 0.458; GA — Medicaid 0.552, Employed 0.975, Uninsured 0.194, CPS Medicaid 0.975. No outcome survives at conventional levels in either case study.

A. Treatment classification

The full state-month treatment panel and event-history file are at:

- `data/raw/state_work_req_treatment_panel.csv` (5,508 rows; 51 jurisdictions × 108 months, January 2016 – December 2024)
- `data/raw/treated_state_event_history.csv` (49 rows; one row per state-event with day-precision date and source URL)
- `data/raw/treatment_panel_notes.md` (construction notes, ambiguous-classification decisions, source documentation)
- `data/clean/treatment_correction_log.md` (the April 2026 correction restricting ever-treated to AR and GA)

B. Cross-language implementation comparison

To verify that the Python implementations are not subject to language-specific numerical or default-parameter artifacts, I re-estimated the wild-cluster boot-

strap and the HonestDiD bounds in R and compared.

- `analysis/tables/python_r_bootstrap_comparison.csv` — Python `boottest` vs R `fwildclusterboot` p-values across all case-study outcomes; agreement within 0.005 on all reported quantities.
- `analysis/tables/python_r_honestdid_comparison.csv` — Python implementation vs R `HonestDiD` package, comparison of robust confidence interval half-widths at $M \in \{0.5, 1.0, 2.0\}$; agreement within 0.0005.
- `analysis/tables/r_wild_bootstrap_results.csv` — full R wild-cluster bootstrap output.
- `analysis/tables/r_honestdid_results.csv` — full R HonestDiD output.

C. Romano-Wolf step-down multiple-testing correction

Full Romano-Wolf step-down output is in `analysis/tables/romano_wolf_summary.csv`. The step-down was applied independently for the AR family (four outcomes) and the GA family (four outcomes), with 1,999 wild bootstrap draws. No outcome in either family survives step-down correction at the 5% level.

D. HonestDiD sensitivity bounds

Full HonestDiD output is in `analysis/tables/honestdid_case_studies.csv`. The table includes both the CR1 standard error variant and the wild-rescaled standard error variant, for $k \in \{0, 1\}$ post-period horizons and $M \in \{0.5, 1.0, 2.0\}$ multipliers of the maximum pre-period violation. For both case studies, the headline outcomes' robust confidence intervals at $M = 0.5$ already include zero under the wild-rescaled standard error.

E. Sommers et al. (2020) reconciliation

Full reconciliation table is in `analysis/tables/sommers_reconciliation.csv`. The reconciliation propagates the Sommers et al. individual-panel point estimates (Medicaid -6.8 pp; uninsurance $+4.0$ pp; employment ≈ 0) through to state-year ACS aggregates under a uniform-effect assumption, and compares the implied diluted effect to the case-study point estimates. The state-year ACS uninsurance point estimate is consistent in sign and magnitude with the implied diluted Sommers effect; the Medicaid coverage point estimate is the wrong sign but small in magnitude.

F. Georgia non-expansion-only sensitivity

Full output is in `analysis/tables/ga_non_expansion_only_summary.csv`. The table reports the GA case-study DiD specification under two restricted comparison groups: (a) strict non-expansion (seven states: AL, FL, MS, SC, TN, TX, WY); (b) broad non-expansion (nine states adding KS and WI). All four outcomes are flat-to-positive under both restricted comparison groups,

indicating that the full-pool specification's apparent uninsurance reduction is driven by ACA-expansion states' divergent 2023–2024 trajectories.

G. Specification registry and analysis scripts

- `analysis/01_main_did.py` — main case-study TWFE estimation
- `analysis/02_two_case_studies.py` — case-study specification with cluster-robust, wild-bootstrap, and Fisher RI p-values
- `analysis/03_honestdid_case_studies.py` — Python HonestDiD implementation
- `analysis/04_sommers_reconciliation.py` — diluted Sommers reconciliation
- `analysis/05_ga_non_expansion_only.py` — Georgia non-expansion-only sensitivity
- `analysis/06_romano_wolf.py` — Romano-Wolf step-down across outcome family
- `analysis/07_r_wild_cluster_bootstrap.R` — R wild-cluster bootstrap cross-language replication
- `analysis/08_python_r_comparison.py` — Python-vs-R comparison table builder
- `analysis/09_r_honestdid.R` — R HonestDiD cross-language replication
- `analysis/10_honestdid_comparison.py` — Python-vs-R HonestDiD comparison

All scripts run from `data/clean/analytic_panel_state_year.parquet` and produce outputs in `analysis/tables/`. Logs are in `analysis/log/`.

H. Data sources

- IPUMS USA (ACS 1% public-use microdata): `data/raw/ipums/usa_*` (Ruggles et al. 2024)
- IPUMS CPS (CPS-ASEC): `data/raw/ipums/cps_*` (Flood et al. 2024)
- CDC BRFSS: `data/raw/brfss/LLCP*.XPT` (CDC BRFSS 2016–2023)
- KFF Medicaid Waiver Tracker: https://www.kff.org/medicaid/medicaid-waiver-tracker-approved-and-pending-section-1115-waivers-by-state/
- NASHP State Tracker (work requirements): https://nashp.org/state-tracker/a-snapshot-of-state-proposals-to-implement-medicaid-work-requirements-nationwide/
- Individual CMS approval and rescission letters, federal-court orders: URLs recorded per-row in `data/raw/treated_state_event_history.csv`