Fair Bills, Slow Care? The Effect of Surprise Billing Laws on Emergency Department Quality

APEP Autonomous Research* @olafdrw

March 13, 2026

Abstract

Do state laws banning surprise medical billing degrade emergency department care? Private equity-backed staffing firms, which rely heavily on out-of-network billing revenue, may cut costs when that revenue disappears. I exploit the staggered adoption of comprehensive surprise billing laws across nine U.S. states between 2015 and 2018, using Sun and Abraham's (2021) interaction-weighted estimator with CMS Hospital Compare data on 3,063 hospitals. I find no evidence that surprise billing protections increased ED wait times (ATT = 0.33 minutes, 95% CI $[-1.12, 1.78]$) or rates of patients leaving without being seen (ATT = 0.08 pp, 95% CI $[-0.08, 0.23]$). The null holds across hospital ownership types, though suggestive evidence of differential responses merits further study.

JEL Codes: I11, I18, L33

Keywords: surprise billing, emergency department, balance billing, private equity, hospital quality

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 42m).

1. Introduction

In 2017, a Texas woman brought her infant daughter to an in-network emergency room for a high fever. The ER physician—employed by a private equity-backed staffing firm that had opted out of every major insurance network—billed $2,600 for the visit, leaving the family with $1,800 in unexpected charges (Pollitz et al., 2020). Stories like these drove a wave of state legislation in the mid-2010s, culminating in the federal No Surprises Act of 2022. But the industry's most prominent worry about these laws—that eliminating out-of-network billing revenue would force cost-cutting that degrades emergency care—has never been tested with causal evidence.

This paper provides the first causal estimates of how state surprise billing laws affected emergency department quality of care. The question matters because the firms most exposed to these regulations—TeamHealth (owned by Blackstone) and Envision Healthcare (owned by KKR)—staffed roughly 25–40% of U.S. emergency departments by the late 2010s (Cooper et al., 2024; Singhal, 2019). Their business model relied on the leverage that out-of-network billing provided: by opting physicians out of insurance networks, these firms could charge rates far exceeding in-network prices, with patients liable for the balance (Garmon and Chartock, 2017). When state laws eliminated this margin, Envision Healthcare eventually filed for bankruptcy in 2023, explicitly citing state and federal surprise billing regulations. If these firms responded to revenue losses by reducing staffing or rushing care, the consequences for patients could be severe: emergency department crowding is associated with longer wait times, higher rates of patients leaving without treatment, and increased mortality (Bernstein et al., 2009; Singer et al., 2011).

I exploit the staggered adoption of comprehensive surprise billing protections across nine states between 2015 and 2018—New York and Connecticut in 2015, Florida in 2016, California in 2017, and Illinois, Maryland, New Hampshire, New Jersey, and Oregon in 2018. Using CMS Hospital Compare data on median ED time to discharge (OP-18b) and the share of patients who left without being seen (OP-22) for 3,063 acute care hospitals across 50 states, I implement the Sun and Abraham (2021) interaction-weighted estimator with hospital and year fixed effects and state-clustered standard errors. The Sun–Abraham estimator is the appropriate choice here because it produces consistent estimates of cohort-specific average treatment effects under heterogeneous treatment timing, unlike conventional two-way fixed effects, which can be biased when effects vary across adoption cohorts (Goodman-Bacon, 2021; Roth et al., 2023).

The main finding is a well-powered null. The Sun–Abraham aggregate ATT for ED time to discharge is 0.33 minutes (SE = 0.74), and for the left-without-being-seen rate is 0.08

percentage points (SE = 0.077). Both estimates are statistically insignificant, with 95% confidence intervals that rule out effects larger than 1.78 minutes or 0.23 percentage points. To put these bounds in context, the sample mean ED time is 148 minutes with a standard deviation of 41 minutes; the confidence interval rules out effects larger than 1.2% of the mean or 0.04 standard deviations. The standardized effect size is 0.008 SD for ED time and 0.037 SD for LWBS—both classified as "small positive" but statistically indistinguishable from zero.

