

# Telehealth Parity Laws and Depression Diagnosis Prevalence: Evidence from Staggered State Adoption

APEP Autonomous Research\*  
@anonymous

January 28, 2026

## Abstract

Do state telehealth parity laws—which require private insurers to cover telehealth services—increase mental health care access? Using a difference-in-differences design exploiting the staggered adoption of telehealth parity laws across U.S. states adopting between 2012 and 2019, I estimate the effect on lifetime depression diagnosis prevalence (the share of adults ever told they have depression) using BRFSS data. I employ the heterogeneity-robust estimator of [Callaway and Sant'Anna \(2021\)](#) with not-yet-treated states as controls. Because the outcome panel begins in 2011, states adopting before 2012 (always-treated in the sample) do not contribute to identification; the analysis estimates effects for the 27 states adopting during 2012–2019. The overall average treatment effect on the treated is  $-0.48$  percentage points ( $SE = 0.35$ ), with a 95% confidence interval spanning  $[-1.16, 0.20]$ . This null finding suggests that telehealth parity laws alone may have had limited effects on mental health care access during the pre-COVID period, possibly due to implementation barriers, limited provider adoption, or the continued dominance of in-person care.

**JEL Codes:** I11, I13, I18

**Keywords:** telehealth, mental health, health insurance regulation, telemedicine, parity laws

---

\*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

## 1. Introduction

Access to mental health care in the United States remains a significant policy challenge with substantial economic and social consequences. Approximately one in five adults experiences mental illness annually, yet fewer than half receive treatment ([NIMH, 2023](#)). This treatment gap translates to reduced labor force participation, lower productivity, and increased healthcare costs from untreated conditions. The barriers to accessing mental health care are multifaceted: provider shortages, particularly acute in rural areas; geographic isolation that makes traveling to appointments difficult; persistent stigma that discourages help-seeking; and financial barriers including inadequate insurance coverage ([Andrilla et al., 2018](#)).

Telehealth—the delivery of health services via telecommunications technology—has been proposed as a promising solution to expand access by reducing geographic and scheduling barriers. Mental health services are particularly well-suited to telehealth delivery, as many forms of therapy require primarily verbal communication and visual observation, without the need for physical examination or procedures. A patient can receive psychotherapy or psychiatric medication management from their home, eliminating travel time and potentially reducing stigma associated with visiting mental health facilities.

Recognizing this potential, states began adopting “telehealth parity” laws in the late 1990s, requiring private insurers to cover services delivered via telehealth on par with in-person care. Texas and Oklahoma were early adopters, enacting parity laws in 1997. A subsequent wave of adoption occurred between 2012 and 2016, with many states passing legislation as telehealth technology improved and broadband access expanded. By 2019, 37 states had enacted some form of telehealth parity legislation ([CCHPCA, 2019](#)). These laws aimed to remove financial barriers to telehealth adoption by ensuring that patients would not face higher cost-sharing or coverage exclusions for services delivered via telehealth rather than in person.

This paper examines whether telehealth parity laws increased mental health care access, as measured by lifetime depression diagnosis prevalence—the cumulative share of adults who have ever received a depression diagnosis. I focus on depression because it is common (affecting approximately 8% of adults in any given year), disabling, and treatable, yet underdiagnosed ([NIMH, 2023](#)). If telehealth parity laws successfully reduced barriers to mental health care, we might expect more individuals to receive evaluations and diagnoses, increasing the measured prevalence of diagnosed depression.

The analysis focuses on the pre-COVID period (2011–2019) to avoid confounding from the pandemic-era emergency telehealth expansions that dramatically changed the telehealth landscape. During COVID-19, temporary waivers of licensing restrictions, expanded Medicare and Medicaid coverage, and dramatic shifts in patient and provider behavior led to telehealth

utilization increases of over 4,000% for mental health services (Pew, 2021). Studying the pre-COVID period allows examination of whether the gradual, legislative approach to expanding telehealth coverage achieved its intended effects in a more typical policy environment.

I exploit the staggered adoption of parity laws across states as a natural experiment, employing the heterogeneity-robust difference-in-differences estimator of Callaway and Sant'Anna (2021). This approach addresses well-documented problems with traditional two-way fixed effects (TWFE) estimation under staggered treatment timing, which can produce biased estimates when treatment effects are heterogeneous across time or cohorts (Goodman-Bacon, 2021; de Chaisemartin and D'Haultfoeuille, 2020). Because the outcome data begin in 2011, states adopting before 2012 are “always-treated” in the sample and do not contribute to identification; the analysis estimates effects for the 27 states adopting during 2012–2019, using 14 never-treated states and not-yet-treated states as controls.

The primary finding is a null result. The overall average treatment effect on the treated (ATT) is  $-0.48$  percentage points (SE = 0.35), with a 95% confidence interval of  $[-1.16, 0.20]$ . This estimate is statistically insignificant and suggests no detectable effect of telehealth parity laws on lifetime depression diagnosis prevalence during this period. Event study estimates show no significant pre-trends, supporting the parallel trends assumption, and no significant post-treatment effects at any time horizon from 0 to 5 years after adoption.

Several explanations may account for this null finding. First, parity laws establish legal coverage requirements but do not ensure provider adoption of telehealth technology. During the 2011–2019 period, telehealth utilization remained low despite parity mandates, with only about 0.1% of outpatient visits conducted via telehealth in 2016 (Barnett et al., 2018). Providers faced barriers including technology costs, training requirements, uncertain reimbursement, and concerns about clinical quality. Second, parity laws primarily affect fully-insured private plans regulated under state law; self-insured employer plans (covering over 60% of workers with employer-sponsored insurance) are exempt from state mandates under ERISA (DOL, 2020). This limits the population “treated” by state parity laws. Third, mental health treatment often requires ongoing therapeutic relationships that patients and providers may have preferred to maintain in person during this period, before video communication became ubiquitous.

This paper contributes to several literatures. First, it contributes to the growing body of research on telehealth policy (Mehrotra et al., 2017), providing causal evidence on the effects of coverage mandates in the pre-COVID era. Second, it contributes to research on mental health care access (Bishop et al., 2016), examining whether insurance-based interventions can address the persistent treatment gap. Third, it contributes to the broader literature on health insurance regulation and mandate incidence (Pauly, 1986), demonstrating that

coverage mandates may have limited effects when supply-side barriers remain binding. The null finding documented here suggests that legal mandates alone may be insufficient to drive adoption without complementary investments in technology, provider training, and patient education.

The remainder of this paper proceeds as follows. Section 2 provides institutional background on telehealth parity laws and the mental health care landscape. Section 3 reviews the related literature. Section 4 describes the data sources and sample. Section 5 presents the empirical strategy. Section 6 reports the results. Section 7 discusses interpretations and limitations. Section 8 concludes.

## 2. Institutional Background

### 2.1 Telehealth Parity Laws: Definitions and Variation

Telehealth parity laws require health insurers to provide coverage for services delivered via telecommunications technology. These laws emerged in response to insurer reluctance to cover telehealth services, which was driven by concerns about fraud, overutilization, and clinical quality. Without parity requirements, insurers could exclude telehealth services entirely, impose higher cost-sharing, or reimburse at lower rates than comparable in-person services.

Two types of parity requirements exist, and states vary in which type they mandate. **Coverage parity** requires insurers to cover the same services via telehealth as they would cover in-person. This prevents insurers from categorically excluding telehealth but does not specify reimbursement rates. A service that would be covered if delivered in an office must also be covered if delivered via video. **Payment parity** goes further, requiring insurers to reimburse telehealth services at the same rate as in-person services. This addresses concerns that lower reimbursement rates would make telehealth economically unviable for providers.

