

Managerial Accountability and Frontline Worker Safety: Evidence from New York City Policing Reform

Kaiyi Wen^a Hyeran Chung^a Xiaolong Hou^b

Yichu Li^c Rhet A. Smith^a

October 5, 2026

^a Department of Economics and Finance, University of Texas at El Paso.

^b Department of Economics, Lahore University of Management Sciences.

^c Jinhe Center for Economic Research, Xi'an Jiaotong University.

Corresponding author: Kaiyi Wen, kwen@utep.edu

Abstract

Managerial accountability reforms are commonly implemented to improve organizational performance by increasing transparency and reducing information asymmetry. Although motivated by efficiency, such reforms may also generate complementary improvements in worker safety. This paper explores these second-order effects in a law enforcement setting. We leverage New York City's 1994 implementation of a data-driven police management system (Compstat) in a difference-in-differences framework, using data from the FBI's Law Enforcement Officers Killed and Assaulted (LEOKA) program and a control group of comparable large-city police departments to estimate a 30% decline in officer assaults following Compstat adoption. The effects strengthen over time, are concentrated in team-patrol encounters and evening hours, and are robust to multiple specifications. Thus, while reforms are efficiency-driven, the returns to such policies are enhanced by improvements in workplace safety.

JEL Codes: **K42, D73, H11, J28**

Keywords: **Police Reform, Public-sector Incentives, Principal-Agent, Police Accountability, Workplace Safety, Compstat**

1 Introduction

Organizations routinely invest in accountability systems to improve productivity, operational efficiency, and organizational performance. These systems mitigate principal-agent problems by reducing information asymmetries and strengthening incentives for managers and employees to pursue organizational objectives. With tangible performance metrics and regular evaluation, accountability reforms increase the costs of poor performance while rewarding effective management. Although these reforms are typically implemented to improve efficiency, the same mechanisms that improve organizational performance may likewise influence worker safety. If improvements in accountability-oriented management similarly improve workplace safety, then the returns to organizational reform extend beyond efficiency alone.

Although the returns to managerial accountability are often studied in private-sector organizations, similar incentive and information problems arise throughout the public sector. These challenges may be particularly pronounced in bureaucratic organizations, where hierarchical structures and delegated decision-making can weaken monitoring and create divergence between managerial and employee incentives, making the design of effective performance incentives particularly important (Mookherjee, 2006; Burgess et al., 2017). Police departments, in particular, operate in environments where managerial oversight influences not only organizational performance but also the safety of frontline officers. It makes them a unique setting for examining whether accountability reforms generate benefits beyond their intended objectives.

Officer safety is a central concern of police administration. Assaults against law enforcement officers impose direct costs, including physical injury, psychological trauma, and lost duty time; as well as indirect costs in reduced morale, higher turnover, and diminished recruitment. In 2022, the FBI’s Law Enforcement Officers Killed and Assaulted (LEOKA) program reported over 79,000 officer assaults in the United States, resulting in nearly 18,000 injuries (Federal Bureau of Investigation, 2022). Yet despite this scale, little causal evidence exists on whether organizational reforms within police departments can systematically reduce

officer victimization. This paper addresses this gap by leveraging the sharp, agency-specific introduction of Compstat within the New York Police Department (NYPD) to examine how accountability reforms affected officer safety.

Compstat, short for Computer Statistics, is a management information and accountability system introduced by New York City Police Commissioner William Bratton in January 1994 (Bratton and Knobler, 1998). The program established a weekly cycle of precinct-level crime-data reporting, real-time crime mapping, and intensive performance reviews in which precinct commanders, for the first time, were held publicly accountable for precinct-level crime outcomes during regular meetings. Prior to Compstat adoption, performance measurements and incentives for precinct commanders were limited. Misaligned incentives between police officers, commanders, and departmental objectives along with information asymmetry created principal-agent issues and hindered efficiency improvements. Compstat alleviated these principal-agent distortions by requiring accurate and timely crime statistics, targeted responses, and bi-weekly reviews with opportunity for commander advancement or demotion (Weisburd et al., 2003; Willis et al., 2003). By increasing performance transparency and managerial oversight, Compstat better aligned the incentives facing precinct leadership and is widely credited with improving organizational performance as crime declined in New York City following its implementation (Eck and Maguire, 2000; Zimring, 2012).

Although the motivation for Compstat reform was to improve efficiency and reduce crime, the leadership accountability reforms also likely affected other dimensions of worker behavior. In this paper, we examine whether these managerial reforms also affected officer safety. Using annual agency-level data of officer assaults from the FBI’s LEOKA dataset for the years 1988–1999 (Federal Bureau of Investigation, 2023), we use a difference-in-differences (DD) design to causally estimate the impact of the Compstat adoption on NYPD officer assault rates. Relative to a control group of 31 large-city police departments that did not implement comparable reforms, Compstat adoption is associated with roughly 6 fewer officer assaults per 100 police officers per year. This reduction equates to a 30% decline in officer assaults

relative to the pre-Compstat mean. The effects on more severe assaults in which an officer sustains an injury is quantitatively similar, suggesting the effects are not driven by changes in administrative reporting.

We supplement the DD approach with the synthetic control method of Abadie et al. (2010), which constructs a data-driven counterfactual for NYPD by reweighting control cities to match NYPD’s specific pre-reform trajectory. The estimated time-varying effects from each approach strengthen over time suggesting that the Compstat program effectiveness increased with officers and supervisors becoming increasingly familiar with managerial practices and integrating them more fully into routine operations. While this initiative notably improved efficiency and reduced crime (Eck and Maguire, 2000; Zimring, 2012), our results indicate that the Compstat reforms simultaneously generated second-order improvements in officer safety.

This paper contributes to multiple strands of literature across labor, public choice, and crime. Although ours is a public sector setting, economically significant improvements in worker safety resulting from greater supervisor accountability may extend to other industries. Workplace injuries in the U.S. impose significant economic costs that were estimated to exceed \$250 billion in 2007 (Leigh, 2011). During the last year of our sample period, the Bureau of Labor Statistics estimated there were 6.3 nonfatal occupational injuries and illnesses per 100 full time equivalent workers (U.S. Bureau of Labor Statistics, 2000). While law enforcement comprises a small fraction of the U.S. labor force, the rate of law enforcement officers treated in emergency departments from workplace events was three times higher than the rates from all other U.S. workers (Tiesman et al., 2018).¹ Although private employers can internalize much of these costs through workers’ compensation insurance, compliance with government safety regulations, and protection against retaliatory firings (Johnson et al., 2024), costs of workplace injuries in public sector settings are only partially internalized with taxpayers bearing much of the financial burden.

¹BLS does not publish an incidence rate for nonfatal injuries and illnesses for law enforcement agencies.

Accordingly, this paper also contributes to the public choice literature on bureaucratic incentives and organizational governance (Adamson and Rentschler, 2023; Fegley and Mur-tazashvili, 2023). Within public organizations especially, principal-agent problems can lead to inefficiency when monitoring is imperfect and leadership lacks effective mechanisms to align employee behavior with organizational objectives (Dixit, 2002; Mookherjee, 2006). Compstat represents an institutional innovation that strengthened monitoring, performance measurement, and managerial accountability within a large public bureaucracy. Our results suggest that the benefits of these reforms extend beyond their intended objective of reducing crime and lead to safer working conditions for police officers.

Finally, this paper contributes to a broader literature examining how performance measurement, monitoring technologies, and managerial accountability influence organizational performance. While accountability reforms are often met with reluctance by employees because they increase managerial oversight and limit discretion, greater performance transparency can improve worker behavior (Bernstein and Li, 2025) and organizational effectiveness (Burgess et al., 2017). Our paper extends these findings to a law enforcement setting by documenting an improvement in workplace safety. Given that reforms emphasizing individual-performance and accountability in the public sector can also induce gaming or adverse behavioral responses (Martins and Ferreira, 2025), the beneficial second-order effect we document becomes all the more notable. Like other technologies to improve police accountability, such as body-worn cameras, which improved police outcomes by reducing complaints against officers (Lum et al., 2020) and instances of police use of force (Campbell, 2024; Kim, 2025), Compstat strengthened accountability and transparency within the NYPD while also improving officer safety. As such, evaluations of the social impacts of policing reforms such as the Compstat adoption will understate the benefits if they do not incorporate corresponding improvements in officer safety.

The rest of the paper proceeds as follows. Section 2 provides a background of the Compstat reform and develops the conceptual framework of how the reform impacts officer safety.

Section 3 and Section 4 describe the data and the difference-in-differences estimation strategy. Section 5 provides the main results, numerous robustness tests, and sub-analyses to explore potential mechanisms. Section 6 details the synthetic control model approach and provides the results while Section 7 concludes.

2 Background and Conceptual Framework

With approximately 27,000 officers, the NYPD is the largest municipal police department in the United States. Prior to Compstat adoption, the department was hindered by delayed crime information, inefficient resource deployment, and limited managerial accountability. Specifically, crime data were compiled monthly through a paper-based system leading to lags between field conditions and administrative response, officer deployment was governed primarily by tradition and labor agreements rather than real-time evidence, and precinct commanders had wide operational discretion but faced limited scrutiny of their results (Bratton and Knobler, 1998).