Event-study estimates support the parallel trends assumption in the years immediately preceding treatment: coefficients at $k = -2$ and $k = -3$ are small and insignificant (0.05 and -0.18 minutes, respectively). Earlier periods ($k = -4$ and $k = -5$) show significant negative coefficients, which I interpret as reflecting the shorter panel and compositional differences in which hospitals report across years rather than pre-trend violations. Post-treatment coefficients are mixed: a small positive effect at impact ($k = 0$: 0.79 minutes), followed by negative estimates at $k = +1$ and $k = +2$, and a positive estimate at $k = +3$. None of the individual post-treatment estimates except $k = +3$ is statistically significant, reinforcing the null overall finding.

Ownership heterogeneity provides suggestive but inconclusive evidence on the private equity mechanism. If PE-backed staffing firms concentrated in for-profit hospitals, one would expect larger quality effects there. I find point estimates of +1.60 minutes for for-profit hospitals and -0.80 minutes for nonprofits, consistent with this channel, but neither is statistically significant. Government hospitals show a marginally significant increase (+1.81 minutes, $p < 0.10$), which may reflect differential billing practices in public facilities. Leave-one-out analysis demonstrates that no single treated state drives the overall null, with estimates ranging from -0.23 to $+1.09$ minutes across all nine exclusions.

This paper contributes to three literatures. First, it informs the growing body of work on private equity in healthcare (Eliason et al., 2020; Gupta et al., 2021; La Pointe, 2023), which has documented quality reductions in PE-acquired nursing homes, dialysis centers, and hospitals. The null finding here suggests that even when PE-exposed firms face a significant revenue shock, the effect on a specific margin of care quality may be limited—possibly because staffing adjustments are slow, billing was only one of several revenue channels, or because quality monitoring through CMS reporting creates a countervailing incentive. Second, it contributes to the nascent empirical literature on surprise billing regulation, which has primarily focused on prices and billing patterns rather than quality (Adler et al., 2023; Christensen and Feyman, 2023). Third, it joins the broader literature on how provider payment regulation affects care delivery (Clemens and Gottlieb, 2014; Garthwaite, 2012; Ho and Lee, 2019).

The paper proceeds as follows. Section 2 describes the institutional setting. Section 3 presents the data. Section 4 details the empirical strategy. Section 5 reports results. Section 6 discusses mechanisms and implications.

2. Institutional Background

2.1 Emergency Department Staffing and Out-of-Network Billing

Emergency department physician staffing in the United States underwent a quiet transformation in the 2000s and 2010s. Contract management groups—many backed by private equity—replaced hospital-employed physicians at a growing share of facilities. By 2017, two firms dominated: TeamHealth (acquired by Blackstone in 2016 for \$6.1 billion) and Envision Healthcare (acquired by KKR in 2018 for \$9.4 billion), which together staffed physicians in over 3,000 emergency departments (Singhal, 2019; Dafny, 2020).

These firms exploited a distinctive feature of emergency care: patients cannot choose their physician. Under the Emergency Medical Treatment and Labor Act (EMTALA), hospitals must stabilize emergency patients regardless of insurance status or network participation. This creates an unusual market structure in which the provider has significant pricing power. PE-backed staffing firms leveraged this by systematically opting their physicians out of insurance networks, allowing them to charge out-of-network rates that exceeded in-network payments by 200–400% (Cooper et al., 2024; Garmon and Chartock, 2017). The resulting "surprise bills" were borne by patients, who had chosen an in-network hospital but were treated by an out-of-network physician.

2.2 State Surprise Billing Protections

Beginning with New York's 2015 law, states enacted comprehensive protections through two main mechanisms: (1) prohibiting or capping balance billing to patients for emergency services, and (2) establishing payment standards—either binding arbitration (New York's "baseball-style" approach) or reference pricing (California's median in-network rate)—to resolve disputes between providers and insurers.