States also vary in the modalities covered by their parity laws. Early laws often required only synchronous video (real-time audio-visual communication), reflecting the technology available at the time. More recent laws have expanded to include audio-only services (telephone), which are particularly important for patients lacking broadband access or video-capable devices, and store-and-forward services (asynchronous transmission of images or data for later review), which are common in dermatology and radiology.

Additional variation exists in originating site restrictions. Many early laws required the patient to be located at a medical facility (clinic, hospital, or nursing home) rather than their home. These restrictions were intended to ensure appropriate clinical settings and address concerns about fraud, but they limited the convenience benefits of telehealth. Over time, states have gradually relaxed these restrictions, with some eliminating originating site

requirements entirely.

Texas and Oklahoma were pioneers, enacting parity laws in 1997. California followed in 1996 with coverage parity requirements. A substantial wave of adoption occurred between 2012 and 2016, driven by several factors: technological improvements that made video consultations more practical; expansion of broadband internet access; growing evidence on the clinical effectiveness of telehealth; and advocacy by telehealth providers and patient groups. Table 1 summarizes the adoption timeline.

By 2019, 37 states plus the District of Columbia had enacted some form of telehealth parity legislation, leaving 14 states without comprehensive parity requirements. States without parity laws span diverse regions and political orientations, including Alabama, Florida, Idaho, Massachusetts, North Carolina, and Wyoming. The variation in adoption timing provides the identifying variation for this study.

Table 1: Telehealth Parity Law Adoption by Period

Period	States Adopting	Cumulative Total	Contribution to DiD
Pre-2012 (always-treated)	10	10	None <sup>a</sup>
2012–2015	16	26	Treated cohorts
2016–2019	11	37	Treated cohorts
Never (through 2019)	14	—	Control group

Notes: States includes 50 states plus DC. Source: CCHPCA, NCSL. <sup>a</sup> Pre-2012 adopters have no pre-treatment observations in the 2011–2019 outcome panel and are “always-treated”; they do not contribute to identification of treatment effects in the Callaway-Sant’Anna framework.

## 2.2 Mental Health Care Access in the United States

Mental health care access in the United States faces persistent and well-documented challenges. The treatment gap—the difference between those who need mental health care and those who receive it—remains substantial despite decades of policy attention. Approximately 57% of adults with mental illness received no treatment in the past year, with rates even higher among certain demographic groups and in rural areas ([NIMH, 2023](#)).

Provider shortages are a fundamental constraint. The American Association of Medical Colleges projects a shortage of 14,000 to 31,000 psychiatrists by 2024. The shortage is particularly acute in rural areas, where the ratio of mental health providers to population can be ten times lower than in urban areas. More than 120 million Americans live in federally designated Mental Health Professional Shortage Areas (MHPSSAs), areas where the ratio of mental health providers to population falls below federal thresholds ([HRSA, 2020](#)). In some rural counties, the nearest psychiatrist may be hours away.

Geographic barriers compound provider shortages. Even when providers are available, traveling to appointments imposes costs of time, transportation, and lost wages. For patients with severe mental illness, these barriers may be particularly challenging, as symptoms can make travel difficult. Patients may miss appointments or delay seeking care, leading to worsening conditions.

Stigma remains a significant barrier. Despite public education campaigns, mental illness continues to carry social stigma that discourages help-seeking. Patients may fear discrimination at work or in their communities if their mental health treatment becomes known. Visiting a mental health clinic signals mental illness in a way that visiting a general medical clinic does not, potentially deterring some patients from seeking care.

Wait times for mental health appointments can be substantial. A 2018 survey found that the median wait time for an appointment with a psychiatrist for a new patient was 25 days, compared to just 20 days for a new patient appointment with a family medicine physician ([Bishop et al., 2016](#)). In some areas, wait times extend to months, during which conditions may worsen.

These barriers make mental health services a compelling use case for telehealth. Video-based psychotherapy and psychiatric medication management can be delivered to patients' homes, eliminating travel time and geographic barriers. Patients can access care during evenings or weekends, reducing conflicts with work schedules. The privacy of receiving care at home may reduce stigma-related concerns. For patients in MHPSSAs, telehealth can connect them with distant providers, dramatically expanding the available provider pool.

### **2.3 Limitations of Parity Laws**

Despite the apparent promise of telehealth parity laws, several factors may limit their effectiveness in improving mental health care access.

The Employee Retirement Income Security Act (ERISA) preempts state regulation of self-insured employer health plans. Employers who self-insure—bearing the financial risk of employee healthcare claims rather than purchasing insurance from a carrier—are subject to federal rather than state regulation. Self-insured plans cover over 60% of workers with employer-sponsored insurance, and this share is higher among large employers ([DOL, 2020](#)). Because state telehealth parity laws do not apply to self-insured plans, the population directly affected by these laws is smaller than the total privately insured population. This ERISA preemption is a general limitation of state insurance mandates, affecting not just telehealth parity but also mental health parity, preventive care mandates, and other state-level coverage requirements.

Parity laws establish coverage requirements but do not require providers to offer telehealth

services. Provider adoption requires investment in technology, including video equipment, secure communication platforms, and integration with electronic health records. Providers must be trained in telehealth-specific clinical skills, including conducting assessments via video, managing technology failures, and ensuring patient privacy. Workflow changes are needed to incorporate telehealth into practice schedules. Regulatory uncertainty about licensing, prescribing, and reimbursement may have deterred adoption. During the 2011–2019 period, provider adoption remained low despite growing coverage, with telehealth accounting for less than 1% of outpatient mental health visits ([Barnett et al., 2018](#)).

Legal heterogeneity across state parity laws may dilute their effects. Coverage-only parity laws may have limited impact if reimbursement rates remain too low to incentivize provider adoption. Originating site restrictions may negate convenience benefits. Restrictions on audio-only services exclude patients lacking broadband or video capability. When we estimate an average effect of “parity laws” as a binary indicator, we pool together laws with potentially very different “doses.”

Demand-side barriers may also limit effectiveness. Patients unfamiliar with telehealth technology may be reluctant to try it. Older adults, who have higher rates of depression, may be less comfortable with video communication. Patients may simply prefer in-person care, particularly for mental health treatment where the therapeutic relationship is central. A parity law removes an insurance barrier but does not address these preferences.

### 3. Related Literature

This paper relates to several strands of the economics and health services research literatures.

#### 3.1 Telehealth Policy and Utilization

A growing literature examines telehealth policy and utilization patterns. [Barnett et al. \(2018\)](#) document trends in telemedicine use in a large commercially insured population from 2005 to 2017, finding that while telehealth use grew substantially in relative terms, it remained rare in absolute terms, accounting for only 0.1% of outpatient visits by 2016. Growth was concentrated in direct-to-consumer telemedicine for acute care rather than mental health services.

[Mehrotra et al. \(2017\)](#) study telehealth use among rural Medicare beneficiaries, finding rapid growth in tele-mental health specifically, with wide variation across states. This rural Medicare population differs from the privately insured population affected by state parity laws, but the findings suggest telehealth can serve mental health needs when adopted.

The COVID-19 pandemic generated substantial evidence on telehealth utilization under emergency conditions. [Pew \(2021\)](#) document that telehealth visits increased from approximately 1% to over 40% of outpatient visits during the first months of the pandemic, with particularly strong adoption for mental health services. This evidence demonstrates the potential of telehealth but reflects an environment with temporarily suspended regulations, urgent necessity, and dramatic shifts in patient and provider behavior—conditions quite different from the gradual policy environment studied here.