In the pre-reform regime, precinct commanders operated in an environment broadly characterized by three institutional constraints. First, *information asymmetry*: the Commissioner observed precinct-level outcomes only with long lags: monthly paper-based crime reports that lagged field conditions by weeks, making it difficult to distinguish commander effort from environmental noise (Holmström, 1979). Second, *multitask incentive distortion*: commanders were responsible for a broad portfolio of outputs, including crime levels, community relations, and departmental politics. In the absence of objective, frequent performance metrics, effort is allocated toward tasks most visible to superiors (Holmström and Milgrom, 1991). Officer safety investments require costly attention to deployment configuration and patrol protocols but generated limited career rewards in a low-transparency environment. Third, *weak career-concern incentives*: without regular, transparent performance review, reputational updating about managerial ability is slow (Holmström, 1999). Public sector

constraints further attenuate incentives: monetary rewards are limited, promotions partly reflect seniority, and external labor market discipline is absent (Dixit, 2002).

Together, these constraints created a setting in which the commissioner had limited ability to monitor managerial effort and commanders had weak incentives to prioritize dimensions of performance that were difficult to observe or reward. We formalize the resulting accountability and information asymmetry problem through a two-tier principal-agent framework. The principal is the Police Commissioner, who sets departmental objectives but cannot directly observe the effort and tactical decisions of subordinates. The middle agent is the precinct commander, who makes deployment and operational decisions that affect the exposure of line officers to assault risk. Patrol officers are the line agents whose safety outcomes depend on the commander's choices. Figure 1 summarizes the causal structure. Prior to Compstat adoption, precinct commanders under-invest in officer-safety-enhancing activities, including optimized deployment patterns, tactical coordination, and backup protocols, all of which fall short of the social optimum, because the costs of such investment fall on the commander while the accountability benefits are diffuse and slow to materialize.

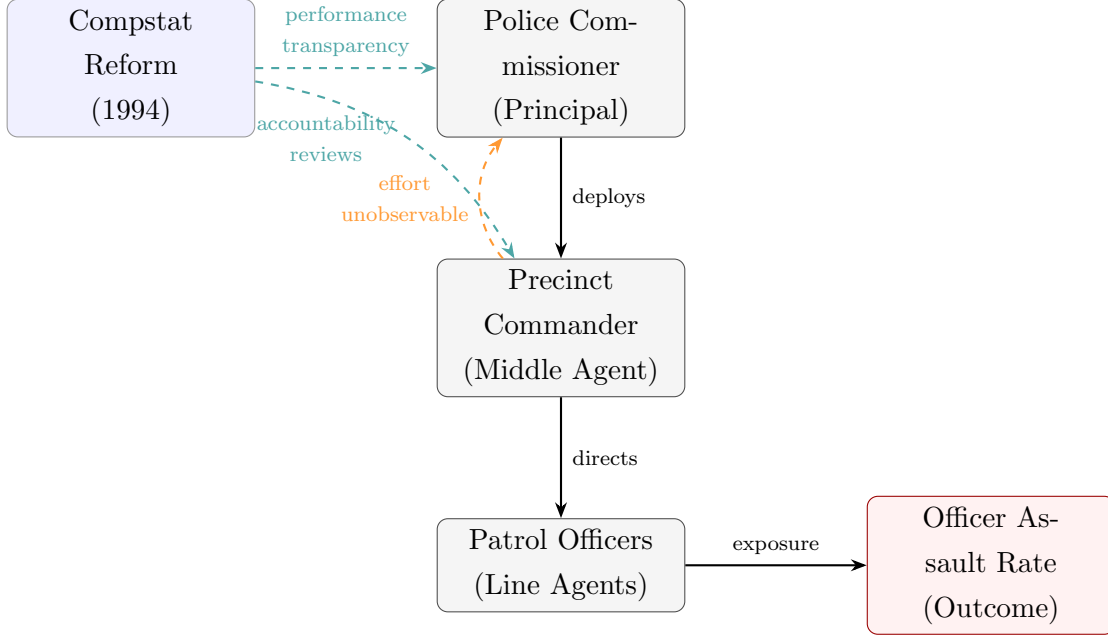


Figure 1: Principal-Agent Structure: Compstat and Officer Safety

Notes: Solid arrows indicate organizational authority. Dashed teal arrows show information flows introduced by Compstat (performance transparency and accountability reviews). Dashed gray arrow indicates the pre-reform information asymmetry between the Commissioner and precinct commanders that Compstat disrupts.

Compstat was introduced in January 1994 under Commissioner William Bratton and Deputy Commissioner Jack Maple (Bratton and Knobler, 1998). The system had four interdependent components: (1) *accurate and timely intelligence*, with weekly crime statistics compiled at the precinct level and distributed department-wide within 24 hours; (2) *rapid deployment*, requiring commanders to respond immediately to emerging patterns with targeted officer deployment; (3) *effective tactics*, with precinct-specific enforcement strategies required for each crime category; and (4) *relentless follow-up*, with bi-weekly reviews at One Police Plaza in which commanders were questioned intensively about crime trends, resource allocations, and tactical responses in front of senior leadership (Weisburd et al., 2003; Willis et al., 2003).

These four components alleviated the principal-agent issues by requiring greater performance transparency and enforcing greater accountability among precinct commanders for measured outcomes (dashed arrows in Figure 1). Weekly crime mapping and bi-weekly

command accountability sessions at One Police Plaza generated high-frequency, publicly observable signals about precinct-level outcomes and reduced the information asymmetry between the commissioner and precinct commanders. Commanders were held accountable on all visible precinct outcomes, including officer-involved incidents. Strong reviews could lead to advancement opportunities, while sustained underperformance could result in demotion, strengthening career-concern incentives in a setting where monetary incentives were limited (Holmström, 1999; Dixit, 2002).

Although these policing strategy reforms were designed to improve officer efficiency and reduce crime, they simultaneously enhanced officer safety through related mechanisms. For instance, real-time crime mapping and weekly hotspot analysis allowed commanders to concentrate team deployments in high-risk areas under controlled, prepared conditions. While this strategy combats crime more efficiently, it likewise reduces the frequency of unanticipated, higher-risk, solo confrontations in which officer assaults are more common (Garner et al., 1995). In addition, the accountability culture increased attention to tactical preparation, including backup protocols and situational briefings, which prior research links to reduced assault escalation risk (Smith, 1986). The concerted preparation efforts likely spill over into officer safety (Smith, 1986), reducing multitask incentive issues and placing greater weight on officer safety in performance evaluations (Holmström and Milgrom, 1991). Last, to the extent that Compstat increased visible police presence and apprehension risk in high-crime areas, potential assailants may have been deterred from attacking officers as a byproduct of the broader crime suppression mechanism. Collectively, Compstat adoption incentivized commanders to reallocate deployment toward configurations that reduce officer exposure to high-risk encounters by increasing team-based patrol (which reduces vulnerability from unanticipated solo confrontations), concentrate coverage during high-assault time windows, and improve tactical protocols for frequent-incident settings. We explore these specific margins in extensions of our main results in Section 5 with sub-analyses by assault incident characteristics.

Greater information management and utilization following Compstat adoption further facilitated a broader set of NYPD initiatives. For instance, Mayor Giuliani and Commissioner Bratton soon encouraged a *broken windows* policing philosophy that directed patrol officers to enforce quality-of-life violations such as public intoxication and fare evasion that had previously been deprioritized (Kelling and Coles, 1996). Compstat provided commanders with timely information on the geographic distribution of these and other offenses, allowing them to direct patrol resources toward areas of concentrated disorder while holding commanders accountable for addressing local crime problems. The NYPD’s *broken windows* approach also expanded stop-and-frisk practices. However, much of the increase in police stops under the stop-and-frisk approach occurred after our sample period when police stops became a performance measure in 2002 (Ferrandino, 2012; Meares, 2014). Still, these methods of utilization of Compstat data management may affect officer safety in theoretically ambiguous manner. While each policy increases the volume of discretionary police-citizen contacts and may escalate otherwise routine encounters into assault-involved events, the proactive policing techniques may also reduce the amount of higher-risk encounters through police avoidance by offenders.²

The composition and size of the NYPD also underwent considerable changes during the 1990s. In 1994, the NYPD’s uniformed headcount grew by several thousand through the federally funded COPS Hiring Program. While the headcounts in other comparison police departments also grew, the NYPD expansion was more significant. In a similar way that workplace injury risk declines with worker age and experience (Guerin et al., 2020), assaults against police officers likewise become less common as officers gain experience (Tiesman et al., 2018). The influx of new officers in the post-Compstat era puts upward pressure on officer assaults and positively influence estimated effects of the Compstat reform on officer assaults. In addition, the NYC Transit Police Department (approximately 4,000 officers) and the NYC Housing Authority Police Department (approximately 2,500 officers) merged into

²We separately examine assaults that involve officer injuries as well as less severe assaults to assess the relative importance of these potential channels.

the NYPD in 1995. Similar to increased headcounts, the inclusion of assault metrics from other departments in the NYPD count during the post-Compstat time period will likewise put upward pressure on estimated effects on officer assaults.³

3 Data

Officer assault data are obtained from the FBI’s Law Enforcement Officers Killed and Assaulted (LEOKA) Data Collection (LEOKA) program. These data provide annual agency-level counts of officers killed and assaulted for participating law enforcement agencies nationwide and include counts of total assault, assaults involving injuries or deaths, assaults by patrol circumstance (solo, assisted), the time of day the assault occurred, and the type of weapon involved (firearm, knife, personal weapons, other). Our outcomes of interest include total assaults as well as more severe assaults in which officers sustain injuries. These outcomes are scaled into rates per 100 officers using the number of sworn officers within each agency that are also provided in the data.