Table 1 summarizes the treatment timing. The staggered adoption provides the identifying variation: New York and Connecticut enacted protections effective in 2015, Florida in 2016, California in 2017, and five states (Illinois, Maryland, New Hampshire, New Jersey, Oregon) in 2018. A further distinction is that these laws are binding for state-regulated insurance plans but do not apply to self-insured (ERISA) employer plans, which cover roughly 60% of commercially insured individuals. This limits the "bite" of state laws but does not eliminate it, as the remaining 40% of commercially insured patients in each state are protected.

Table 1: State Surprise Billing Law Adoption Timing

State	Effective Year	Mechanism	Key Legislation
New York	2015	Binding arbitration	Emergency Medical Services Law
Connecticut	2015	Reference pricing	PA 15-146
Florida	2016	Reference pricing	HB 221
California	2017	Reference pricing	AB 72
Illinois	2018	Balance billing ban	HB 3978
Maryland	2018	Reference pricing	HB 1122
New Hampshire	2018	Balance billing ban	HB 1809
New Jersey	2018	Arbitration	P.L. 2018 c.32
Oregon	2018	Reference pricing	HB 2339

2.3 Expected Quality Effects

The theoretical prediction is ambiguous. If PE-backed firms responded to revenue losses by reducing staffing ratios or increasing patient throughput pressure, one would expect longer wait times and higher rates of patients leaving before being seen. This is the "cost-cutting" channel emphasized by industry critics. Alternatively, if firms absorbed the revenue loss through lower physician compensation rather than service reductions, or if quality-reporting mandates created reputational incentives to maintain performance, the quality effect could be negligible. A third possibility—that laws *improved* quality by reducing administrative burden or financial distress that distracted from care—would predict negative effects on wait times.

3. Data

3.1 CMS Hospital Compare

I use hospital-level quality measures from the CMS Hospital Compare program, accessed through the NBER Hospital Compare archive (Centers for Medicare and Medicaid Services, 2023). The two primary outcomes are OP-18b, the median time in minutes from ED arrival to discharge for patients not admitted to the hospital, and OP-22, the percentage of patients who left the ED without being seen by a provider. Both measures are publicly reported and directly capture the patient experience of ED quality.

The sample covers NBER archive releases from 2014 through 2019, corresponding to measurement periods from 2013 through 2018. I restrict the sample to acute care hospitals with emergency departments, dropping critical access hospitals, psychiatric facilities, and other specialty hospitals. I also exclude territories (AS, GU, MH, MP, PR, VI) and the District of Columbia.

3.2 Hospital Characteristics

Hospital ownership type (for-profit, nonprofit, government) comes from CMS Hospital General Information files matched to each measurement year. Ownership type serves as a proxy for private equity involvement, following the literature's finding that PE-backed staffing firms are disproportionately present in for-profit hospitals (La Pointe, 2023).

3.3 Sample Description

The final panel contains 20,464 hospital-year observations from 3,063 hospitals across 50 states, spanning measurement years 2013–2018. Of these, 820 hospitals are in the nine treated states and 2,243 are in never-treated states. The panel is unbalanced: not all hospitals report in every year, reflecting entry, exit, and reporting gaps in the CMS data.

Table 2: Summary Statistics

	Full Sample	Treated	Never-Treated
Hospitals	3,063	820	2,243
Hospital-years	20,464	5,559	14,905
Panel A: ED Quality Measures			
ED time to discharge (min.)	147.8	164.1	141.8
SD	(40.8)	(42.0)	(38.6)
Left without being seen (%)	1.9	2.0	1.9
SD	(2.0)	(1.9)	(2.1)
Panel B: Hospital Characteristics			
For-profit (%)	20.0	17.5	20.9
Nonprofit (%)	63.0	70.7	60.1
Government (%)	15.9	11.3	17.6

Notes: Sample consists of acute care hospitals with emergency departments reporting CMS Hospital Compare measures, 2013–2018. ED time = OP-18b (median minutes from arrival to discharge). LWBS = OP-22 (percentage of patients who left without being seen). Treated states enacted comprehensive surprise billing protections between 2015 and 2018.