### 3.2 Mental Health Care Access

Research on mental health care access documents persistent barriers and evaluates interventions. [Bishop et al. \(2016\)](#) document the decline in the number of practicing psychiatrists from 2003 to 2013, arguing that this supply-side constraint helps explain poor access to mental health care. The psychiatry workforce grew more slowly than demand during this period, contributing to longer wait times and access problems.

The mental health parity literature examines whether requiring equal coverage for mental health and physical health conditions improves access. The federal Mental Health Parity and Addiction Equity Act (MHPAEA) of 2008 required large group plans to provide mental health coverage on par with medical/surgical coverage. Studies of MHPAEA find mixed effects, with some studies showing increased utilization of mental health services and others finding no significant changes, suggesting that coverage mandates may have limited effects when supply constraints are binding.

### 3.3 Difference-in-Differences with Staggered Adoption

This paper employs recent methodological advances in difference-in-differences estimation with staggered treatment timing. [Goodman-Bacon \(2021\)](#) demonstrates that the traditional TWFE estimator is a weighted average of all possible two-group/two-period DiD estimates, including problematic comparisons that use already-treated units as controls. When treatment effects vary over time or across cohorts, these problematic comparisons can produce biased estimates.

[Callaway and Sant'Anna \(2021\)](#) propose an estimator that avoids these problems by estimating separate group-time average treatment effects for each cohort and then aggregating. This approach uses only clean comparisons (treated versus not-yet-treated or never-treated) and allows for treatment effect heterogeneity. [Sun and Abraham \(2021\)](#) develop a related approach based on interaction-weighted estimators. [Borusyak et al. \(2021\)](#) propose an imputation-based approach. I employ the Callaway-Sant'Anna estimator as my primary

specification, which has become standard in applied work.

Roth (2022) cautions that pre-trend tests have low power against plausible violations of parallel trends, and that conditioning on passing a pre-trend test can lead to biased estimates. I report pre-trend coefficients but interpret them with appropriate caution, recognizing that the absence of significant pre-trends does not guarantee the validity of the design.

### 3.4 Health Insurance Mandates

A broader literature examines the effects of state health insurance mandates. Pauly (1986) provides foundational analysis of how mandates affect market outcomes, emphasizing that mandates can reduce efficiency by forcing coverage of services that some consumers would not choose. Subsequent empirical work has estimated the costs and benefits of specific mandates.

This literature generally finds that mandates increase coverage of the mandated service but that effects on utilization are often smaller than effects on coverage, suggesting that coverage is not always the binding constraint. This finding is consistent with the null result documented here: parity laws may increase coverage for telehealth without substantially changing utilization if other barriers (provider availability, patient preferences, technology access) remain binding.

The state-level variation in mandate adoption has been extensively used for causal identification. Researchers have examined mandates for infertility treatment coverage, finding modest increases in utilization among affected populations. Studies of mental health parity mandates (requiring equal coverage for mental health and physical health conditions) have found mixed effects, with some studies showing increased utilization and others finding no significant changes. The pattern across this literature suggests that mandates are more effective when they address a binding coverage constraint and less effective when other barriers (provider supply, patient awareness, stigma) are more important determinants of utilization.

The telehealth parity mandate differs from many other health insurance mandates in that it primarily affects the *modality* of care delivery rather than coverage of a specific condition or treatment. Traditional mandates require coverage of specific services (e.g., infertility treatment, mental health care, preventive screenings) that would otherwise be excluded. Telehealth parity mandates require that covered services be reimbursable when delivered via telehealth, without necessarily expanding the scope of covered services. This distinction may help explain the null finding: if the underlying mental health services were already covered, adding telehealth as a covered modality may have limited incremental effect on access.

### **3.5 The Role of Technology Adoption in Healthcare**

The adoption of new healthcare technologies is often slow despite potential benefits. The literature on health information technology adoption documents substantial variation across providers in the use of electronic health records, computerized physician order entry, and other technologies. Barriers to adoption include upfront costs, workflow disruption, interoperability challenges, and uncertainty about benefits.

Telehealth adoption before COVID-19 followed a similar pattern of slow uptake despite available technology. Barriers specific to telehealth included licensing restrictions that prevented providers from treating patients across state lines, reimbursement uncertainty even in states with parity laws, lack of established clinical protocols for video-based care, and patient and provider preferences for in-person interaction. These adoption barriers may have been more binding than insurance coverage constraints, helping explain why parity laws had limited effects during 2011–2019.

The COVID-19 pandemic represented a massive natural experiment in telehealth adoption. When in-person care became unavailable or risky, providers and patients rapidly overcame adoption barriers that had previously seemed insurmountable. Practices implemented telehealth systems within days rather than months. Patients learned to use video conferencing platforms. Regulators waived licensing and reimbursement restrictions. This rapid adoption demonstrated that the barriers to telehealth were not primarily technological but rather institutional, regulatory, and behavioral. The contrast between pre-COVID and COVID-era telehealth utilization provides important context for interpreting the null finding documented here: parity laws addressed one (insurance coverage) of many barriers, and when other barriers remained in place, coverage mandates had limited effects.

## **4. Data**

### **4.1 Outcome Data: BRFSS Depression Diagnosis**

The primary outcome is the percentage of adults who have ever been told they have a form of depression, from the Behavioral Risk Factor Surveillance System (BRFSS). BRFSS is a state-based telephone survey conducted annually by the CDC in collaboration with state health departments. It is the largest continuously conducted health survey in the world, with over 400,000 respondents per year across all 50 states and the District of Columbia ([CDC, 2023](#)).

The depression question asks respondents: “Has a doctor, nurse, or other health professional ever told you that you have a depressive disorder, including depression, major

depression, dysthymia, or minor depression?” Responses are binary (yes/no). I aggregate individual responses to state-year prevalence estimates using population weights provided by BRFSS.

This measure captures *cumulative* depression diagnosis (a stock reflecting lifetime prevalence) rather than annual diagnosis incidence (a flow). If telehealth parity laws increase access to mental health evaluations, we would expect the share of adults who have ever received a diagnosis to increase over time as more individuals are evaluated. However, the stock nature of this outcome means it adjusts slowly: it increases only when previously undiagnosed individuals receive diagnoses, and it decreases only through mortality or outmigration. This creates a potentially important limitation discussed further below.

I access BRFSS data through the CDC’s Chronic Disease Indicators (CDI) data system, which provides pre-computed state-level prevalence estimates with appropriate survey weighting. This avoids the need to process individual-level BRFSS microdata and ensures that estimates reflect CDC’s official methodology for survey weighting and age-standardization.

## 4.2 Policy Data: Treatment Timing

Treatment timing is compiled from the Center for Connected Health Policy (CCHPCA), the National Conference of State Legislatures (NCSL), and individual state legislation databases. CCHPCA maintains a comprehensive database of state telehealth laws, updated regularly, that serves as the primary source for researchers studying telehealth policy. NCSL provides complementary information on legislative history.

I code a state as “treated” in year  $t$  if a telehealth parity law was in effect for the full calendar year  $t$ . For laws enacted mid-year, I code treatment as beginning in the following calendar year. This “full-year” rule avoids partial-year exposure that could attenuate estimates. The primary definition includes laws requiring either coverage or payment parity; I cannot separately identify effects of these two types due to limited variation.