The NYPD is the sole police department with jurisdiction over all five New York City boroughs. Notably, the NYPD does not report to the LEOKA program in 2000 or 2001. Thus, our sample period begins in 1988 to allow for 6 years of pre-intervention observations and spans through 1999. For comparison law enforcement agencies, we comprise the rest of the sample with large-city police departments that most resemble the New York City Police Department. Specifically, police departments in our sample must have an urban jurisdiction population exceeding 500,000 people and be classified as a local police department. While the 500,000 population threshold is chosen for greater comparability, our results are robust to the inclusion of city police agencies with jurisdiction populations of at least 300,000. Our sample also includes county-wide departments serving urban cores, such as Las Vegas Metropolitan Police Department and Miami-Dade Police Department in the final sample.

³For comparability across police departments, officer assaults are converted into per officer rates in our analysis. To accommodate changes in NYPD composition, we also scale assaults by local population and estimate qualitatively similar effects.

However, suburban county-level sheriff departments that serve a more dispersed residential population are omitted.⁴

Because LEOKA reporting is not mandatory at the federal level, reporting practices may vary across agencies and over time. Agencies submit their officer-employment counts and their assault counts on separate forms, and an agency that files the former but not the latter in a given year is recorded as reporting zero assaults rather than as missing. In the raw file a non-report is therefore indistinguishable from a genuine zero-assault year. Because every department in our sample is a large-city agency employing hundreds to thousands of sworn officers, a reported count of exactly zero assaults in a year is not credible and instead signals a missing observation.⁵ Missing observations are observed across ten police departments over our sample period.⁶

The final sample consists of an unbalanced panel of 32 agencies (the NYPD and 31 control departments) observed from 1988 through 1999. After the reporting correction described above, 352 of the 384 agency-years have non-missing assault data. The summary statistics for both the pre-reform (1988–1993) and post-reform (1994–1999) periods are provided in Table 1. In the pre-reform period, NYPD’s assault rate (19.7 per 100 officer-years) is broadly comparable to the control-group mean (23.7). In the post-reform period, NYPD’s assault rate falls sharply to 8.8 per 100 officer-years, a 55 percent reduction from its pre-reform mean. Although officer assaults among the other major city police departments also falls, the decline is more modest (20 percent). The pattern is similar for injury-involved assaults, for which the decline in NYPD rates is more pronounced than the corresponding change among other major city police departments. These univariate comparisons suggest a significant

⁴Ten suburban county departments are excluded: Suffolk County PD and Nassau County PD (suburban Long Island), Fairfax County PD (suburban Washington, DC), Montgomery County PD, Baltimore County PD, Prince George’s County PD, and Anne Arundel County PD (suburban Baltimore and DC), and Gwinnett County PD, Cobb County PD, and DeKalb County PD (suburban Atlanta).

⁵The Chicago Police Department, for example, reports roughly 900 assaults per year through 1994 and then exactly zero for 1995–1999 despite employing more than 13,000 officers. We set these non-reported cells to missing rather than treat them as true zeros.

⁶Uneven reporting and missing observations are common in crime data. Maltz and Targonski (2002) show that imputation processes can create additional issues. As such, we do not impute the missing values and our sample is unbalanced.

improvement in NYPD officer safety during second half of the 1990s that is not attributable to nationwide trends in officer assaults.

Table 1: Officer Assault Summary Statistics: Pre- and Post-Reforms

	NYPD				Other Major City PDs			
	Pre (1988–93)		Post (1994–99)		Pre (1988–93)		Post (1994–99)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>Primary outcomes (per 100 officers)</i>								
Assault rate	19.71	1.63	8.79	3.76	23.68	17.77	18.91	15.16
Injury assault rate	18.11	1.35	8.18	3.65	7.81	5.14	5.72	4.83
<i>Patrol circumstance (per 100 officers)</i>								
Solo-patrol assault rate	2.41	0.22	0.92	0.47	6.57	6.47	4.86	4.19
Team-patrol assault rate	17.30	1.45	6.65	3.61	17.12	13.40	14.06	12.37
Solo share of assaults	0.122	0.005	0.122	0.012	0.279	0.153	0.277	0.164
<i>Time of day (per 100 officers)</i>								
Day-shift assault rate	4.1	0.2	1.6	0.9	3.23	2.43	2.95	2.27
Evening-shift assault rate	8.3	0.9	3.2	1.7	8.92	6.79	7.55	6.08
Night-shift assault rate	7.4	0.8	2.8	1.5	11.53	9.36	8.41	7.61
<i>Agency characteristics</i>								
Female officer share	0.133	0.010	0.149	0.004	0.117	0.039	0.140	0.044
Total sworn officers	27,310	1,251	37,139	3,869	2,490	2,395	2,742	2,602
Observations (agency-years)	6		6		171		169	

Officer assault data are obtained from the FBI’s LEOKA data program for the years 1988–1999. The first two columns describe officer assault rates for the NYPD (ORI NY03030) separated by pre- and post-Compstat reforms. The last two columns provide comparable statistics from the 31 other large city police departments in the sample. Assault rates are expressed per 100 sworn officers per year while solo share and female officer share are proportions (0–1). Day shifts are defined as events that occur between 06:01–14:00, evening assaults occur between 14:01–22:00, and night assaults occur 22:01–06:00. NYPD 1995 assault rate and injury assault rate are imputed based on Jan–Apr LEOKA reports and NYPD 1992–1994 seasonal share; 1995 breakdown variables (solo, team, time-of-day) reflect Jan–Apr 1995 only and may modestly understate the post-reform means for those components.

Further disaggregation by patrol circumstances suggest that the reduction in total assault rates may be driven by the significant decline in NYPD’s team-patrol assault rate. Last, the disproportionate increase in department size of the NYPD during the post reform years can be largely attributed to the 1995 Transit and Housing Police merger. While this affects

assault rates mechanically by increasing the denominator, we note that other major city police departments are also larger in post-reform period. Further, we show in sensitivity checks that our results are unaffected when scaling officer assaults by underlying jurisdiction populations that would not be affected by changes in agency size.

We supplement the LEOKA panel with officer workforce demographics from the Law Enforcement Management and Administrative Statistics (LEMAS) survey, conducted by the Bureau of Justice Statistics. LEMAS provides agency-level counts of full-time sworn officers by race and sex. We use the 1993 (ICPSR 6708) and 1997 (ICPSR 2700) waves for the NYPD.⁷ Figure 2 provides a visual summary of the racial composition differences between NYPD and the control group in the pre- and post-reform LEMAS waves.

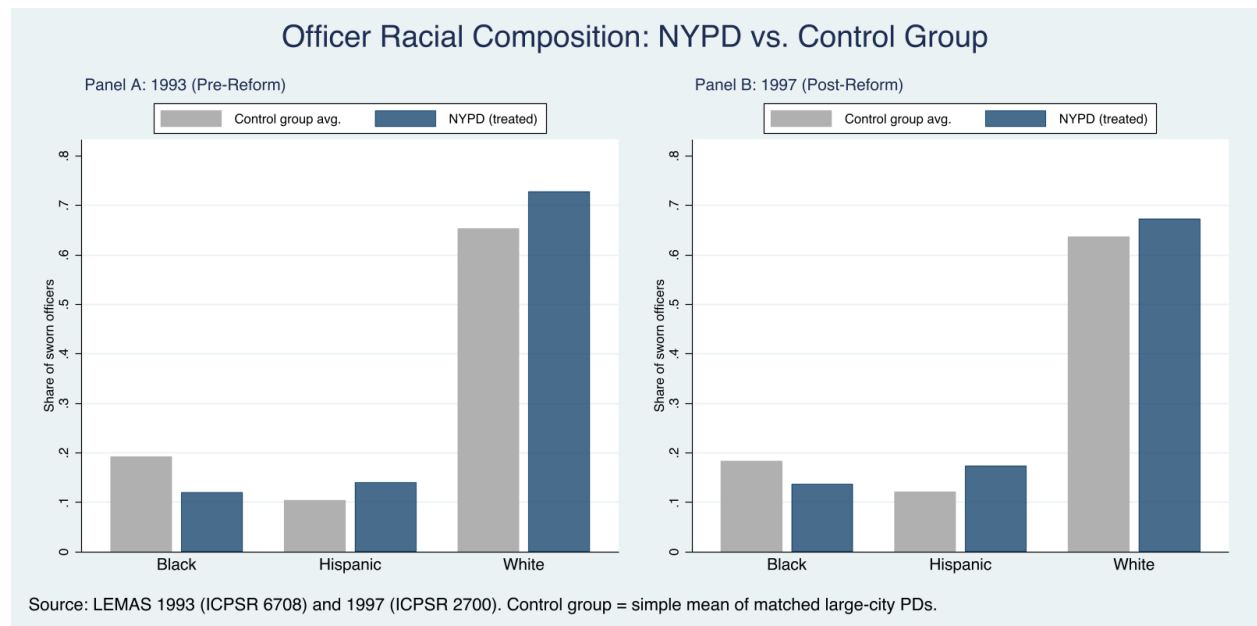


Figure 2: Officer Racial Composition: NYPD vs. Control Group, 1993 and 1997

Notes: Gray bars show the simple mean of matched large-city control departments; navy bars show NYPD. Shares are computed from LEMAS 1993 (ICPSR 6708) and LEMAS 1997 (ICPSR 2700). *White* includes White non-Hispanic male and female sworn officers; *Black* includes Black/African American male and female; *Hispanic* includes Hispanic male and female. Shares sum to less than one because Asian/Pacific Islander and other race categories are omitted. Control group comprises the same large-city departments used in the main synthetic control analysis.