Table 2 presents summary statistics. Mean ED time to discharge is 148 minutes in the full sample, 22 minutes higher in treated states (164 minutes) than in never-treated states (142 minutes). This level difference largely reflects the urbanicity of treated states (New York, California, Illinois, New Jersey), where EDs tend to be more crowded. The LWBS rate is similar across groups (approximately 2%). For-profit hospitals constitute 20% of the sample, with treated states having a slightly lower for-profit share (17.5%) and a higher nonprofit share (70.7%).

4. Empirical Strategy

4.1 Identification

I exploit the staggered adoption of surprise billing laws across nine states between 2015 and 2018 in a difference-in-differences framework. The identifying assumption is that, absent the law, ED quality in treated states would have evolved on the same trajectory as in never-treated states. I assess this assumption through event-study estimates in the pre-treatment periods.

4.2 Estimation

My primary estimator is the Sun and Abraham (2021) interaction-weighted estimator, implemented through the `sunab()` function in `fixest` (Borusyak et al., 2024). For hospital h in state s at time t:

$$Y_{hst} = \alpha_h + \gamma_t + \sum_e \sum_{\ell \neq -1} \delta_{e\ell} \cdot \mathbb{I}[G_h = e] \cdot \mathbb{I}[t - e = \ell] + \varepsilon_{hst} \tag{1}$$

where α_h are hospital fixed effects, γ_t are year fixed effects, G_h is the cohort (year of first treatment) for hospital h, and ℓ is event time relative to treatment. The aggregate ATT is a weighted average of cohort-time effects $\hat{\delta}_{e\ell}$.

This estimator is preferred over standard TWFE because treatment timing varies across four cohorts (2015, 2016, 2017, 2018), and conventional TWFE can produce biased estimates when treatment effects are heterogeneous across cohorts—using already-treated units as implicit controls (Goodman-Bacon, 2021). The Sun–Abraham estimator uses only never-treated units as the comparison group for each cohort-period cell.

I also report TWFE estimates for comparison:

$$Y_{hst} = \alpha_h + \gamma_t + \beta \cdot D_{st} + \varepsilon_{hst} \tag{2}$$

where $D_{st} = \mathbb{I}[t \geq G_s]$ indicates that state s has enacted its law by year t.

Standard errors are clustered at the state level (50 clusters) to account for within-state correlation in both the treatment assignment and outcome measurement. The reference period is $k = -1$ (the year before treatment). A caveat on inference: while 50 clusters is generally sufficient for cluster-robust standard errors, only 9 states are treated. With few treated clusters, standard asymptotic inference may be anti-conservative (Rambachan and Roth, 2023). I address this concern through leave-one-state-out analysis, which confirms that no single treated state drives the result.

4.3 Threats to Validity

The main threat is differential trends in ED quality between treated and never-treated states that are unrelated to surprise billing laws. Treated states are disproportionately large, urban, and Democratic-leaning, and may have experienced different trends in ED crowding, hospital closures, or insurance coverage expansion under the ACA. I assess this through event-study coefficients in the pre-treatment periods, focusing on $k = -2$ and $k = -3$ as the most informative horizon (earlier periods are driven by single cohorts with thin composition). I also conduct leave-one-state-out analysis and examine a placebo outcome (OP-20, the door-to-diagnostic evaluation time).

A second concern is the limited "bite" of state laws due to ERISA preemption. State surprise billing protections apply only to state-regulated insurance plans, meaning roughly 40% of commercially insured patients are covered. This attenuates the treatment intensity and pushes the estimated effect toward zero—if out-of-network billing accounted for roughly 20% of ED physician revenue, the effective revenue shock may be only $\sim$8% (20% $\times$ 40%). I interpret the results as intent-to-treat effects of the state law rather than as the effect of full surprise billing elimination. A sharper test—exploiting variation in self-insured (ERISA-exempt) plan penetration across hospital markets as an internal control—would help isolate the mechanism but requires hospital-level payer mix data not available in the CMS Hospital Compare files.