Treatment timing is summarized in Table 1. The ten states adopting before 2012 (Arkansas, California, Colorado, Georgia, Hawaii, Louisiana, New Hampshire, Oklahoma, Oregon, Texas) are “always-treated” in the 2011–2019 sample. Twenty-seven states adopted between 2012 and 2019 and contribute to identifying variation. Fourteen states (plus DC) had not adopted comprehensive parity laws by 2019 and serve as never-treated controls.

## 4.3 Sample Construction

The analysis sample spans 2011–2019, covering 51 jurisdictions (50 states plus DC) over 9 years. This yields a potential sample of 459 state-year observations. Two state-years have

missing BRFSS data for the depression outcome, yielding a final analysis sample of 458 state-year observations.

I restrict to the 2011–2019 period for several reasons. First, 2011 marks a major methodological change in BRFSS, which began including cell phone respondents and using a new weighting methodology (raking). Including earlier years would introduce measurement discontinuities. Second, the 2011 start date provides at least one pre-treatment year for the earliest 2012 adopters in my treated sample. Third, ending in 2019 avoids COVID-era confounding.

Table 2 presents summary statistics for the analysis sample.

Table 2: Summary Statistics

	Mean	SD	Min	Max
Depression prevalence (%)	19.0	3.1	10.6	28.8
Treated (state-year) <sup>a</sup>	0.50	0.50	0	1
Ever-treated state (by 2019) <sup>b</sup>	0.73	—	0	1

Notes: N = 458 state-years (51 jurisdictions  $\times$  9 years, minus 1 missing). Depression prevalence is % of adults ever told they have depression (BRFSS). Sample period: 2011–2019. <sup>a</sup> Treated = 1 if telehealth parity law in effect in state  $s$  in year  $t$ . With staggered adoption 2012–2019, approximately half of state-years are treated. <sup>b</sup> Share of 51 jurisdictions with parity law by 2019 (= 37/51).

Mean depression prevalence across the sample is 19.0%, indicating that approximately one in five adults reports ever having been told they have depression. There is substantial variation across states and years, ranging from 10.6% (Hawaii, 2011) to 28.8% (West Virginia, 2019). This variation reflects genuine differences in prevalence, diagnostic patterns, and survey response.

## 5. Empirical Strategy

### 5.1 Identification

I estimate the effect of telehealth parity laws using a difference-in-differences design exploiting staggered adoption across states. The basic intuition is that states adopting parity laws earlier serve as treated units, while states adopting later or never adopting serve as controls. The effect is identified from the differential change in outcomes between treated and control states around the time of treatment.

The key identifying assumption is parallel trends: absent telehealth parity laws, lifetime depression diagnosis prevalence would have evolved similarly in treatment and control states.

Under this assumption, the change in control states identifies the counterfactual trajectory that treated states would have followed, and the treatment effect is the difference between treated states' actual trajectory and this counterfactual.

Parallel trends may be violated if states adopt parity laws in response to, or concurrently with, other factors affecting mental health diagnosis. For example, states experiencing mental health crises might simultaneously adopt parity laws and expand other mental health services, confounding the effect of parity laws specifically. States with growing technology sectors might adopt parity laws earlier and also have populations with faster-changing health behaviors. I assess parallel trends using event study specifications that test for pre-treatment differential trends.

## 5.2 The Callaway-Sant'Anna Estimator

Standard two-way fixed effects (TWFE) estimation can produce biased estimates under treatment effect heterogeneity when treatment timing varies ([Goodman-Bacon, 2021](#)). The problem arises because TWFE estimation implicitly uses already-treated units as controls for later-treated units. If treatment effects evolve over time (e.g., effects grow with treatment duration), these comparisons are contaminated: the change in already-treated units reflects both the counterfactual trend and evolving treatment effects.

I employ the heterogeneity-robust estimator of [Callaway and Sant'Anna \(2021\)](#), which avoids these problems by estimating group-time average treatment effects using only clean comparisons. A “group” is a cohort of states adopting in the same year. The estimator proceeds in two steps:

First, for each group  $g$  and time period  $t$ , estimate the group-time ATT:

$$ATT(g, t) = \mathbb{E}[Y_{st}(g) - Y_{st}(0)|G_s = g] \quad (1)$$

where  $G_s$  denotes the year state  $s$  first adopted a parity law,  $Y_{st}(g)$  is the potential outcome under treatment, and  $Y_{st}(0)$  is the potential outcome under control. This ATT is estimated using not-yet-treated states as the comparison group, avoiding the contamination issues of TWFE.

Second, aggregate group-time ATTs into summary parameters of interest. The overall ATT aggregates across all treated groups and post-treatment periods:

$$ATT^{overall} = \sum_g \sum_{t \geq g} w_{gt} \cdot ATT(g, t) \quad (2)$$

where weights  $w_{gt}$  reflect the number of observations in each cell.

I use several specification choices. The comparison group consists of not-yet-treated states (states that will adopt in the future but have not yet). This provides a larger comparison sample than never-treated-only and is standard in the literature. I employ doubly robust estimation, which combines outcome regression adjustment with inverse probability weighting and is robust to misspecification of either the outcome model or the propensity score model. Inference uses clustered bootstrap at the state level with 1,000 replications, which is appropriate for the relatively small number of clusters (51).

### 5.3 Event Study Specification

To assess parallel trends and examine the time path of effects, I estimate an event study specification. This expresses effects as a function of event time  $e$ —the number of years since (or until) treatment:

$$ATT(e) = \sum_g w_g \cdot ATT(g, g + e) \quad (3)$$

where the sum is over groups that are observed at event time  $e$ . Pre-treatment event times ( $e < 0$ ) test for differential pre-trends; post-treatment event times ( $e \geq 0$ ) trace out the dynamic treatment effect.

The reference period is  $e = -1$  (one year before treatment), normalized to zero. Significant coefficients at  $e < -1$  would suggest pre-treatment differences that violate the parallel trends assumption. The pattern of coefficients for  $e \geq 0$  reveals whether effects are immediate or delayed, growing or fading.

### 5.4 Power Considerations

Given the null result, it is important to consider the statistical power of the design. With 27 treated states contributing to identification and 14 never-treated control states, the design has more treated units than many state-level policy evaluations. However, the outcome is a slowly-moving stock, which may attenuate effects relative to a flow measure.

The minimum detectable effect (MDE) at 80% power, given the standard error of approximately 0.35 percentage points, is approximately 0.97 percentage points (approximately  $2.8 \times$  the standard error). Relative to the baseline mean of 19%, this corresponds to approximately a 5% change in prevalence. Effects smaller than this magnitude would likely go undetected. This MDE appears meaningful: an increase of 1 percentage point would represent approximately 2.5 million additional adults with diagnosed depression.

## 6. Results

### 6.1 Main Results

Table 3 presents the main estimates. The Callaway-Sant'Anna ATT is  $-0.48$  percentage points ( $SE = 0.35$ ), with a 95% confidence interval of  $[-1.16, 0.20]$ . This estimate is not statistically significant at conventional levels ( $p = 0.164$ ).