In both 1993 (pre-reform) and 1997 (post-reform), the NYPD's officer workforce was

⁷The 1990 NYPD survey entry was not completed and contains only missing values. The 1999 wave (ICPSR 3079) reports total sworn personnel but omits the officer-race breakdown.

broadly comparable to the control group average on White and Black shares, while having a somewhat higher Hispanic representation. Importantly, no discrete compositional break distinguishes the NYPD from the control group between the two waves that would suggest officer assault rates are responding to changes in racial composition of law enforcement officers.

4 Empirical Strategy

We use a difference-in-differences (DD) framework to estimate the impact of the NYPD Compstat reforms on officer assaults relative to police departments in similar US cities that did not experience an intervention. We first estimate an event study model that simultaneously assesses pre-reform comparability in outcomes between the NYPD and the control group while flexibly allowing for post-reform effects to vary over time. This model takes the form

$$A_{it} = \alpha_i + \gamma_t + \sum_{k \neq -1} \beta_k \cdot \text{NYPD}_i \times \mathbf{1}\{t - 1994 = k\} + \varepsilon_{it}, \quad (1)$$

where A_{it} is the assault rate per 100 police officers in police department i in year t . β_k is vector of time-relative coefficients relative to the 1994 reform that ranges from -6 to $+6$. The period prior to the reform is normalized to zero and omitted as the reference period. The estimated effects of the policing reform on officer assaults are reflected in $\hat{\beta}_{k \geq 0}$, while trends in the pre-reform coefficients would indicate potential confounding omitted variables that would otherwise threaten the validity of our estimates. Agency and year fixed effects are represented by α_i and γ_t and absorb time-invariant differences across departments as well as national trends in officer assault rates common to all agencies. ε_{it} is the mean zero error term and standard errors are clustered at the agency level.

We also estimate the average annual effect of the reform on officer assaults using a

traditional DD model by adjusting the previous equation as follows

$$A_{it} = \alpha_i + \gamma_t + \beta \cdot \text{NYPD}_i \times \text{Post}_t + \varepsilon_{it}. \quad (2)$$

Here, $\text{NYPD}_i \times \text{Post}_t$ is equal to one for the NYPD in years after 1993 and is otherwise equal to zero. We estimate similar specifications across different outcomes, patrol conditions, and time of day to shed light on the potential mechanisms through which the reform affects officer safety. The key identifying assumption to our approach is parallel trends. That absent Compstat, NYPD’s officer assault rate would have evolved in parallel with the control group’s. Although this assumption is not directly testable, the estimates of pre-reform coefficients from the event study approach can provide credibility this assumption is not violated.

5 Reform Effects on Officer Assaults

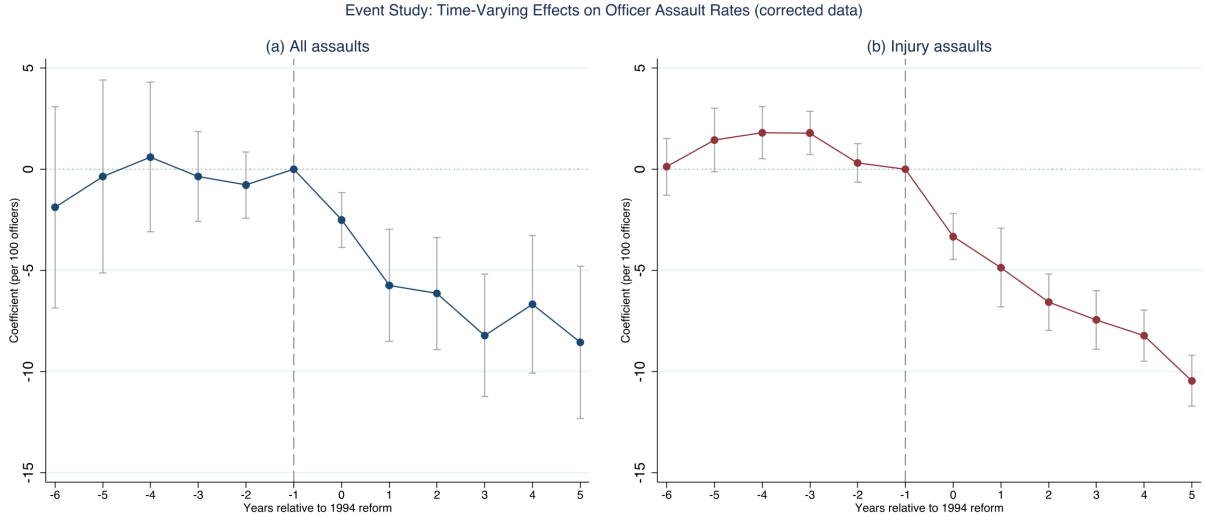


Figure 3: Time-Varying Effects of NYPD Reform on Officer Assault Rates

The figure depicts the estimated time-varying effects of the 1994 NYPD Compstat reform on assault rates per 100 officers. The whiskers describe the corresponding 95% confidence intervals. Standard errors are clustered at the agency level.

Estimates of the time-varying effects of the Compstat reform from the event study model

are depicted in Figure 3. The first figure describes the effects on all officer assaults while the second figure depicts the impact on more severe assaults that result in police injuries. Although we observe statistically significant pre-reform coefficients in the latter outcome, the pre-treatment coefficients are generally flat in both instances and provide little evidence of pre-reform trends that would weaken the credibility of the analysis. Following the intervention, the reduction in assault rates is immediate with the decline in total officer assaults leveling off at roughly 6-7 fewer assaults per 100 officers per year. The reduction in injury involved assaults is comparable in magnitude and grow stronger over time. This suggests that not only are officer assaults occurring less frequently, the severity of the assaults is also decreasing.

The corresponding average treatment effects of the reform on officer assault rates estimated from the DD model are reported Column (1) of Table 2. The 1994 reform centered on Compstat reduced the NYPD officer assault rate by 5.8 per 100 officer-years. This equates to a 30 percent decline relative to NYPD’s pre-reform mean. The effect of the reform on the more severe injury-related assaults rates is quantitatively similar, if not slightly larger. Injury related assaults fall by 7.7 assaults per 100 officers per year (a roughly 43% decline relative to the pre-treated mean). In addition to reductions in assault severity, the decline in injury-involved assaults also overcome concerns of reform-induced changes in assault reporting. Thus, the quantitatively similar effects across outcomes suggest the reduction in assault rates reflect improvements in officer safety and are not a result of changes in officer reporting standards.

As evident in columns (2)-(6) in Table 2, the estimated effects are stable across various specifications and sample sizes. Column (2) augments the baseline model by controlling for state-specific linear time trends. While the effects attenuate in magnitude, the estimates remain statistically significant and directionally consistent with the baseline estimates. Moreover, the estimated 18-19 percent reduction in officer assaults is still economically meaningful. Because the inclusion of time trends likely absorbs variation caused by the exogenous inter-

Table 2: DD Estimated Effects of the Compstat Reform on Officer Assaults

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	State Trends	Population Scaled	Sample >300,000	IHS	Poisson
<i>Total assaults</i>						
Compstat Reform	-5.838*** (1.976)	-3.479*** (0.782)	-22.574*** (4.835)	-8.226*** (1.840)	-0.612*** (0.095)	-0.563*** (0.069)
Implied % decline	30%	18%	31%	42%	46%	43%
<i>Injury assaults</i>						
Compstat Reform	-7.719*** (0.500)	-3.402*** (0.759)	-23.636*** (1.079)	-8.494*** (0.360)	-0.496*** (0.098)	-0.469*** (0.071)
Implied % decline	43%	19%	35%	47%	39%	38%
Agency FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	352	352	352	655	352	352
Clusters	32	32	32	56	32	32

Each column describes the DD estimated effects of the 1994 CompStat reform on the number of assaults and injury assaults per 100 officers. Pre-reform mean and implied % decline computed relative to NYPD pre-1994 average. Standard errors (in parentheses) are clustered at agency level, $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

vention (Wolfers, 2006), we view these estimates conditioned on time trends as lower bound estimates.

Recall that coinciding with the NYPD reform, in 1995 the NYC Transit and Housing Police Departments were merged with the NYPD. While the 6,500 officer increase unambiguously raises the count of officer assaults, the influx also affects the denominator of the scaled assault per 100 sworn officers. If Transit and Housing officers faced systematically lower assault rates than patrol officers, the merger could mechanically reduce NYPD's per-officer assault rate independent of any reform effect. We address this concern in column (3) in Table 2 with a per-population specification that uses the served population as the denominator instead of sworn officer count.⁸ When scaling officer assaults by jurisdiction population, the post-reform reduction on officer assaults and injury-involved assaults are quantitatively

⁸Population is reported in the LEOKA data and is entirely unaffected by the institutional merger.

similar to the baseline estimates (a 31% and 35% reduction relative to pre-reform means). Because the post-reform merger will put upward pressure on officer assaults per jurisdiction population and positively bias any estimated effect, the comparability across models alleviates concerns the reduction in assault rates are mechanically driven by an expanding officer base and provides further evidence of reform-induced improvements in officer safety.⁹

For better comparability, we limit our preferred sample to large city jurisdictions consistent with the urban policing literature’s standard definition (Weisburd et al., 2003). In column (4) of Table 2, we examine the sensitivity of this arbitrary threshold by expanding the sample to include smaller cities with jurisdiction populations of at least 300,000 people. The effect sizes are slightly larger in this expanded sample.¹⁰

In columns (5) and (6) of Table 2 we use an inverse hyperbolic sine transformation of the outcome to account for potential skewness and also estimate a Poisson to accommodate the count nature of the data.¹¹ In both specifications across both outcomes, the effects are comparable to the estimates from baseline linear model in column (1).