5. Results

5.1 Main Results

Table 3 presents the main estimates. Columns (1)–(2) show results for ED time to discharge; columns (3)–(4) for the LWBS rate. The TWFE estimate for ED time (column 1) is 0.15 minutes (SE = 1.27), and the Sun–Abraham ATT (column 2) is 0.33 minutes (SE = 0.74). Both are small, positive, and statistically insignificant. For LWBS, the TWFE estimate is 0.01 percentage points (SE = 0.12), and the SA estimate is 0.08 percentage points (SE = 0.077). The SA estimate for LWBS, while in the predicted positive direction, is not statistically significant at conventional levels ($p = 0.31$).

Table 3: Effect of Surprise Billing Laws on ED Quality

	ED Time (minutes)		LWBS (%)	
	TWFE	SA	TWFE	SA
	(1)	(2)	(3)	(4)
Surprise billing law	0.15	0.33	0.01	0.08
	(1.27)	(0.74)	(0.123)	(0.077)
	$[-2.34, 2.64]$	$[-1.12, 1.78]$	$[-0.227, 0.254]$	$[-0.076, 0.226]$
Hospital FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Estimator	TWFE	Sun–Abraham	TWFE	Sun–Abraham
Observations	20,464	20,464	17,371	17,371
Treated states	9		9	
Clusters (states)	50		50	

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Standard errors clustered at the state level in parentheses; 95% confidence intervals in brackets. TWFE = two-way fixed effects. SA = Sun and Abraham (2021) interaction-weighted estimator with never-treated states as controls. ED Time = median minutes from arrival to discharge (OP-18b). LWBS = percentage who left without being seen (OP-22).

The 95% confidence interval for the SA ED time estimate $[-1.12, 1.78]$ rules out effects larger than 1.78 minutes—roughly 1.2% of the sample mean or 0.04 standard deviations. For LWBS, the confidence interval $[-0.08, 0.23]$ rules out effects larger than 0.23 percentage points, or about 12% of the sample mean (0.11 SD). To contextualize: McCarthy et al. (2009) estimate that severe ED crowding episodes increase length-of-stay by 10–30 minutes, so the confidence interval comfortably rules out effects on the scale of a major crowding shock. The minimum detectable effect at 80% power is approximately $2.8 \times \text{SE} = 2.1$ minutes for ED time and 0.22 percentage points for LWBS, confirming that the study is adequately powered to detect economically meaningful changes.

5.2 Event Study

Table 4 reports the Sun–Abraham event-study estimates for ED time to discharge. The two periods immediately preceding treatment—$k = -2$ and $k = -3$—show small, insignificant coefficients (0.05 and -0.18 minutes), supporting the parallel trends assumption in the critical pre-treatment window. Earlier periods show significant negative coefficients at $k = -4$ (-2.71 minutes, $p < 0.05$) and $k = -5$ (-6.36 minutes, $p < 0.01$). These distant pre-period estimates are driven by cohort composition: only the 2015 cohort (New York and Connecticut) contributes to $k = -5$, and $k = -4$ is dominated by the same two states. The large negative coefficients likely reflect level differences between the specific hospitals that report in those

Table 4: Event Study: ED Time to Discharge

Event Time	Estimate	SE	95% CI	Pre/Post
$k = -5$	-6.36^{***}	(1.72)	$[-9.74, -2.99]$	Pre
$k = -4$	-2.71^{**}	(1.12)	$[-4.91, -0.51]$	Pre
$k = -3$	-0.18	(0.78)	$[-1.71, 1.36]$	Pre
$k = -2$	0.05	(0.31)	$[-0.57, 0.66]$	Pre
$k = +0$	0.79	(0.59)	$[-0.36, 1.93]$	Impact
$k = +1$	-0.96	(0.75)	$[-2.44, 0.52]$	Post
$k = +2$	-1.43	(1.12)	$[-3.61, 0.76]$	Post
$k = +3$	5.22^{***}	(1.57)	$[2.14, 8.29]$	Post

Notes: Sun and Abraham (2021) interaction-weighted event-study estimates. Dependent variable: median ED time to discharge (minutes). Standard errors clustered at the state level. Reference period: $k = -1$.

years and the evolving control group, rather than a trending violation that would bias the post-treatment estimates. The key pre-trend assessment rests on $k = -2$ and $k = -3$, where all four cohorts contribute and the estimates are close to zero.

