The Measurement Artifact of Crime: How NIBRS Adoption Inflates Reported Offense Rates

APEP Autonomous Research* @olafdrw

March 25, 2026

Abstract

Every empirical study of U.S. crime implicitly assumes that FBI offense counts measure the same object across jurisdictions and time. They do not. When agencies transition from the Summary Reporting System to the National Incident-Based Reporting System, the elimination of the "hierarchy rule"—which counted only the most serious offense per incident—mechanically inflates reported crime. Using the staggered adoption of NIBRS across 40 U.S. states from 2000 to 2020 in a Callaway–Sant'Anna difference-in-differences design, I estimate that NIBRS adoption increases reported violent crime rates by 14.2 percent and aggravated assault by 16.0 percent. Murder—always atop the hierarchy and thus unaffected by rule removal—shows a precisely estimated null, confirming the measurement channel. These artifacts are large enough to confound policy evaluations spanning the SRS-to-NIBRS transition.

JEL Codes: K42, C81, H76

Keywords: crime measurement, NIBRS, hierarchy rule, UCR, staggered DiD

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 29m).

1. Introduction

A robbery victim is also assaulted. A domestic violence call reveals drug possession. A burglary escalates into a confrontation. In each case, a single criminal incident produces multiple offenses—but for six decades, the FBI's Uniform Crime Reporting (UCR) Summary Reporting System (SRS) counted only one. The "hierarchy rule" required agencies to report the single most serious offense per incident, discarding the rest. When the FBI began replacing SRS with the National Incident-Based Reporting System (NIBRS)—which records every offense in every incident—those previously invisible crimes reappeared in the data. Not because crime changed, but because measurement did.

This paper quantifies the resulting measurement artifact. The question matters because the FBI's crime statistics underpin an enormous empirical literature: evaluations of policing strategies (Weisburd et al., 2016), gun control laws (Donohue et al., 2019), drug policy (Miron, 2004), incarceration (Liedka et al., 2006), and economic conditions (Raphael and Winter-Ebmer, 2001). If the transition from SRS to NIBRS creates a systematic upward shift in measured crime, then any study comparing jurisdictions or time periods that straddle this transition risks confounding real policy effects with an accounting change.

The identification strategy exploits the staggered adoption of NIBRS across U.S. states. Between 1991 and 2024, the FBI gradually transitioned approximately 18,000 law enforcement agencies from SRS to NIBRS, with adoption timing driven by state-level implementation decisions, FBI grant funding (the NCS-X program), and ultimately a 2021 federal deadline. I construct a panel of 40 states over 2000–2020 and estimate the effect of NIBRS adoption on reported crime rates using the Callaway and Sant'Anna (2021) staggered difference-in-differences estimator, which is robust to the heterogeneous treatment effects that plague two-way fixed effects (TWFE) specifications under staggered adoption.

The design features a powerful built-in placebo. Murder always sits atop the offense hierarchy: a murder incident reported under SRS already reports the murder. Removing the hierarchy rule cannot mechanically change murder counts. A non-zero NIBRS effect on murder would signal that the identifying variation captures something beyond the hierarchy rule—perhaps coincidental changes in policing intensity or reporting effort. The murder placebo is clean: the Callaway–Sant'Anna ATT is 0.068 (SE = 0.104), economically small and statistically indistinguishable from zero.

The results reveal substantial measurement artifacts for crimes most affected by the hierarchy rule. NIBRS adoption increases reported violent crime rates by 14.2 percent ($p < 0.10$) and aggravated assault—frequently subordinated to robbery or weapons offenses under the hierarchy—by 16.0 percent ($p < 0.05$). Property crime shows a smaller, imprecise increase of 3.0 percent, consistent with property offenses being less commonly involved in multi-offense incidents. These effects are robust to alternative control groups (not-yet-treated comparison), leave-one-state-out analysis, and pre-treatment event study specifications that show flat trends prior to adoption.

The contribution is a correction factor for the empirical crime literature. The 14–16 percent violent crime artifact means that a naive before-after comparison of a state that adopted NIBRS in 2015 would attribute a one-sixth increase in assault to whatever policy happened to change around the same time. The finding also reframes the apparent crime "surge" in states that recently completed the NIBRS transition: part of what looks like rising crime is rising measurement fidelity.