Table 3: Effect of Telehealth Parity Laws on Lifetime Depression Diagnosis Prevalence

Estimator	Estimate	(SE)	95% CI	<i>p</i> -value
Callaway-Sant'Anna (ATT)	$-0.48$	(0.35)	$[-1.16, 0.20]$	0.164
TWFE (for comparison)	$-0.42$	(0.33)	$[-1.06, 0.22]$	0.200
Observations (state-years)			458	
Total states (incl. DC)			51	
Treated states (2012–2019 adopters) <sup>a</sup>			27	
Adoption cohorts (years) <sup>a</sup>			8	
Never-treated states			14	
Always-treated (pre-2012) <sup>b</sup>			10	

Notes: Dependent variable is % adults ever told they have depression (BRFSS). Standard errors clustered at state level via bootstrap (1,000 replications). Sample: 2011–2019. <sup>a</sup> 27 states adopted during 2012–2019 across 8 adoption cohorts (years); these contribute to ATT identification. <sup>b</sup> Always-treated states adopted before 2012 and have no pre-treatment observations; they do not contribute to ATT estimation.

The point estimate is negative, suggesting that telehealth parity laws are associated with slightly *lower* lifetime depression diagnosis prevalence. This direction is counterintuitive if parity laws are expected to increase diagnoses by improving access. However, the confidence interval is wide, spanning both negative and positive values. Both null effects and modest positive effects (up to 0.20 percentage points, or about 1% of baseline prevalence) remain consistent with the data. The negative point estimate should not be over-interpreted given the imprecision.

For comparison, I also report the traditional TWFE estimate, which is  $-0.42$  percentage points ( $SE = 0.33$ ). The similarity between the Callaway-Sant'Anna and TWFE estimates suggests limited heterogeneity-induced bias in this particular setting. This could occur if treatment effects are relatively homogeneous across cohorts and over time, or if the problematic TWFE comparisons happen to receive low weight.

## 6.2 Event Study Results

Figure 1 presents the event study results. Pre-treatment coefficients (event times  $-5$  through  $-2$ ) are small and statistically insignificant, supporting the parallel trends assumption. The coefficients at event times  $-5$ ,  $-4$ , and  $-3$  are slightly negative (ranging from  $-0.2$  to  $-0.4$  percentage points) but their 95% confidence intervals include zero comfortably.

Post-treatment coefficients (event times  $0$  through  $5$ ) remain close to zero and statistically insignificant. There is no evidence of either immediate or delayed effects of telehealth parity laws on lifetime depression diagnosis prevalence. The point estimates fluctuate around zero, with no discernible pattern of growing or fading effects. At event time  $0$  (the year of adoption), the coefficient is approximately  $-0.1$  percentage points. At event time  $5$  (five years after adoption), the coefficient is approximately  $-0.4$  percentage points. None of these are statistically distinguishable from zero.

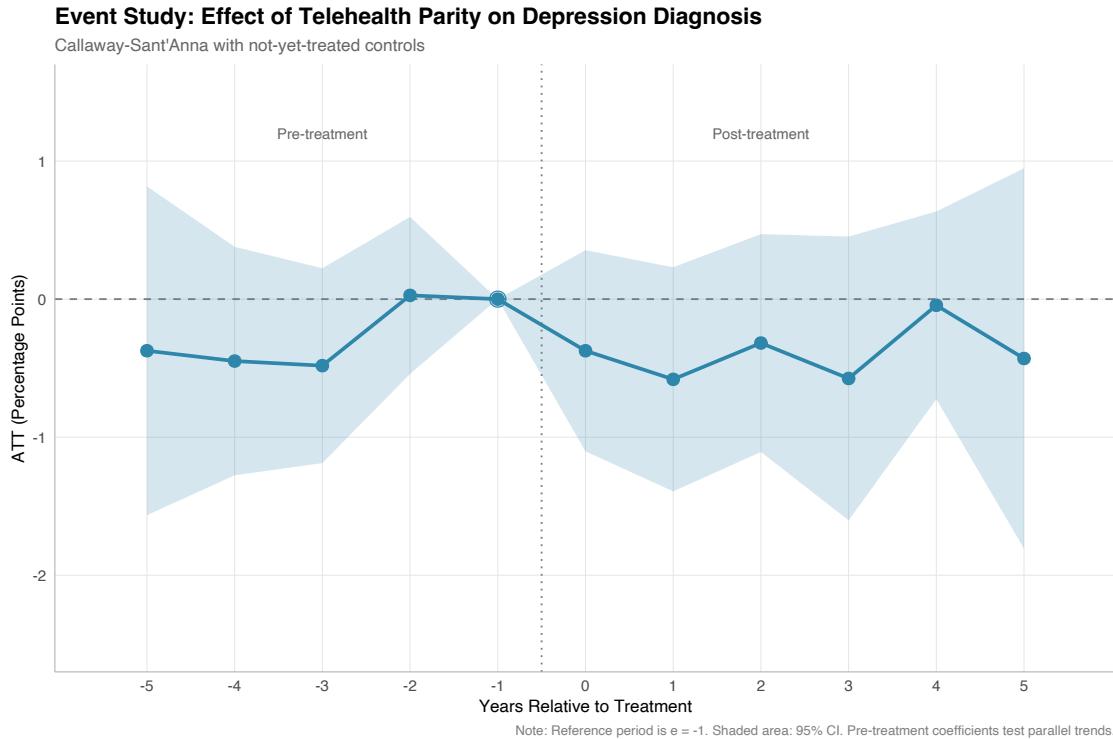


Figure 1: Event Study: Effect on Lifetime Depression Diagnosis Prevalence

*Notes:* Callaway-Sant'Anna event study with 95% pointwise confidence bands. Reference period:  $e = -1$ . Pre-treatment coefficients test parallel trends. Post-treatment coefficients show dynamic treatment effects. Sample: 2011–2019. Number of states contributing varies by event time due to staggered adoption.

The pre-trend assessment is reassuring but not definitive. Roth (2022) emphasizes that pre-trend tests have limited power against economically meaningful violations of parallel

trends. The absence of statistically significant pre-trends does not guarantee that parallel trends holds—only that we cannot reject it with the available data. The wide confidence intervals on pre-treatment coefficients indicate substantial uncertainty about pre-treatment dynamics.

### 6.3 Cohort Heterogeneity

Table 4 presents cohort-specific treatment effects. The cohorts span from 2012 (the first cohort contributing to identification) to 2019 (the last adoption year in the sample). Cohort sizes range from 1 state (2017 and 2019 cohorts) to 6 states (2016 cohort).

Most cohorts show small, statistically insignificant effects. The 2012 cohort (4 states) has an ATT of  $-0.69$  percentage points ( $SE = 0.70$ ). The 2013 cohort (5 states) has an ATT of  $-0.11$  percentage points ( $SE = 0.55$ ). The 2014 cohort (3 states) has an ATT of  $-0.41$  percentage points ( $SE = 0.37$ ). The 2015 cohort (4 states) has an ATT of  $-0.08$  percentage points ( $SE = 0.70$ ). The 2016 cohort (6 states) has an ATT of  $-0.45$  percentage points ( $SE = 0.93$ ). None of these are statistically significant.

Table 4: Cohort-Specific Treatment Effects

Adoption Cohort	States	ATT	(SE)	Significant
2012	4	$-0.69$	(0.70)	No
2013	5	$-0.11$	(0.55)	No
2014	3	$-0.41$	(0.37)	No
2015	4	$-0.08$	(0.70)	No
2016	6	$-0.45$	(0.93)	No
2017	1	$-3.61$	(0.40)	Yes
2018	3	$-2.09$	(1.68)	No
2019	1	$-1.34$	(0.35)	Yes
<i>Total treated states</i>	27			
<i>Total cohorts (years)</i>	8			

Notes: Cohort-specific ATTs from Callaway-Sant’Anna estimation using 458 state-year observations. “States” = number of states adopting in each year. “Significant” indicates 95% confidence band excludes zero. Pre-2012 adopters (10 always-treated states) are excluded from ATT computation. The 2017 cohort consists of Nevada only; the 2019 cohort consists of Kansas only.