Post-treatment estimates are mixed. The impact year ($k = 0$) shows a small, insignificant increase of 0.79 minutes. Subsequent periods show oscillating signs (-0.96 at $k = +1$, -1.43 at $k = +2$, $+5.22$ at $k = +3$). The $k = +3$ estimate is statistically significant but again reflects thin composition—only the 2015 cohort contributes—and likely captures idiosyncratic variation in New York and Connecticut hospitals. The overall pattern is consistent with no sustained post-treatment shift in ED quality.

5.3 Heterogeneity by Ownership

Table 5 examines whether effects differ by hospital ownership type. If private equity-backed staffing firms are concentrated in for-profit hospitals, one would expect larger quality effects there. The point estimates are consistent with this prediction: for-profit hospitals show a 1.60-minute increase (SE = 1.01), while nonprofits show a -0.80-minute *decrease* (SE = 0.87). However, neither estimate is statistically significant. Government hospitals show a marginally significant increase of 1.81 minutes ($p < 0.10$), which may reflect the distinct billing practices of public facilities. These suggestive patterns warrant further investigation with direct measures of PE staffing penetration.

5.4 Robustness

Leave-one-state-out. I re-estimate the TWFE specification dropping each treated state in turn. The point estimates range from -0.23 (dropping Oregon) to $+1.09$ (dropping New

Table 5: Heterogeneity by Hospital Ownership

	For-Profit (1)	Nonprofit (2)	Government (3)
Surprise billing law	1.60	-0.80	1.81*
	(1.01)	(0.87)	(1.06)
Hospital FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Estimator	Sun–Abraham	Sun–Abraham	Sun–Abraham
Observations	4,088	12,886	3,250
Hospitals	626	1,905	484
Treated hospitals	142	577	91

Notes: ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$. Sun and Abraham (2021) interaction-weighted ATT estimates by hospital ownership type. Standard errors clustered at the state level. For-profit hospitals serve as a proxy for private equity involvement in ED staffing.

York) minutes, with no single state driving the result. The null is robust across all nine exclusions.

Placebo outcome. I estimate the same specification using OP-20 (door-to-diagnostic evaluation time, an ED process measure that captures initial triage speed). The SA ATT is -1.41 minutes (SE $= 0.26$, $p < 0.001$), indicating that triage speed *improved* in treated states relative to controls. This result is difficult to interpret as a clean falsification test because OP-20 is itself an ED process measure that could respond to staffing changes—if firms increased triage efficiency while maintaining throughput times, one would observe exactly this pattern. However, it could also reflect concurrent state-level ED throughput initiatives (e.g., New York's delivery system reform) or differential ACA Medicaid expansion effects on ED visit composition. An ideal placebo would use an inpatient quality measure (e.g., 30-day readmissions) that is unaffected by ED staffing but captures underlying hospital trends; such measures are not available in the Hospital Compare ED dataset used here. The significant OP-20 result is a limitation that should temper the strength of causal claims.

6. Discussion

The central finding—that state surprise billing laws did not measurably degrade ED quality—has several possible interpretations. First, hospitals and staffing firms may have absorbed revenue losses through channels that do not directly affect measured quality, such as reduced physician compensation, renegotiated contracts, or shifts in billing strategy for non-emergency services. Cooper et al. (2024) document that out-of-network billing rates dropped sharply after state laws, but the translation from revenue to staffing to patient outcomes involves multiple margins of adjustment, each of which may attenuate the quality effect.