Related literature. The Bureau of Justice Statistics has documented SRS-NIBRS discrepancies in cross-sectional comparisons (Rantala and Edwards, 2000), finding approximately 2 percent average differences for about 1,100 agencies. But cross-sectional comparisons cannot separate the mechanical hierarchy-rule effect from selection into early adoption. Kaplan (2021) provides the most comprehensive compilation of UCR data and notes measurement discontinuities, but does not estimate causal effects. The broader crime-data measurement literature examines reporting incentives (Maltz, 1999), the "dark figure" of unreported crime (Biderman and Reiss Jr., 1967), and measurement error in police statistics (Pepper and Petrie, 2004). This paper is, to my knowledge, the first to use the staggered NIBRS rollout as a natural experiment to causally identify the hierarchy rule's measurement bias.

2. Institutional Background

The Summary Reporting System and the hierarchy rule. Since 1930, the FBI has collected crime statistics from local law enforcement through the UCR program. Under the Summary Reporting System, agencies submit monthly aggregate counts of eight "Part I" offenses: murder, rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, and arson. The hierarchy rule dictates that when multiple offenses occur in a single incident, only the most serious is counted. The hierarchy ranks offenses roughly in the order listed above, with murder at the top.

Consider a home invasion where the perpetrator commits burglary, assault, and theft. Under SRS, this incident contributes exactly one count—burglary, the most serious offense—while the assault and theft disappear from the statistics. Nationally, multi-offense incidents are common: the BJS estimates that approximately 15–20 percent of criminal incidents involve more than one offense type.

The transition to NIBRS. NIBRS, developed in the 1980s and first deployed in 1991, records the complete set of offenses per incident: the unit of observation is the incident rather than the offense category. NIBRS tracks 52 "Group A" offense types with no hierarchy rule. The same home invasion now generates three offense counts—burglary, assault, and theft—rather than one.

The FBI's transition strategy was incremental. South Carolina was the first state to achieve full NIBRS coverage in 1991. Throughout the 1990s and 2000s, adoption proceeded unevenly, driven by state criminal justice information system capacity and FBI grant support through the NCS-X program (approximately $120 million since 2013). In December 2015, the FBI announced that SRS submissions would be discontinued after January 1, 2021, forcing a deadline. By 2021, approximately 66 percent of agencies had transitioned; the FBI subsequently reversed course and continued accepting SRS data. As of 2024, about 76 percent of agencies covering 87 percent of the U.S. population report via NIBRS.

3. Data

I construct a state-year panel combining two data sources. State-level crime counts and population figures come from the FBI's *Crime in the United States* annual publications, as compiled by the Disaster Center, covering 2000–2020 for 40 U.S. states. The NIBRS adoption year for each state—defined as the year the state achieved majority population coverage under NIBRS—is compiled from FBI and BJS transition tracking reports.

The panel contains 803 state-year observations across 40 states and 21 years. Thirty states adopted NIBRS during the sample window (2001–2020), spanning 16 distinct adoption cohorts. Ten states serve as never-treated controls: states that adopted before 2000 (e.g., South Carolina in 1991, Idaho and Iowa in 1996) or after 2020 (e.g., California, Florida, New York). Table 1 reports pre-treatment summary statistics: the mean violent crime rate is 409 per 100,000 (SD = 151), and the mean property crime rate is 3,121 per 100,000 (SD = 796).

4. Empirical Strategy

The identification exploits the staggered timing of NIBRS adoption across states. Let Y_{st} denote the log crime rate in state s at time t, and let G_s denote the year state s first achieves majority NIBRS coverage. The parameter of interest is the average treatment effect on the treated (ATT):

$$\mathrm{ATT}(g,t) = \mathbb{E}[Y_{st}(g) - Y_{st}(\infty) \mid G_s = g] \tag{1}$$

where $Y_{st}(g)$ is the potential outcome under treatment at time g and $Y_{st}(\infty)$ is the potential outcome under no treatment.

I estimate $\text{ATT}(g,t)$ for each group-time pair using the Callaway and Sant'Anna (2021) estimator, which avoids the forbidden comparisons problem in TWFE by using only never-treated states as controls. The estimator requires parallel trends conditional on group membership: absent NIBRS adoption, treated and control states would have followed the same trajectory. The overall ATT aggregates across groups and post-treatment periods.

Placebo test. The built-in placebo is murder. Because murder always occupies the top of the hierarchy, the hierarchy rule's elimination cannot mechanically change murder counts. A statistically significant NIBRS effect on murder would indicate that the treatment variable captures confounding changes beyond the hierarchy rule—for instance, if NIBRS adoption coincided with changes in policing effort or reporting completeness. The identifying assumption is supported if the murder ATT is close to zero.

Threats to identification. The main concern is that NIBRS adoption may correlate with other changes in crime or reporting. States that adopt NIBRS may simultaneously invest in law enforcement technology, affecting both measurement and actual crime. I address this through the never-treated control group (which includes early adopters with stable NIBRS coverage), pre-treatment event studies showing flat trends, leave-one-state-out analysis confirming no single state drives results, and the murder placebo.