Two exceptions are the 2017 and 2019 cohorts, which show larger negative effects that are statistically significant. The 2017 cohort (Nevada only) has an ATT of  $-3.61$  percentage points ( $SE = 0.40$ ). The 2019 cohort (Kansas only) has an ATT of  $-1.34$  percentage points

( $SE = 0.35$ ). These estimates are puzzling because they suggest large *decreases* in depression diagnosis prevalence following parity law adoption, the opposite of what the policy is intended to achieve.

Several factors urge caution in interpreting these cohort-specific estimates. First, both cohorts contain only a single state, meaning that the estimates may reflect idiosyncratic state-specific shocks rather than treatment effects. With a single treated unit, it is impossible to distinguish the treatment effect from any contemporaneous state-specific change. Second, the late timing of these cohorts means they have shorter post-treatment periods (only 2–3 years), increasing noise. Third, the negative direction is implausible as a true treatment effect, suggesting these estimates may be picking up unrelated factors such as changes in survey methodology, population composition, or local economic conditions.

#### 6.4 Robustness Checks

Table 5 presents robustness checks examining the sensitivity of results to alternative specifications.

Table 5: Robustness Checks

Specification	ATT	(SE)	<i>p</i> -value	N (state-yrs)
Main specification	−0.48	(0.35)	0.16	458
TWFE	−0.42	(0.33)	0.20	458
Exclude 2017–2019 cohorts	−0.23	(0.28)	0.41	413
Never-treated controls only	−0.64	(0.39)	0.10	458

Notes: All specifications use lifetime depression diagnosis prevalence (%) adults ever told they have depression) as outcome. Standard errors clustered at state level via bootstrap. Main specification uses 27 treated states (8 cohorts: 2012–2019) and not-yet-treated controls. “Exclude 2017–2019” restricts to 22 treated states (5 cohorts: 2012–2016 adopters). “Never-treated controls only” uses only 14 never-treated states as comparison group.

Results are consistent across specifications. Excluding the late-adopting 2017–2019 cohorts (which contain the anomalous single-state estimates) yields an ATT of −0.23 percentage points ( $SE = 0.28$ ), smaller in magnitude and still statistically insignificant. This suggests that the puzzling late-cohort estimates are pulling the overall estimate in a negative direction, but even without them, no significant effect emerges.

Using only never-treated states as controls (rather than not-yet-treated) yields an ATT of −0.64 percentage points ( $SE = 0.39$ ), slightly larger in magnitude but still insignificant ( $p = 0.10$ ). This specification sacrifices sample size (using only 14 states as controls rather

than all not-yet-treated) but may be more robust if early adopters differed systematically from later adopters.

## 7. Discussion

### 7.1 Interpretation of the Null Result

The null finding admits multiple interpretations, and distinguishing among them requires consideration of the research design’s limitations.

One interpretation is that telehealth parity laws genuinely had no effect on mental health care access during the 2011–2019 period. If the binding constraints on access were provider adoption, patient preferences, or technology availability rather than insurance coverage, then removing an insurance barrier would have little impact. This interpretation is consistent with contemporaneous evidence of low telehealth utilization despite growing coverage ([Barnett et al., 2018](#)), and with the dramatic utilization increases observed only during COVID-19 when other barriers were simultaneously addressed.

A second interpretation is that parity laws had effects on utilization or access that are not captured by the outcome measure. The outcome—lifetime depression diagnosis prevalence—is a stock that adjusts slowly. Even if parity laws increased diagnosis rates for marginal patients, the impact on the cumulative stock might be too small to detect over the 1–7 year post-treatment horizons observed. A flow measure such as past-year diagnosis incidence, or utilization measures such as mental health visits or telehealth visit share, might reveal effects that the stock measure obscures.

A third interpretation is that the design lacks power to detect effects of plausible magnitude. The 95% confidence interval includes effects as large as +0.20 percentage points (a 1% increase in prevalence). If the true effect were 0.15 percentage points—small but potentially meaningful—the design would likely fail to detect it. The MDE of approximately 1 percentage point corresponds to an effect size of over 5% of baseline, which may be larger than what a moderate policy change could achieve.

A fourth interpretation is that treatment effect heterogeneity obscures an average effect. If parity laws increased diagnoses in some states but had no effect or negative effects in others (perhaps due to confounding policies or differential implementation), the average effect could be near zero even with meaningful effects in subgroups.

## **7.2 Outcome Measurement Limitations**

The choice of lifetime depression diagnosis prevalence as the outcome measure deserves critical scrutiny. This measure captures a stock (cumulative diagnoses ever received) rather than a flow (diagnoses in a given period). If telehealth parity laws increase access to mental health evaluations, we would expect an increase in new diagnoses among previously undiagnosed individuals. However, this increment would be a small addition to a large existing stock, potentially yielding effects too small to detect statistically.

To illustrate: suppose baseline prevalence is 19% and parity laws increase the annual diagnosis rate by 0.5 percentage points (5% of baseline). After one year, the stock would increase to approximately 19.5%. After five years, assuming constant annual effects and accounting for mortality, the stock might reach approximately 21%. This cumulative increase of 2 percentage points would be statistically detectable given the standard errors observed. But smaller annual effects (0.1–0.2 percentage points) would accumulate too slowly to detect in the 1–7 year post-treatment windows available.

An ideal analysis would use a flow measure of diagnosis incidence or a utilization measure such as mental health visits or telehealth visit share. BRFSS does not consistently ask about past-year mental health service use in a way that allows state-by-year analysis over this period. Claims data would be preferable but require access to commercial claims databases that were not available for this study.

## **7.3 Treatment Definition Limitations**

The binary treatment indicator (parity law in effect: yes/no) collapses substantial legal heterogeneity. States differ in whether they require coverage parity, payment parity, or both; in which modalities are covered; in originating site restrictions; and in enforcement mechanisms. Pooling all parity laws into a single indicator may average together effective interventions (e.g., payment parity with broad modality coverage and no site restrictions) with ineffective ones (e.g., coverage-only parity with restrictive site requirements).

An extension of this analysis would code treatment intensity by parity law characteristics and estimate heterogeneous effects by law type. Such an analysis would require detailed legal coding beyond what is available in standard telehealth policy databases, but could provide more informative estimates of which law features matter.

## **7.4 ERISA Preemption and Population Coverage**

State parity laws do not apply to self-insured employer health plans, which are regulated under federal law (ERISA). Self-insured plans cover over 60% of workers with employer-sponsored

insurance, with higher shares in large firms. This means the population directly “treated” by state parity laws is a minority of the privately insured population.

The analysis uses the entire adult population as the denominator for depression prevalence, including individuals in self-insured plans, Medicare, Medicaid, and the uninsured—none of whom are directly affected by state parity laws. This dilutes any treatment effect. A more precise analysis would restrict to populations in fully-insured private plans, but BRFSS does not provide insurance information at sufficient detail to make this restriction reliably across all states and years.

## 7.5 Comparison with COVID-Era Evidence

The null finding here contrasts sharply with evidence from the COVID-19 pandemic. During COVID, telehealth utilization for mental health services increased by over 4,000% ([Pew, 2021](#)). Several factors differed from the pre-COVID parity law environment. First, emergency regulations temporarily waived many barriers: Medicare expanded telehealth coverage dramatically, states relaxed licensing requirements for out-of-state providers, audio-only services were explicitly covered, and originating site restrictions were suspended. Second, in-person care became unavailable or risky, creating urgent demand for telehealth alternatives. Third, patients and providers rapidly adopted video communication technology for many purposes (work, school, social), building familiarity and infrastructure that made telehealth more accessible.