Last, agency-level clustered standard errors are provided in Table 2 for inference. However, because the treatment is assigned to a single unit and the number of groups small, we complement the previous asymptotic inference approximations by implementing a variant of Fisher’s (1935) permutation test as done in similar empirical work (Buchmueller et al., 2011; Cunningham and Shah, 2017; Johnson and Smith, 2023). In this approach, treatment is assigned to each agency in the control group as if it implemented the 1994 reform and collect the resulting placebo estimates. By comparing the effects of the reform on NYPD assaults to the distribution of placebo effects, we can evaluate the likelihood our estimated effects are attributable to the reform and as opposed to by chance.

The placebo distributions are plotted in Figure 4. The vertical solid black line indicates

⁹Appendix Figure A7 plots the corresponding event-study coefficients; pre-reform trends are flat and post-1994 coefficients turn sharply negative, mirroring the pattern in Figure 3.

¹⁰The estimates are consistent across varying population thresholds (400k, 250k, etc.). These results are available upon request.

¹¹The asinh transformation closely parallels the natural log transformation, but is well defined at zero (Card and DellaVigna, 2020).

the actual estimated effects for the NYPD as reported in column (1) in Table 2. The tan bars plot the distribution of the placebo estimates and the red vertical dashed lines represent the 5th and 95th percentiles of each distribution. Although the estimated effect of the reform on total officer assault rates in the NYPD does not fall below the 5th percentile, the estimate lies in the left tail of the placebo distribution, with 6 of the 31 control agencies experiencing larger declines ($p\text{-value} = 7/32 = 0.219$). With a single treated department and a modest number of placebo estimates, conventional thresholds for statistical significance are demanding (Ferman and Pinto, 2019). In our context, the most extreme possible estimate would yield a p -value of approximately $1/32 = 0.031$. However, Romer (2020) cautions against relying on conventional significance thresholds and argues that an estimate’s distance from the placebo distribution, measured in standard deviations, is more informative, with roughly one standard deviation providing moderate evidence of an effect. The NYPD’s total-assault estimate lies about 0.6 standard deviations below the placebo mean. Because this dispersion is inflated by a small number of poorly-fit control agencies with extreme placebo estimates in both tails, as a sensitivity check, a trimmed standard deviation that excludes these outliers places the estimate roughly one standard deviation below the mean. Combined with the stability of the estimates across specifications in Table 2, this provides reasonable evidence that the Compstat reform reduced officer assaults overall.

The NYPD’s estimate (-7.72) for injury assault rates is the most negative of all 32 agencies and lies roughly 2.8 standard deviations below the mean of the placebo distribution. Thus, while the placebo-based inference for total assault rates is weaker than the evidence from the other approaches, the more pronounced separation of the NYPD’s injury-assault estimate from the placebo distribution provides strong evidence of a reduction in injury-related assaults and reinforces the broader evidence that the Compstat reform improved officer safety.

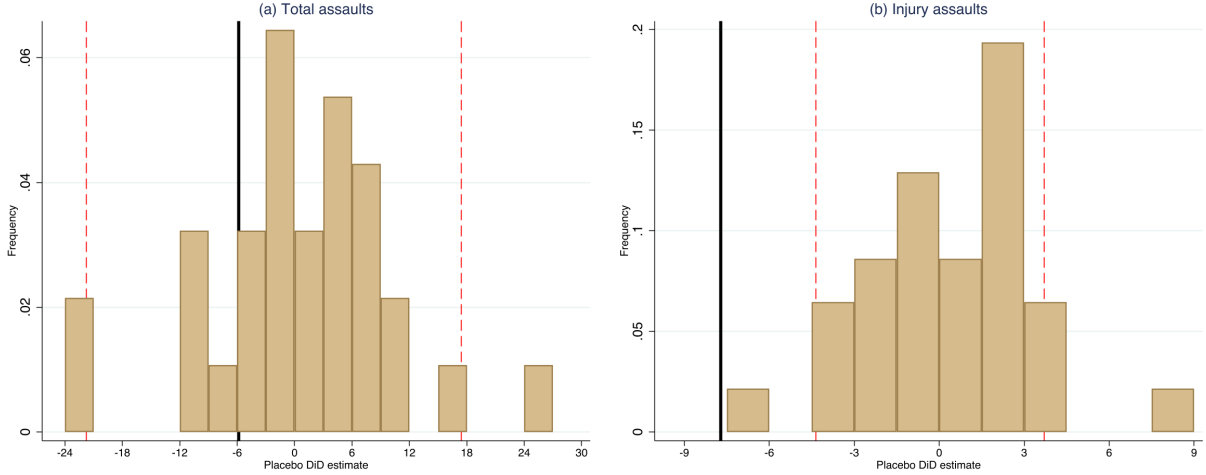


Figure 4: Randomization Inference: Placebo Distribution of DiD Estimates

Notes: Each tan bar is a placebo estimate obtained by reassigning the NYPD- $\times$ -post treatment to a control agency and re-estimating the TWFE difference-in-differences model (Table 2, columns 1 and 2). The bold black vertical line marks NYPD’s actual estimate; red dashed lines mark the 5th and 95th percentiles of the placebo distribution. Left panel: total assault rate; right panel: injury assault rate. Estimates are in per 100 officer-years and centered near zero under the null.

5.1 Mechanisms of Reform Effects on Officer Safety

To gain insight into the channels through which officer safety improves, we extend the analysis and disaggregate assaults by patrol circumstance, the time of day assaults occur, weapons involved, and whether gender composition moderates the effect. The results of these sub-analyses are provided in Table 3.

Three patterns emerge. First, the reduction is concentrated in team-patrol encounters and is absent in solo-patrol contacts (Panel A). This pattern is consistent with the mechanism that Compstat improved the preparation and positioning of team deployments. Greater information accessibility and real-time crime mapping may have allowed commanders to direct teams to known hotspots under more controlled conditions. Conversely, solo contacts, which are harder to pre-position and more idiosyncratic, show no corresponding improvement.

Second, although reductions in officer assaults are observed across all eight-hour windows (Panel B), the decline is most pronounced during evening hours when street activity is highest

Table 3: Mechanism Analysis: Decomposition of Assault Rate Reduction

Outcome	Coeff.	SE
<i>Panel A: Patrol Circumstance</i>		
Solo-patrol assault rate	+0.395	(0.783)
Team-patrol assault rate	−7.454***	(1.503)
Solo share of assaults ($\times 100$)	−0.538	(1.513)
<i>Panel B: Time of day (8-hour shifts)</i>		
Day (06:01–14:00)	−2.176***	(0.249)
Evening (14:01–22:00)	−3.613***	(0.698)
Night (22:01–06:00)	−1.270	(1.105)
<i>Panel C: Weapon type</i>		
Firearm assault rate	−1.736***	(0.124)
Knife assault rate	−0.122*	(0.070)
Unarmed assault rate	−3.621*	(1.785)
Other-weapon assault rate	−1.581***	(0.194)
<i>Panel D: Gender heterogeneity</i>		
Compstat Reform (main)	−5.598**	(2.599)
Compstat Reform $\times$ Female share (std.)	+0.277	(1.805)

Each row in Panels A–C: separate TWFE regression per equation (2). Panel D: augmented regression including z-scored female share and interaction. All regressions include agency and year FE; SE clustered at agency level. Outcomes scaled to per 100 sworn officers per year, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

(Cohen and Felson, 1979) and data-driven resource allocation can have the largest impact. Although negative, the effects on assaults attenuate slightly and lose statistical significance during night hours. Similar to solo-patrol environments, post-reform improvements in officer safety again may be mitigated by idiosyncratic or erratic late night environments.

Third, the reduction is broad-based across all weapon types (Panel C) suggesting a general reduction in confrontational encounters rather than any specific tactical change. The largest decline in assaults are those not involving a weapon. While this strengthens the argument of reductions in hostile interactions overall, it also importantly alleviates concerns the reduction in assaults may be driven by other policies (such as firearm restrictions).

Last, we augment equation (2) with the interaction of the standardized female officer share and the post-reform indicator (Panel D). The estimated effect of the reform on officer assaults is quantitatively similar to the baseline estimates and the interaction term is not distinguishable from zero providing no evidence that departments with higher female representation experienced differential treatment effects. The safety improvement appears to operate through organizational practices rather than workforce composition channels.

Taken together, the results in Table 3 suggest the Compstat reforms improved officer safety by improving the tactical preparation and deployment positioning of officer teams, particularly during high-activity evening hours, and by reducing the frequency of uncontrolled confrontations across all encounter types. This evidence is consistent with accountability and deployment channels discussed in Section 2. However, because of data limitations, we interpret these compositional patterns as suggestive. The LEOKA data report assault events but not officer *exposure* by patrol circumstance, and the underlying rate of officer-citizen contact in each setting may also be changing. Thus, while the concentration of the decline in team-patrol and evening encounters is consistent with a deployment-optimization channel, we cannot fully distinguish a genuine change in deployment from a shift in the composition of encounters without an exposure measure. We therefore read the mechanism evidence as indicative of where officer safety improves rather than as conclusively identifying the channel through which it did.