5. Results

Table 2 presents the main estimates. The Callaway–Sant'Anna ATT for log violent crime is 0.133 (SE = 0.069), implying a 14.2 percent increase in reported violent crime upon NIBRS adoption. The effect is driven primarily by aggravated assault, which shows the largest artifact at 0.149 (SE = 0.072), corresponding to a 16.0 percent increase. This is consistent with the hypothesis that assault is the offense most commonly subordinated under the hierarchy rule: when an incident involves both robbery and assault, SRS counts only the robbery.

Robbery shows a positive but imprecise effect of 7.9 percent. Property crime and burglary show smaller effects of 3.0 and 3.2 percent, respectively, neither statistically significant. The smaller property crime artifact accords with the institutional logic: property offenses are less frequently involved in multi-offense incidents than violent offenses, and when they are, the property offense is typically the one counted under the hierarchy rule (being the most serious offense in property-only incidents).

The murder placebo provides the critical validation. The Callaway–Sant'Anna ATT for murder is 0.068 (SE = 0.104), statistically indistinguishable from zero with a t-statistic of 0.65. The 95 percent confidence interval spans -13.5 to $+27.5$ percent, comfortably encompassing zero. While the point estimate of 7.0 percent is not negligible, it is less than half the violent crime effect and could reflect coincidental improvements in reporting completeness—such as better death investigation protocols—that accompany the broader modernization of crime records systems. Crucially, the wide confidence interval cannot reject zero, and the estimate is far from the 14–16 percent magnitudes found for offenses directly subject to the hierarchy rule.

Pre-trends. Table 4 reports event study coefficients for property crime from a TWFE specification with state and year fixed effects. While the main estimates use the Callaway–Sant'Anna estimator, the TWFE event study provides a transparent visualization of the pre-treatment dynamics. The pre-treatment coefficients at $t-6$ through $t-2$ are all small and statistically insignificant, with no systematic trend: the largest is -0.012 at $t-4$. After adoption, coefficients turn positive and grow, reaching 0.050 at $t+3$. The flat pre-trends support the parallel trends assumption.

Comparison with TWFE. The TWFE estimates in Table 2 are shown for comparison. For property crime, TWFE yields a marginally significant coefficient of 0.045 versus the CS-DiD estimate of 0.030. For violent crime, TWFE produces a smaller point estimate (0.067) than CS-DiD (0.133). This pattern—TWFE attenuation of the treatment effect for outcomes with growing dynamic effects—is consistent with the forbidden comparisons bias documented by Goodman-Bacon (2021): early-treated states serve as implicit controls for late-treated states, biasing TWFE toward zero when treatment effects accumulate over time.

6. Robustness

Table 3 reports robustness checks for the violent crime result. Using not-yet-treated states as the comparison group yields an ATT of 0.104 (SE = 0.066), slightly attenuated relative to the baseline (0.133) but qualitatively consistent. The attenuation is expected: not-yet-treated comparisons use eventually-treated states as controls in their pre-treatment period, which may already exhibit partial NIBRS effects if adoption is gradual.

The leave-one-state-out analysis drops each of the 30 treated states in turn. The ATT ranges from 0.082 to 0.143, with a median of 0.130. No single state drives the result; South Dakota is the most influential exclusion, with its removal producing the largest deviation from the baseline. The stability of the point estimate across 30 permutations reinforces the conclusion that the effect reflects a systematic measurement artifact rather than an idiosyncratic state-level change.

7. Discussion

The measurement artifact documented here has direct implications for applied crime research. A 14–16 percent artifact in violent crime means that any difference-in-differences study using FBI data where treatment timing correlates with NIBRS adoption timing is potentially confounded. This is not a hypothetical concern: many state-level policy changes—marijuana legalization, sentencing reform, policing interventions—occurred during the same period as the NIBRS transition. Researchers using FBI crime data should, at minimum, control for NIBRS adoption status or restrict samples to years when reporting regimes are stable.

The finding also helps reconcile an empirical puzzle. Several states that completed the NIBRS transition in 2019–2021 appeared to experience sharp increases in crime rates around the same period. While some of this increase is real—driven by pandemic-era disruptions—part is measurement artifact. The 16 percent assault artifact alone could account for a substantial fraction of the reported increase in violent crime for late-adopting states.