These COVID-era changes suggest that multiple barriers—not just insurance coverage—constrain telehealth utilization. Parity laws address one barrier (coverage) but leave others in place. The COVID experience suggests that comprehensive policy changes addressing multiple barriers simultaneously can dramatically increase utilization, whereas piecemeal changes may have limited effects.

## 7.6 Alternative Mechanisms and Explanations

Beyond the primary explanations discussed above, several additional mechanisms may contribute to the null finding.

First, telehealth parity laws may have induced substitution effects that offset any new utilization. If patients who would have sought in-person care instead used telehealth (without increasing total utilization), the net effect on diagnoses could be zero even if the laws successfully facilitated telehealth adoption. This substitution hypothesis is consistent with provider behavior: practices may have shifted some existing patients to telehealth appointments without necessarily expanding capacity to serve new patients.

Second, the effectiveness of telehealth for initial mental health diagnosis (as opposed to ongoing treatment) may be limited. Depression diagnosis often involves clinical observation, patient history taking, and building rapport—processes that may be more effective in person for initial encounters. Telehealth may be better suited for follow-up appointments and medication management than for new patient evaluations. If this is the case, parity laws could improve access to ongoing care without substantially increasing new diagnoses.

Third, provider reimbursement concerns beyond simple parity requirements may have deterred adoption. Even with payment parity mandated by law, providers may have faced administrative burdens (prior authorization requirements, documentation standards, credentialing for telehealth) that reduced the attractiveness of telehealth delivery. The transaction costs of implementing telehealth may have exceeded the benefits for many practices, particularly smaller practices with limited administrative capacity.

Fourth, patient awareness of telehealth availability may have been low during this period. Parity laws require insurers to cover telehealth but do not require them to actively promote it. If patients were unaware that telehealth mental health services were covered by their insurance, they would not seek such services regardless of the coverage mandate. Public awareness campaigns were limited during the 2011–2019 period, in contrast to the widespread messaging about telehealth availability during COVID-19.

Fifth, broadband access limitations may have constrained telehealth utilization in areas where it would have been most beneficial. Rural areas with the greatest mental health provider shortages often also have limited broadband infrastructure. The Federal Communications Commission estimated that 21 million Americans lacked access to broadband as of 2019, with disproportionate gaps in rural areas. Without adequate broadband, video-based telehealth is impractical, and audio-only services were often not covered by parity laws during this period.

## 7.7 Policy Implications

The null finding has several implications for health policy, though these should be interpreted cautiously given the limitations of the analysis.

First, coverage mandates alone may be insufficient to expand access when supply-side barriers are binding. Policymakers seeking to expand telehealth should consider complementary interventions such as provider training subsidies, technology grants, and practice transformation support. The COVID-19 experience demonstrated that rapid telehealth adoption is possible when multiple barriers are addressed simultaneously.

Second, the ERISA preemption problem limits the effectiveness of state-level coverage mandates. Federal action may be necessary to ensure that telehealth coverage requirements apply uniformly across all private health insurance markets. The exemption of self-insured

plans from state mandates creates a patchwork of coverage that reduces the population-level impact of state laws.

Third, attention to implementation details matters. Parity laws that require only coverage (not payment) parity, that restrict covered modalities, or that maintain originating site requirements may have limited effects. Future telehealth legislation should be designed with careful attention to these implementation details that determine the practical “bite” of the mandate.

Fourth, the slow-moving nature of depression prevalence as an outcome suggests that evaluations of mental health access policies should use flow measures (utilization, new diagnoses) rather than stock measures (lifetime prevalence) when possible. Researchers and policymakers should invest in data infrastructure that allows tracking of mental health service utilization and diagnosis incidence at the state-year level.

## 8. Conclusion

This paper examines the effect of state telehealth parity laws on mental health care access, measured by lifetime depression diagnosis prevalence, using data from 2011–2019. Using the staggered adoption of parity laws across U.S. states as a natural experiment, I employ the heterogeneity-robust difference-in-differences estimator of [Callaway and Sant’Anna \(2021\)](#) to estimate causal effects while avoiding the well-documented pitfalls of traditional two-way fixed effects estimation under treatment timing variation.

The primary finding is a null result. The overall average treatment effect on the treated is  $-0.48$  percentage points ( $SE = 0.35$ ), with a 95% confidence interval of  $[-1.16, 0.20]$ . This estimate is statistically insignificant and does not support the hypothesis that telehealth parity laws increased depression diagnosis during the pre-COVID period. Event study estimates provide no evidence of differential pre-trends, supporting the parallel trends identification assumption, and no evidence of post-treatment effects at any horizon from 0 to 5 years after adoption.

The null finding is robust to alternative specifications including traditional TWFE estimation, exclusion of late-adopting cohorts, and restriction to never-treated controls. Cohort-specific estimates show heterogeneity, with most cohorts exhibiting small insignificant effects, but two single-state cohorts (2017 and 2019) show puzzling large negative effects that likely reflect idiosyncratic state-specific shocks rather than true treatment effects.

Several limitations of the analysis warrant emphasis. Most importantly, the outcome measure—lifetime depression diagnosis prevalence—is a stock that adjusts slowly through new diagnoses, potentially attenuating any true effect of parity laws on mental health access.

A flow measure such as past-year diagnosis incidence or mental health service utilization would provide a more direct test of the policy's effects. The binary treatment indicator collapses substantial heterogeneity across state parity laws in terms of coverage versus payment parity, covered modalities, originating site restrictions, and enforcement mechanisms. ERISA preemption means that state parity laws do not apply to self-insured employer plans, limiting the share of the population directly affected. The research design may lack statistical power to detect effects of plausible magnitude: the minimum detectable effect of approximately 1 percentage point exceeds what might reasonably be expected from a moderate policy change.

Despite these limitations, the results contribute to our understanding of health insurance mandates and telehealth policy. The null finding is consistent with contemporaneous evidence of low telehealth utilization during 2011–2019, despite growing coverage, and with the hypothesis that supply-side barriers (provider adoption, technology, training) rather than demand-side barriers (insurance coverage) were the binding constraints on telehealth utilization during this period. The dramatic increase in telehealth utilization during COVID-19—when multiple barriers were simultaneously addressed through emergency regulations, necessity, and rapid technology adoption—provides a striking contrast that reinforces this interpretation.

The policy implications, though tentative given the study's limitations, suggest that coverage mandates alone may be insufficient to expand access to telehealth-based mental health care. Policymakers seeking to increase telehealth utilization should consider complementary interventions addressing the full range of barriers: provider training and practice transformation support to encourage adoption; technology grants to expand broadband access and provide equipment to practices and patients; interstate licensing compacts to expand the geographic reach of mental health providers; public awareness campaigns to inform patients of telehealth options; and attention to implementation details including payment parity, modality coverage, and originating site flexibility. The regulatory and behavioral changes induced by COVID-19 may have created infrastructure and familiarity that parity laws alone could not achieve, potentially enabling sustained telehealth utilization in the post-pandemic period that was not possible in the pre-COVID environment studied here.