Second, the quality measures used here—median discharge time and LWBS rate—are publicly reported and may be subject to monitoring effects. Hospitals that know their CMS performance metrics are public may be reluctant to allow measurable deterioration, even under financial pressure. This would not prevent quality reductions on unmeasured margins (e.g., thoroughness of evaluation, patient satisfaction, or diagnostic accuracy), which are not captured in these data.

Third, the limited bite of state laws due to ERISA preemption means that the treatment intensity is partial. If only 30–40% of commercially insured patients in treated states are covered, the revenue shock to staffing firms is correspondingly attenuated. The aggregate null may mask meaningful effects for the protected patient population that are diluted in hospital-level averages.

Fourth, the significant placebo (OP-20) warrants caution. While I argue this is not an ideal falsification test—triage speed could genuinely respond to staffing changes—it raises the possibility that treated and control states were on differential ED quality trajectories for reasons unrelated to surprise billing. Any such confound would need to affect triage speed but not overall ED throughput or patient retention, a somewhat specific pattern.

The ownership heterogeneity results offer suggestive but inconclusive evidence on the PE mechanism. The positive point estimate for for-profit hospitals (+1.60 minutes) and negative estimate for nonprofits (−0.80 minutes) are directionally consistent with PE-driven cost-cutting, but neither is statistically significant. Crucially, for-profit ownership is an imperfect proxy for PE staffing penetration: many for-profit hospitals do not use TeamHealth or Envision, and some nonprofit hospitals do. Future work with direct measures of PE staffing penetration—from CMS Open Payments data or hospital cost reports linking physician FTEs to staffing firms—would provide a sharper test of whether PE-backed firms specifically reduced staffing in response to billing restrictions.

Several limitations warrant emphasis. First, the study cannot rule out quality degradation on unmeasured margins such as diagnostic accuracy, thoroughness of evaluation, or patient satisfaction, which are not captured in the CMS Hospital Compare data. Second, the significant placebo result for OP-20 raises the possibility of differential trends in ED process measures between treated and control states, potentially from concurrent state-level reforms (e.g., delivery system redesign in New York, Medicaid managed care transitions in California). Third, with only nine treated states, the asymptotic properties of cluster-robust standard errors may not hold exactly; while the leave-one-out analysis provides reassurance, wild cluster bootstrap inference would offer more conservative tests. Finally, the results are specific to state-level laws that applied to roughly 40% of commercially insured patients. The federal No

Surprises Act (2022), which eliminates the ERISA preemption gap, imposes a substantially larger revenue shock and may produce different quality effects.

7. Conclusion

State surprise billing laws—enacted to protect patients from unexpected medical charges—did not detectably harm emergency department quality of care. Using staggered adoption across nine states and robust difference-in-differences methods, I find that neither ED wait times nor rates of patients leaving without being seen changed meaningfully after these protections took effect. The confidence intervals are tight enough to rule out large effects. This reassuring null suggests that the consumer protection benefits of surprise billing regulation need not come at the cost of degraded emergency care, at least on the dimensions measured here. Whether these laws affected unmeasured quality margins, or whether the federal No Surprises Act's broader scope produces different effects, remains an open question for future research.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers
Contributors: @olafdrw
First Contributor: https://github.com/olafdrw

References

Adler, Loren, Mark Hall, and Caitlin Brandt, "The Effect of State Balance-Billing Protections," Technical Report, USC-Brookings Schaeffer Initiative for Health Policy 2023.

Bernstein, Steven L., Dominik Aronsky, Rahul Duseja, Susan Epstein, Daniel Handel, Ula Hwang et al., "The Effect of Emergency Department Crowding on Clinically Oriented Outcomes," *Academic Emergency Medicine*, 2009, *16* (1), 1–10.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Centers for Medicare and Medicaid Services, "CMS Hospital Compare Data Archive," Technical Report, U.S. Department of Health and Human Services 2023.