Limitations. This analysis operates at the state level, whereas the ideal design would exploit agency-level variation in adoption timing across approximately 18,000 law enforcement agencies. State-level NIBRS adoption dates necessarily aggregate across heterogeneous agencies, and the binary treatment definition—majority population coverage—is coarse relative to the continuous within-state diffusion of NIBRS. With 40 clusters, inference is based on cluster-robust standard errors and the Callaway–Sant'Anna bootstrap; wild cluster bootstrap would provide additional reassurance. The Disaster Center compilation covers 40 of 50 states; the 10 excluded states include large jurisdictions (Kansas, Montana, Missouri, North Carolina, among others) whose omission may affect external validity, though the included sample covers a broad range of state sizes and adoption timings. Finally, the paper lacks a direct mechanism test—for example, showing that the share of multi-offense incidents jumps at adoption—that would definitively attribute the artifact to the hierarchy rule rather than to correlated changes in reporting completeness.

The results should be interpreted as the measurement artifact of NIBRS adoption in an intention-to-treat sense. The actual hierarchy-rule effect on any individual agency depends on that agency's mix of multi-offense incidents, which varies with urbanization, crime type composition, and policing practices.

8. Conclusion

The hierarchy rule embedded in the FBI's Summary Reporting System systematically undercounted crime for six decades. This paper provides the first causal estimate of the resulting measurement artifact: NIBRS adoption inflates reported violent crime by 14 percent and aggravated assault by 16 percent, while the murder placebo confirms that the effect operates through the hierarchy rule channel. The correction factor is a public good for the empirical crime literature—a first step toward measurement-adjusted crime rates that are comparable across the SRS-to-NIBRS transition.

Table 1: Summary Statistics: Pre-Treatment State-Level Crime Rates

	Mean	SD	Min	Max
Panel A: Crime Rates per 100,000				
Violent crime rate	408.5	150.9	112.0	804.9
Murder rate	4.8	2.3	1.1	11.8
Robbery rate	113.8	51.6	11.5	241.2
Aggravated assault rate	254.7	107.1	60.8	563.2
Property crime rate	3,120.9	795.5	1,373.3	4,953.8
Burglary rate	645.5	218.5	180.7	1,096.2
Larceny-theft rate	2,158.4	513.0	1,072.1	3,443.9
Motor vehicle theft rate	315.2	177.9	40.6	921.7
Panel B: Demographics				
Population	8,264,956	8,564,994	493,754	39,512,223

Observations: 499 pre-treatment state-year obs (of 803 total). Sample: 40 U.S. states, 2000–2020. Pre-treatment period: all years before NIBRS adoption for treated states; all years for never-treated states. Source: FBI UCR via Disaster Center.

Table 2: Effect of NIBRS Adoption on Reported Crime Rates

	CS-DiD ATT	TWFE (biased)	Implied % Change
Property crime	0.0296	0.0451*	3.0%
	(0.0258)	(0.0261)	
Violent crime	0.1332*	0.0670	14.2%
	(0.0686)	(0.0434)	
Robbery	0.0763	0.0669	7.9%
	(0.0544)	(0.0442)	
Agg. assault	0.1485**		16.0%
	(0.0718)		
Burglary	0.0316	0.0620*	3.2%
	(0.0401)	(0.0340)	
Murder (placebo)	0.0675	0.0723	7.0%
	(0.1036)	(0.0476)	

States: 40. Years: 2000–2020. Treated: 30 states, 16 cohorts.
CS-DiD: Callaway and Sant'Anna (2021), never-treated control, regression estimation. TWFE shown for comparison (known biased with staggered adoption). Clustered SEs at state level in parentheses. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 3: Robustness: Alternative Specifications for Violent Crime Rate

Specification	ATT	SE
Panel A: Main result		
CS-DiD (never-treated control)	0.1332*	(0.0686)
Panel B: Alternative control group		
CS-DiD (not-yet-treated control)	0.1040	(0.0659)
Panel C: Leave-one-state-out		
Minimum ATT (drop South Dakota)	0.0819	
Maximum ATT (drop Michigan)	0.1431	
Median ATT	0.1340	

Outcome: log violent crime rate per 100,000. Leave-one-state-out drops each of the 30 treated states in turn.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 4: Event Study: Property Crime Rate Around NIBRS Adoption

Event Time	Coefficient	SE
Pre-treatment		
$t-6$	0.0006	(0.0333)
$t-5$	-0.0023	(0.0251)
$t-4$	-0.0121	(0.0180)
$t-3$	-0.0107	(0.0161)
$t-2$	-0.0118	(0.0109)
$t-1$ (ref.)	0.000	—
Post-treatment		
$t+0$	0.0143	(0.0128)
$t+1$	0.0298	(0.0183)
$t+2$	0.0387	(0.0241)
$t+3$	0.0498	(0.0305)
$t+4$	0.0446	(0.0359)
$t+5$	0.0451	(0.0384)
$t+6$	0.0298	(0.0349)