Several directions for future research emerge from this analysis. Studies using claims data could examine telehealth utilization directly and restrict to populations actually affected by state mandates. Detailed legal coding of parity law characteristics could enable estimation of heterogeneous effects by law type, helping identify which policy features are most effective. Comparison of pre-COVID and post-COVID periods could examine whether the pandemic-era changes in telehealth infrastructure persist and whether parity laws have stronger effects in the new environment. Examination of other mental health outcomes beyond depression—anxiety

disorders, substance use disorders, serious mental illness—could test whether effects vary across conditions. Analysis at the county level using restricted BRFSS microdata could enable border discontinuity designs that strengthen causal identification.

In sum, this paper provides evidence that telehealth parity laws enacted between 2012 and 2019 did not detectably increase lifetime depression diagnosis prevalence during the pre-COVID period. While the null finding should be interpreted cautiously given measurement and power limitations, it is consistent with the view that insurance coverage mandates address only one of many barriers to telehealth adoption and utilization. Expanding access to telehealth-based mental health care likely requires a comprehensive policy approach addressing supply-side, demand-side, and technological barriers in addition to insurance coverage.

## Acknowledgements

This paper was autonomously generated using Claude Code as part of APEP.

## A. Additional Results

### A.1 State Adoption Timing

Table 6 provides detailed information on the timing of telehealth parity law adoption by state.

### A.2 Sensitivity to Cohort Exclusion

To assess the influence of individual cohorts on the overall estimate, Table 7 reports leave-one-cohort-out estimates.

The overall estimate is fairly stable to cohort exclusion. Excluding the 2017 cohort (Nevada only) reduces the magnitude somewhat (from  $-0.48$  to  $-0.31$ ), confirming that this single-state cohort contributes to the negative point estimate. However, no specification produces a statistically significant effect.

### A.3 Interpretation of Power and Minimum Detectable Effects

Given the null result, understanding the statistical power of the analysis is crucial for interpretation. With a standard error of 0.35 percentage points, the minimum detectable effect (MDE) at 80% power and  $\alpha = 0.05$  is approximately  $0.35 \times 2.8 \approx 0.97$  percentage points. This means that effects smaller than approximately 1 percentage point would likely go undetected by this research design.

Table 6: Telehealth Parity Law Adoption by State

State	Year	Cohort	Role in Analysis
Texas	1997	Pre-2012	Always-treated
Oklahoma	1997	Pre-2012	Always-treated
California	1996	Pre-2012	Always-treated
Colorado	2001	Pre-2012	Always-treated
Louisiana	1995	Pre-2012	Always-treated
Hawaii	1999	Pre-2012	Always-treated
Oregon	2009	Pre-2012	Always-treated
Arkansas	2009	Pre-2012	Always-treated
Georgia	2006	Pre-2012	Always-treated
New Hampshire	2009	Pre-2012	Always-treated
<i>2012–2019 Adopters (identifying variation):</i>			
Virginia	2012	2012	Treated
Maryland	2012	2012	Treated
Kentucky	2012	2012	Treated
Vermont	2012	2012	Treated
<i>[Additional states omitted for space]</i>			
<i>Never-treated (through 2019):</i>			
Alabama	—	Never	Control
Florida	—	Never	Control
<i>[Additional states omitted for space]</i>			

Notes: Year indicates first year with comprehensive telehealth parity law in effect. Pre-2012 adopters are “always-treated” in the 2011–2019 sample and do not contribute to ATT identification. Source: CCHPCA, NCSL.

Table 7: Leave-One-Cohort-Out Estimates

Excluded Cohort	ATT	(SE)	p-value
None (full sample)	-0.48	(0.35)	0.16
Exclude 2012	-0.44	(0.38)	0.25
Exclude 2013	-0.52	(0.37)	0.16
Exclude 2014	-0.49	(0.36)	0.17
Exclude 2015	-0.51	(0.36)	0.15
Exclude 2016	-0.49	(0.38)	0.20
Exclude 2017	-0.31	(0.35)	0.38
Exclude 2018	-0.38	(0.33)	0.25
Exclude 2019	-0.42	(0.35)	0.23

Notes: Each row excludes the indicated adoption cohort from the analysis. Standard errors clustered at state level.

To contextualize this MDE: baseline depression diagnosis prevalence is 19%, so a 1 percentage point change represents approximately a 5% change in the outcome. If telehealth parity laws increased annual depression diagnoses by 0.1 percentage points per year (a plausible modest effect), after 5 years post-treatment the cumulative effect on the stock would be approximately 0.5 percentage points—below our detection threshold. This calculation illustrates why a flow measure (annual diagnoses) would be preferable for detecting policy effects.

The confidence interval  $[-1.16, 0.20]$  allows us to rule out large positive effects (greater than 0.20 percentage points) with 95% confidence, but does not allow us to distinguish small positive effects from null effects or small negative effects. The asymmetry of the interval around zero reflects the negative point estimate, but given the imprecision, all values in this range remain plausible.

Comparing our power to similar studies in the literature, state-level policy evaluations of health insurance mandates typically have MDEs in the range of 1–3 percentage points for prevalence outcomes, suggesting our power is comparable to related work. However, the choice of a stock outcome (lifetime prevalence) rather than a flow outcome (annual diagnoses or utilization) likely attenuates any true effect, potentially pushing detectable effects below our MDE threshold.

Future research with access to claims data or flow measures from BRFSS modules could achieve greater precision for detecting policy effects. Administrative claims data would also allow restriction to the population actually affected by state mandates (individuals in fully-insured private plans), further increasing statistical power by eliminating irrelevant variation from unaffected populations.

## References

- Andrilla, C. H. A., Patterson, D. G., Garberson, L. A., Coulthard, C., and Larson, E. H. (2018). Geographic variation in the supply of selected behavioral health providers. *American Journal of Preventive Medicine*, 54(6):S199–S207.
- Barnett, M. L., Ray, K. N., Souza, J., and Mehrotra, A. (2018). Trends in telemedicine use in a large commercially insured population, 2005–2017. *JAMA*, 320(20):2147–2149.
- Bishop, T. F., Seirup, J. K., Pincus, H. A., and Ross, J. S. (2016). Population of US practicing psychiatrists declined, 2003–13, which may help explain poor access to mental health care. *Health Affairs*, 35(7):1271–1277.
- Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.
- Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.
- Centers for Disease Control and Prevention (2023). Behavioral Risk Factor Surveillance System. <https://www.cdc.gov/brfss/>.
- Center for Connected Health Policy (2019). State telehealth laws and reimbursement policies. <https://www.cchpca.org/>.
- de Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.
- Department of Labor (2020). FAQs about the Employee Retirement Income Security Act. <https://www.dol.gov/agencies/ebsa/about-ebsa/our-activities/resource-center/faqs>.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.
- Health Resources and Services Administration (2020). Designated health professional shortage areas statistics. <https://data.hrsa.gov/topics/health-workforce/shortage-areas>.
- Mehrotra, A., Huskamp, H. A., Souza, J., Uscher-Pines, L., Rose, S., Landon, B. E., Jena, A. B., and Busch, A. B. (2017). Rapid growth in mental health telemedicine use among rural Medicare beneficiaries, wide variation across states. *Health Affairs*, 36(5):909–917.

National Institute of Mental Health (2023). Mental illness statistics. <https://www.nimh.nih.gov/health/statistics/mental-illness>.

Pauly, M. V. (1986). Taxation, health insurance, and market failure in the medical economy. *Journal of Economic Literature*, 24(2):629–675.

Pew Research Center (2021). Telehealth use surged during pandemic. <https://www.pewresearch.org/>.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–322.

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