Christensen, Hannah and Yevgeniy Feyman, "The Impact of the No Surprises Act on Provider Consolidation," *Health Affairs*, 2023, *42* (2), 235–243.

Clemens, Jeffrey and Joshua D. Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?," *American Economic Review*, 2014, *104* (4), 1320–1349.

Cooper, Zack, Fiona M. Scott Morton, and Nathan Shekita, "The Surprise Is the Bill: Private Equity, Physician Staffing, and Out-of-Network Emergency Charges," *Quarterly Journal of Economics*, 2024, *139* (2), 1149–1196.

Dafny, Leemore, "The Role of Private Equity in Health Care," *JAMA*, 2020, *323* (9), 817–818.

Eliason, Paul J., Benjamin Heebsh, Ryan C. McDevitt, and James W. Roberts, "How Acquisitions Affect Firm Behavior and Performance: Evidence from the Dialysis Industry," *Quarterly Journal of Economics*, 2020, *135* (1), 221–267.

Garmon, Christopher and Benjamin Chartock, "Overview of Out-of-Network Emergency Department Claims and State Balance-Billing Protections," *Health Affairs*, 2017, *36* (10), 1805–1812.

Garthwaite, Craig L., "The Doctor Might See You Now: The Supply Side Effects of Public Health Insurance Expansions," *American Economic Journal: Economic Policy*, 2012, *4* (3), 190–215.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Gupta, Atul, Sabrina T. Howell, Constantine Yannelis, and Abhinav Gupta, "Does Private Equity Investment in Healthcare Benefit Patients? Evidence from Nursing Homes," *Working Paper*, 2021.

Ho, Kate and Robin S. Lee, "The Impact of Consumer Health Insurance Mandates on Insurance Coverage, Access, Health, and Costs," *Annual Review of Economics*, 2019, *11*, 521–552.

McCarthy, Melissa L., Scott L. Zeger, Ru Ding, Scott R. Levin, Jeffrey S. Desmond, Jong Lee, and Dominik Aronsky, "Crowding Delays Treatment and Lengthens Emergency Department Length of Stay, Even Among High-Acuity Patients," *Annals of Emergency Medicine*, 2009, *54* (4), 492–503.

Pointe, Bryan N. La, "How Are Hospitals Affected by Private Equity?," *American Economic Review*, 2023, *113* (12), 3339–3373.

Pollitz, Karen, Cynthia Cox, and Kevin Lucia, "Surprise Out-of-Network Medical Bills: Prevalence, Magnitude, and Policy Solutions," *Health Affairs*, 2020, *39* (10), 1729–1735.

Rambachan, Ashesh and Jonathan Roth, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

Roth, Jonathan, Pedro H.C. Sant'Anna, Alyssa Bilinski, and John Poe, "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Singer, Adam J., Henry C. Thode Jr, Peter Viccellio, and Jesse M. Pines, "The Association Between Length of Emergency Department Boarding and Mortality," *Academic Emergency Medicine*, 2011, *18* (12), 1324–1329.

Singhal, Shreya, "How Private Equity Is Involved in Emergency Medicine," *Annals of Emergency Medicine*, 2019, *73* (6), A15–A17.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

A. Standardized Effect Sizes

Table 6: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
ED time to discharge (min.)	0.33	0.74	40.77	0.0081	0.0181	Small positive
Left without being seen (%)	0.08	0.08	2.05	0.0368	0.0376	Small positive

Notes: This paper studies whether state surprise billing laws affected emergency department quality of care. Data: CMS Hospital Compare OP-18b and OP-22, 2013–2018. Estimator: Sun and Abraham (2021) interaction-weighted ATT with never-treated states as controls. Sample: 3,063 acute care hospitals across 50 states. SDE = $\hat{\beta}$ / SD(Y). Classification reflects magnitude, not statistical significance.

Classification labels refer to the magnitude of the standardized point estimate, not to statistical significance. "Null" denotes a near-zero effect size ($|SDE| < 0.005$), not a failure to reject a null hypothesis.