6 Reform Effects on Officer Safety: Synthetic Control Approach

Although sample restrictions to major US city police departments eliminates the potential for comparisons of the NYPD to non-comparable police departments with rural or small city jurisdictions, some police departments in major US cities may not be comparable to the NYPD in the pre-reform period. Any systematic differences in characteristics may influence

post-reform differences and corresponding estimated treatment effects. The sythentic control approach overcomes concerns of poor comparability by assigning higher weights to police departments that more closely resemble the NYPD prior to Compstat intervention. The synthetic control approach applies a vector of weights to the pool of control city police departments to construct a synthetic unit that closely matches the pre-reform assault rates for NYPD officers (Abadie et al., 2010, 2015). This synthetic police department represents a counterfactual NYPD and describes how officer assaults would evolve had the Compstat intervention not occurred.

In this approach, the estimated effect of the intervention ($\alpha_{NY,t}$) is the difference in officer assaults predicted by the synthetic counterfactual and what is actually observed in the NYPD following the Compstat reform. Formally, in our scenario with $J + 1$ police departments and $j = 1$ for the NYPD, the treatment effect can be written as

$$\alpha_{1t} = A_{1t}^I - A_{1t}^N \quad (3)$$

where the post-Compstat officer assaults (A_{1t}^I) in the NYPD are observed while the assaults among NYPD officers absent intervention (A_{1t}^N) must be estimated. This data-driven process then constructs a convex combination of control agencies, with weights $\mathbf{w}^* = (w_1^*, \dots, w_J^*)$ satisfying $w_j^* \geq 0$ and $\sum_j w_j^* = 1$, chosen to minimize the pre-reform discrepancy between NYPD and its synthetic counterpart. The estimated treatment effect in post-reform year t is:

$$\hat{\alpha}_{1t} = A_{1t} - \sum_{j=2}^{J+1} w_j^* A_{jt}, \quad t \geq 1994. \quad (4)$$

The predictors used to construct the synthetic NYPD counterfactual include the assault rate in 1988, 1990, 1992, and 1993. We use multiple distinct years to pin down the trajectory shape, including both the level and the curvature of the pre-reform assault rate. Weights are estimated using nested optimization (**synth**, **nested allopt**), which iterates over both the predictor weight matrix V and the unit weight vector W to minimize the pre-treatment

mean squared prediction error. The pre-treatment fit period is 1988–1993.

Inference is permutation-based and similar to the approach used for the DD estimates in Figure 4. Here, we re-estimate the synthetic control for each of the 21 control agencies in the corrected donor pool as if it were the treated unit in 1994, using the same predictor set and optimization.¹² The key summary statistic is the post-to-pre mean squared prediction error (MSPE) ratio:

$$R_i = \frac{\text{Post-MSPE}_i}{\text{Pre-MSPE}_i}. \quad (5)$$

The MSPE ratio has a natural interpretation: the numerator measures how far the actual post-reform outcome deviates from the synthetic counterfactual, and the denominator measures how well the synthetic control fit the pre-reform data. A large ratio indicates an unusual post-reform break relative to how well the method could track the unit before treatment. Dividing by the pre-period MSPE normalizes for differences in pre-treatment fit: a unit whose synthetic control fits poorly before treatment will mechanically show large post-treatment gaps for reasons unrelated to treatment, so it should not count against NYPD in the distribution. Following Abadie et al. (2010), we exclude placebos whose pre-MSPE exceeds five times NYPD’s.

6.1 Synthetic Control Results

The synthetic control model-estimated effects of the reform on total officer assaults and assaults that result in officer injury are described in Figure 5. In the top row, the bold lines describe the observed assault rates of NYPD officers across the sample period while the dashed lines describe the assault rates of the synthetic NYPD constructed from other police agencies in the control pool. The vertical dashed lines separate the pre- and post-reform periods and the synthetic NYPD predicts the level of assaults the NYPD would have

¹²The donor pool comprises 21 of the 31 control agencies. As described in Section 3, ten agencies reported officer employment but non-reported (zero) assault counts in at least one year; because the synthetic control requires a complete outcome series for each donor, these agencies are dropped from the donor pool. The difference-in-differences and randomization-inference estimates, which do not require a complete series, retain all agencies for their non-missing years.

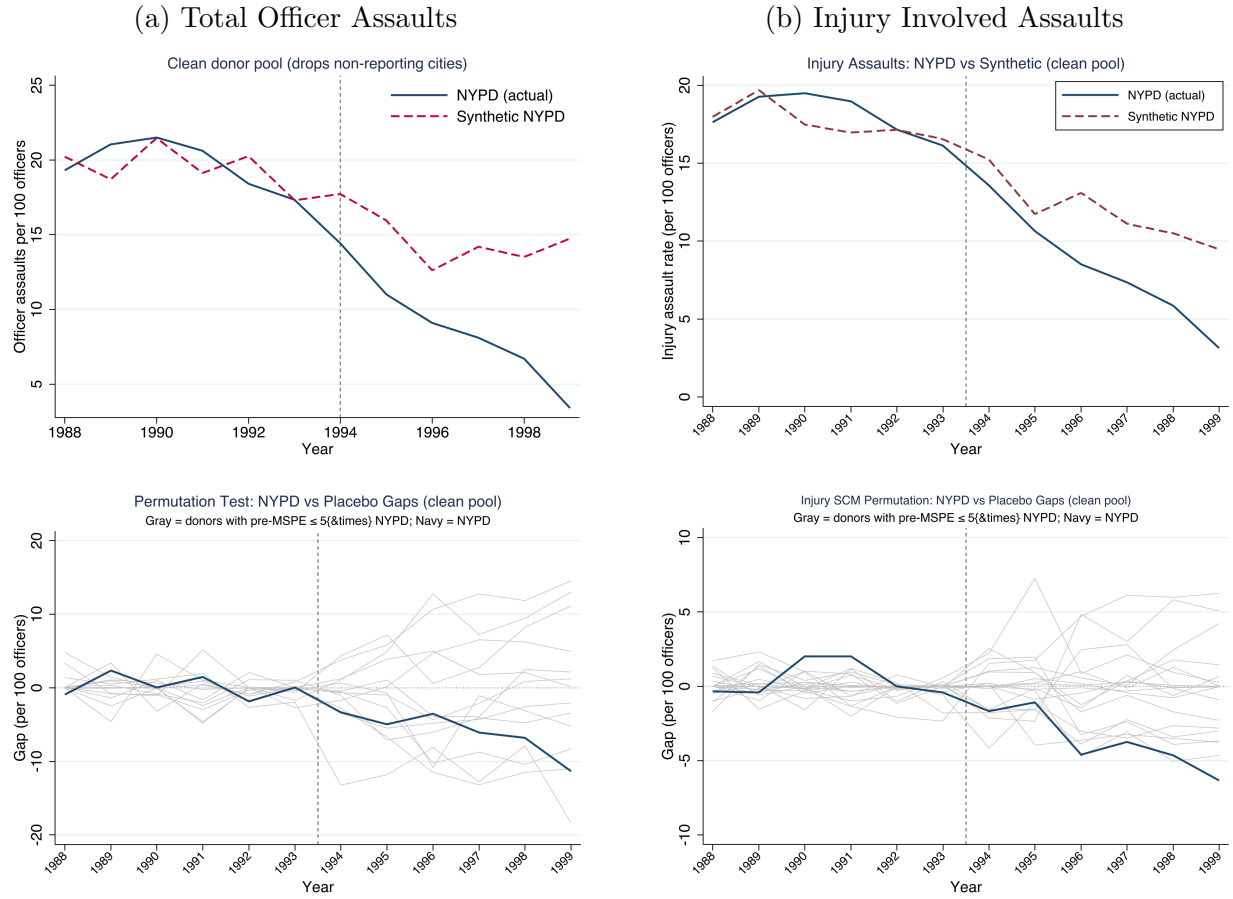


Figure 5: Synthetic Control: Officer Assault Rate

Notes: Left panel: NYPD actual (solid navy) vs. synthetic NYPD (dashed red). Right panel: gap (actual minus synthetic). Weights estimated via Abadie et al. (2010). Vertical dashed line: 1994 Compstat reform.

experienced absent of the 1994 reform. The effects of the policy are estimated as the vertical difference between the actual NYPD and its synthetic counterfactual.

The synthetic control closely tracks NYPD's pre-reform assault rates. Following 1994, however, the two series diverge. Although total assaults and injury-involved assaults decline for both the actual and synthetic NYPD in the post-reform period, the decline in both assault measures is larger for the actual NYPD. These results are consistent with the DD-estimated relative reductions in assaults previously discussed indicating reform-induced improvement in officer safety.