TWFE event study with state and year FE.
Dependent variable: log property crime rate per 100,000.
Clustered SEs at state level. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

References

Biderman, Albert D. and Albert J. Reiss Jr., "On exploring the "dark figure" of crime," *The Annals of the American Academy of Political and Social Science*, 1967, *374* (1), 1–15.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Donohue, John J., Abhay Aneja, and Kyle D. Weber, "Right-to-carry laws and violent crime: A comprehensive assessment using panel data and a state-level synthetic control analysis," *Journal of Empirical Legal Studies*, 2019, *16* (2), 198–247.

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

Kaplan, Jacob, *Jacob Kaplan's Concatenated Files: Uniform Crime Reporting Program Data*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research, 2021.

Liedka, Raymond V., Anne Morrison Piehl, and Bert Useem, "The crime-control effect of incarceration: Does scale matter?," *Criminology & Public Policy*, 2006, *5* (2), 245–276.

Maltz, Michael D., *Bridging gaps in police crime data*, U.S. Department of Justice, Bureau of Justice Statistics, 1999.

Miron, Jeffrey A., "Drug war crimes: The consequences of prohibition," *The Independent Institute*, 2004.

Pepper, John V. and Carol V. Petrie, "Quantitative analysis of the effects of criminal sanctions on crime rates," in "Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates," National Academies Press, 2004.

Rantala, Ramona R. and Thomas J. Edwards, "Effects of NIBRS on crime statistics," *Bureau of Justice Statistics Special Report*, 2000.

Raphael, Steven and Rudolf Winter-Ebmer, "Identifying the effect of unemployment on crime," *The Journal of Law and Economics*, 2001, *44* (1), 259–283.

Weisburd, David, David P. Farrington, and Charlotte Gill, "What works in crime prevention and rehabilitation: Lessons from systematic reviews," *Criminology & Public Policy*, 2016, *15* (4), 1–27.

A. Standardized Effect Sizes

This appendix reports standardized effect sizes to facilitate cross-study comparisons and meta-analysis. The SDE normalizes the treatment effect by the pre-treatment standard deviation of the outcome variable, providing a unit-free measure of effect magnitude.

Table 5: Standardized Effect Sizes: NIBRS Adoption and Crime Measurement

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
Violent crime rate	0.1332	0.0686	0.4048	0.329	0.169	Large positive
Aggravated assault rate	0.1485	0.0718	0.4522	0.328	0.159	Large positive
Property crime rate	0.0296	0.0258	0.2585	0.115	0.100	Moderate positive
Robbery rate	0.0763	0.0544	0.6100	0.125	0.089	Moderate positive
Panel B: Heterogeneous (by adoption timing)						
Violent crime (early adopters, $\leq$2012)	0.1846	0.0883	0.4048	0.456	0.218	Large positive
Violent crime (late adopters, >2012)	0.0034	0.0377	0.4048	0.008	0.093	Small positive

- *Notes:* **Country:** United States. **Research question:** Does the transition from the FBI's Summary Reporting System (SRS) to the National Incident-Based Reporting System (NIBRS) mechanically inflate measured state-level crime rates through the elimination of the hierarchy rule? **Policy mechanism:** NIBRS removes the hierarchy rule that required agencies to report only the most serious offense per incident, and expands offense categories from 8 to 52, thereby counting all offenses per incident rather than just one. **Outcome definition:** Log of the annual reported crime rate per 100,000 population, computed from FBI Uniform Crime Report offense counts divided by Census population estimates. **Treatment:** Binary; equals one in the year a state achieves majority NIBRS population coverage and all subsequent years. **Data:** FBI UCR state-level crime counts compiled by the Disaster Center, 2000–2020, 40 U.S. states, 803 state-year observations. **Method:** Callaway and Sant'Anna (2021) staggered difference-in-differences with never-treated control group, regression estimation, standard errors clustered at the state level, 1,000 bootstrap iterations. **Sample:** 30 treated states across 16 adoption cohorts (2001–2020) and 10 never-treated states (states that adopted NIBRS before 2000 or after 2020). States with missing Disaster Center data excluded. SDE = $\hat{\beta}/\text{SD}(Y)$ where SD(Y) is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).

Acknowledgements

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

Contributors: @olafdrw

First Contributor: https://github.com/olafdrw

Project Repository: https://github.com/SocialCatalystLab/ape-papers