The bottom row of Figure 5 plots the estimated gap, the vertical difference between the actual and synthetic NYPD, in bold navy, overlaid on the gray placebo gaps obtained by

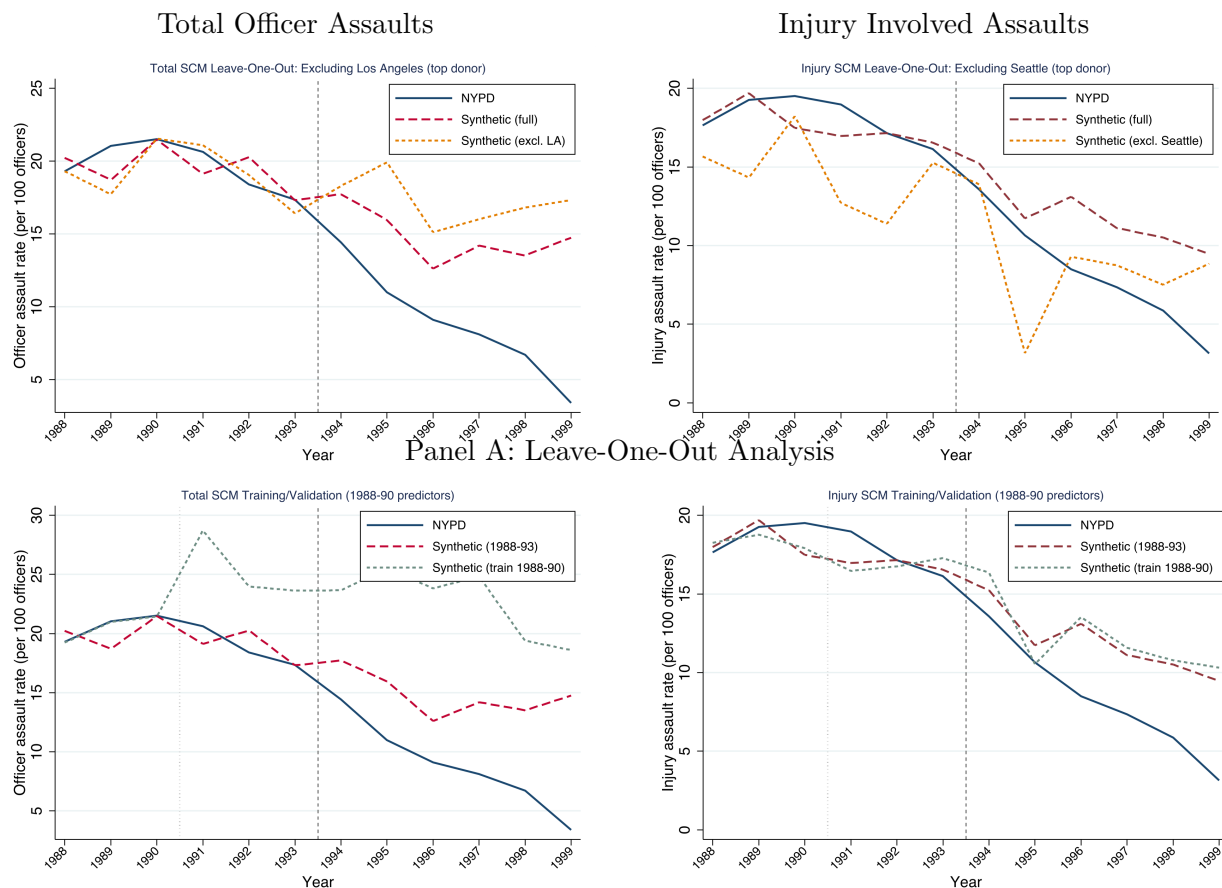
reassigning the 1994 reform to each control agency in turn. For both outcomes, NYPD’s post-reform gap turns negative and widens through 1999, placing it toward the lower end of the placebo distribution.

Formal permutation inference, based on the ratio of post- to pre-reform mean squared prediction error for the entire post-treatment period, yields a p -value of approximately 0.24 for total assaults and 0.33 for injury assaults. NYPD is among the more extreme units but is not the single most extreme in either synthetic-control distribution, and neither p -value attains conventional significance. However, with only 21 placebo units the permutation distribution is coarse and lacks statistical power.¹³ What the synthetic control contributes is not an independent significance test, but rather a transparent, nonparametric counterfactual. NYPD’s gap relative to its synthetic counterpart is near zero through 1993 and grows monotonically after 1994. The SCM results suggest that this dynamic pattern is not simply a product of poor comparability to agencies in the control group: although officer assaults declined nationwide during the 1990s, the reduction was substantially more pronounced in NYPD following the adoption of Compstat.

6.2 Synthetic Control Sensitivity Analyses

We use the synthetic control model as a complement to the traditional DD approach. This alternative approach alleviates concerns of poor comparability between treated and control units by assigning higher weights to untreated police agencies that are more similar to the NYPD prior to the 1994 reform. By constructing a reweighted counterfactual that closely tracks NYPD’s specific pre-reform trajectory, rather than assuming all control units are equally valid comparisons, it makes the identifying assumption transparent and testable. However, the SCM results depend on the pre-intervention predictors chosen by the researchers that may influence results (Huntington-Klein et al., 2025) or encourage specification searching (Ferman et al., 2020; Kaul et al., 2022). As such, in Figure 6 we test the

¹³The smallest attainable p -value is roughly $1/22 = 0.045$.



Panel B: Training/Validation Period Analysis

Figure 6: Synthetic Control Sensitivity Analyses: Officer Assault Rate

Notes: Left panel: NYPD actual (solid navy) vs. synthetic NYPD (dashed red). Right panel: gap (actual minus synthetic). Weights estimated via Abadie et al. (2010). Vertical dashed line: 1994 Compstat reform.

sensitivity of this approach using alternative specifications.

In Panel A in Figure 6, we review the weights each police agency receive in constructing the NYPD counterfactual and remove the city that receives the heaviest weight. Specifically, we remove Los Angeles when estimating the impact of reform on total assaults and omit Seattle when examining the effects of the reform on injury assaults. For convenience, we also include the counterfactual using the entire sample. For total assaults, the new counterfactual appears to mirror assault rates of the NYPD prior to the reform effectively. Post-reform, the estimated effect is larger when the largest donor police department is omitted. For injury assaults, the post reform effects are less pronounced than estimated using the entire sample.

However, this alternative sample does not accurately mirror the NYPD injury assault rates prior to intervention. Therefore, the attenuated post-reform effects are more reflective of poor comparability than sensitivity of the post-reform estimates.

In Panel B, we follow Cavallo et al. (2013) and Galiani and Quistorff (2017) and construct a synthetic counterfactual using only the first years of the pre-treatment period. Here, predictors are constructed from 1988–1990 only (the training window), with 1991–1993 held out as an out-of-sample validation window. If the synthetic control built on 1988–1990 data alone tracks NYPD well in 1991–1993 without having seen those years, the baseline specification’s predictor set is not merely over-fitting the pre-reform period.

The estimates for injury-related assaults using the shortened training/validation window are similar to those from the full pre-intervention period, and the interpretation of a reduction in injury assaults is unchanged. For total assaults, by contrast, the training-only synthetic diverges from the actual NYPD series within the 1991–1993 validation window, indicating that a three-year predictor window is too short to reconstruct NYPD’s trajectory. We emphasize that these are sensitivity checks on the synthetic control, which is itself a secondary, corroborative method rather than our primary basis for inference. A poor fit under a deliberately limited predictor set does not weaken or contradict the difference-in-differences results. If anything, it reinforces why we treat the TWFE estimates and their randomization inference as the primary evidence and the synthetic control as corroborative and supportive of our main results that officer safety improved as a result of the Compstat reform.

7 Discussion and Conclusion

Policies that introduce additional oversight and accountability are often viewed unfavorably. The 1994 reform centered on Compstat implementation in the NYPD is no exception, and the principal-agent framework offers a productive lens for understanding how accountability reforms in law enforcement can impact workplace safety. Compstat was explicitly designed to

reduce crime; officer safety was never mentioned as an objective. Yet the reform’s mechanism of increasing performance transparency, activating career-concern incentives, and inducing commanders to optimize deployment incidentally operated on dimensions that also govern officer-citizen confrontation risk and improved officer safety. These findings have direct relevance to current debates over body-worn cameras, real-time crime analytics platforms, and other technologies that increase the observability of frontline police behavior. Our results suggest that the performance-visibility channel may also protect officers by inducing more cautious and coordinated deployment decisions.

Two limitations bound the interpretation of these results. First, the 1994 reform did not occur in isolation: it coincided with a change in city and departmental leadership, a broken-windows enforcement strategy, an expansion of stop-question-and-frisk activity, and federally funded hiring. Our design cannot fully separate the contribution of Compstat’s information-and-accountability system from this broader transformation, and we accordingly interpret our estimate as the effect of the 1994 NYPD reform *centered on* Compstat rather than of Compstat as an isolated technology. Reassuringly, the most plausible concurrent confounders - intensified enforcement and street contact, would, if anything, raise officer assault risk, working against the decline we estimate. Second, the 1995 merger of the Transit and Housing Police permanently changed the NYPD’s institutional boundary; the donut and per-population specifications address the transitional artifact and the mechanical denominator change but not the enduring shift in force composition, and the LEOKA data do not permit reconstruction of a consistent pre-merger combined series. We regard the post-1996 estimates, which reflect a stable merged department, as the most reliable basis for the long-run effect.

A related concern is that the decline in officer assaults may reflect the broad reduction in New York City crime over this period, and hence fewer risky encounters rather than the accountability and deployment channels. However, much of NYC’s crime decline is itself widely attributed to Compstat adoption (Eck and Maguire, 2000; Zimring, 2012). Thus, a

crime-reduction channel would represent a second-order pathway through which the same reform improved officer safety, rather than a competing explanation, and it may help account for why the estimated effect strengthens over time. Moreover, the reduction in officer assaults appears immediately in 1994 (Figure 3) suggesting that broader crime trends cannot account for the entire effect.¹⁴

While the Compstat reform is credited with reduced crime (Eck and Maguire, 2000; Zimring, 2012), the increased emphasis on metrics also increased pressure to perform (Lu et al., 2021) and manipulate crime statistics (Eterno and Silverman, 2010; Eckhouse, 2022). This paper explores an alternative unintended consequence that aligned incentives reduced information asymmetries improved officer safety. We find that New York City’s 1994 managerial and policing reform centered on Compstat substantially reduced the rate at which NYPD officers were assaulted. The reduction is strongest for injury-involved assaults that are less susceptible to discretionary reporting, concentrated in team-patrol encounters and evening hours, distributed broadly across all weapon types, and not moderated by departmental gender composition or workforce racial demographics.

Our results highlight the importance of organizational structure in determining occupational safety outcomes in high-risk public sector jobs. The 30 percent reduction in officer assaults (larger still for injury-involved assaults) achieved through a management reform that reallocated attention and accountability rather than adding equipment or changing use-of-force policy, suggests that deployment optimization and accountability culture are first-order determinants of frontline worker safety, at least as important as equipment investments such as body armor or weapon policy. For police administrators, these results argue that investments in real-time crime analytics and commander accountability should be evaluated not only for their effects on public safety, but also as instruments for reducing the human costs borne by officers in the line of duty.

¹⁴We do not condition on contemporaneous crime rates, which are themselves an outcome of the reform and would constitute a “bad control” (Cinelli et al., 2024).

Statements and Declarations

Competing interests. The authors have no relevant financial or non-financial interests to disclose.

Funding. The authors received no specific funding for this work.

Data availability. The officer-assault and employment data are from the FBI's Law Enforcement Officers Killed and Assaulted (LEOKA) program; the workforce-composition data are from the Law Enforcement Management and Administrative Statistics (LEMAS) surveys (ICPSR 6708 and 2700). All data are publicly available. Replication code is available from the authors upon request.

References

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller**, “Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program,” *Journal of the American Statistical Association*, 2010, *105* (490), 493–505.
- , —, and —, “Comparative politics and the synthetic control method,” *American Journal of Political Science*, 2015, *59* (2), 495–510.
- Adamson, Jordan and Lucas Rentschler**, “Criminal justice from a public choice perspective: an introduction to the special issue,” *Public Choice*, 2023, *196* (3–4), 223–227.
- Bernstein, Ethan and Shelley Xin Li**, “The Performance Effects of Giving Front-Line Employees Direct Access to Performance Data and Thereby Limiting the Supervisor’s Feedback-Intermediation Role: Evidence from a Field Experiment,” *Management Science*, 2025, *72* (2), 783–1726.
- Bratton, William and Peter Knobler**, *Turnaround: How America’s Top Cop Reversed the Crime Epidemic*, New York: Random House, 1998.
- Buchmueller, Thomas C, John DiNardo, and Robert G Valletta**, “The Effect of an Employer Health Insurance Mandate on Health Insurance Coverage and the Demand for Labor: Evidence from Hawaii,” *American Economic Journal: Economic Policy*, 2011, *3* (4), 25–51.
- Burgess, Simon, Carol Propper, Marisa Ratto, and Emma Tominey**, “Incentives in the Public Sector: Evidence from a Government Agency,” *The Economic Journal*, 2017, *127* (605), F117–F141.
- Campbell, Travis**, “Black Lives Matter’s effect on police lethal use of force,” *Journal of Urban Economics*, 2024, *141*, 103587.
- Card, David and Stefano DellaVigna**, “What Do Editors Maximize? Evidence from Four Economics Journals,” *Review of Economics and Statistics*, 2020, *102* (1), 195–217.
- Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano**, “Catastrophic natural disasters and economic growth,” *Review of Economics and Statistics*, 2013, *95* (5), 1549–1561.
- Cinelli, Carlos, Andrew Forney, and Judea Pearl**, “A Crash Course in Good and Bad Controls,” *Sociological Methods & Research*, 2024, *53* (3), 1071–1104.

- Cohen, Lawrence E. and Marcus Felson**, “Social Change and Crime Rate Trends: A Routine Activity Approach,” *American Sociological Review*, 1979, 44 (4), 588–608.
- Cunningham, Scott and Manisha Shah**, “Decriminalizing Indoor Prostitution: Implications for Sexual Violence and Public Health,” *Review of Economic Studies*, 2017, 85 (3), 1683–1715.
- Dixit, Avinash**, “Incentives and organizations in the public sector: An interpretative review,” *Journal of Human Resources*, 2002, 37 (4), 696–727.
- Eck, John E and Edward R Maguire**, “Have changes in policing reduced violent crime? An assessment of the evidence,” in Alfred Blumstein and Joel Wallman, eds., *The Crime Drop in America*, New York: Cambridge University Press, 2000, pp. 207–265.
- Eckhouse, Laurel**, “Metrics Management and Bureaucratic Accountability: Evidence from Policing,” *American Journal of Political Science*, 2022, 66 (2), 385–401.
- Eterno, John A and Eli B Silverman**, “The NYPD’s Compstat: Compare Statistics or Compose Statistics?,” *International Journal of Police Science & Management*, 2010, 12 (3), 426–449.
- Federal Bureau of Investigation**, “Law Enforcement Officers Killed and Assaulted, 2022,” Technical Report, U.S. Department of Justice, Federal Bureau of Investigation, Washington, DC 2022. Criminal Justice Information Services Division.
- , “Law Enforcement Officers Killed and Assaulted (LEOKA), Uniform Crime Reporting Program,” U.S. Department of Justice. Data distributed via the FBI Crime Data Explorer and ICPSR 2023. Annual agency-level data accessed for 1988–1999.
- Fegley, Tate and Ilia Murtazashvili**, “From defunding to refunding police: institutions and the persistence of policing budgets,” *Public Choice*, 2023, 196 (1–2), 123–140.
- Ferman, Bruno and Cristine Pinto**, “Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity,” *The Review of Economics and Statistics*, 2019, 101 (3), 452–467.
- , —, and **Vitor Possebom**, “Cherry picking with synthetic controls,” *Journal of Policy Analysis and Management*, 2020, 39 (2), 510–532.
- Ferrandino, Joseph**, “The Efficiency of Frisks in the NYPD, 2004–2010,” *Criminal Justice Review*, 2012, 38 (2), 149–168.

- Fisher, Ronald A**, *The Design of Experiments*, Edinburgh: Oliver and Boyd, 1935.
- Galiani, Sebastian and Brian Quistorff**, “The synth_runner package: Utilities to automate synthetic control estimation using synth,” *Stata Journal*, 2017, 17 (4), 834–849.
- Garner, Joel, Thomas Schade, John Hepburn, and John Buchanan**, “Understanding the use of force by and against the police,” *NIJ Research in Brief*, 1995.
- Guerin, RJ, AA Reichard, S Derk, KJ Hendricks, LM Menger-Ogle, and AH Okun**, “Nonfatal Occupational Injuries to Younger Workers — United States, 2012–2018,” *Morbidity and Mortality Weekly Report*, 2020, 69 (35), 1204–1209.
- Holmström, Bengt**, “Moral hazard and observability,” *Bell Journal of Economics*, 1979, 10 (1), 74–91.
- , “Managerial incentive problems: A dynamic perspective,” *Review of Economic Studies*, 1999, 66 (1), 169–182.
- **and Paul Milgrom**, “Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design,” *Journal of Law, Economics, & Organization*, 1991, 7, 24–52.
- Huntington-Klein, Nick, Claus C Portner, Ian McCarthy, and The Many Economists Collaborative on Researcher Variation**, “The Sources of Researcher Variation in Economics,” *NBER Working Paper 33729*, 2025.
- Johnson, Josiah and Rhet A Smith**, “Main street business initiatives and crime in small towns,” *Journal of Economic Behavior and Organization*, 2023, 209, 91–112.
- Johnson, Matthew S, Daniel Schwab, and Patrick Koval**, “Legal Protection Against Retaliatory Firing Improves Workplace Safety,” *Review of Economics and Statistics*, 2024, 106 (5), 1236–1253.
- Kaul, Ashok, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler**, “Standard synthetic control methods: The case of using all preintervention outcomes together with covariates,” *Journal of Business & Economic Statistics*, 2022, 40 (3), 1362–1376.
- Kelling, George L. and Catherine M. Coles**, *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities*, New York: Simon and Schuster, 1996.
- Kim, Taeho**, “Facilitating Police Reform: Body Cameras, Police-Involved Homicides, and Law Enforcement Outcomes,” *Journal of Public Economics*, 2025, 248, 105424.

- Leigh, J. Paul**, “Economic Burden of Occupational Injury and Illness in the United States,” *The Milbank Quarterly*, 2011, 89 (4), 728–772.
- Lu, Yi, Kaifeng Yang, and Blair M Thomas**, “Designing Performance Systems in Anticipation of Unintended Consequences: Experiences and Lessons From the Compstat-Based Performance Regime in NYPD,” *Administration and Society*, 2021, 53 (6), 907–936.
- Lum, Cynthia, Christopher S Koper, David B Wilson, Margaret Stoltz, Michael Goodier, Elizabeth Eggins, Angela Higginson, and Lorraine Mazerolle**, “Body-worn cameras’ effects on police officers and citizen behavior: A systematic review,” *Campbell Systematic Reviews*, 2020, 16 (3), e1112.
- Maltz, Micheal D and Joseph Targonski**, “A Note on the Use of County-Level UCR Data,” *Journal of Quantitative Criminology*, 2002, 18 (3), 297–318.
- Martins, Pedro S. and João R. Ferreira**, “Effects of individual incentive reforms in the public sector: the case of teachers,” *Public Choice*, 2025, 204 (3–4), 457–481.
- Meares, Tracey L**, “The Law and Social Science of Stop and Frisk,” *Annual Review of Law and Social Science*, 2014, 10, 335–352.
- Mookherjee, Dilip**, “Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective,” *Journal of Economic Literature*, 2006, 44 (2), 367–390.
- Romer, David**, “In Praise of Confidence Intervals,” *AEA Papers and Proceedings*, 2020, 110, 55–60.
- Smith, Douglas A**, “The neighborhood context of police behavior,” *Crime and Justice*, 1986, 8, 313–341.
- Tiesman, Hope M., Melody Gwilliam, Srinivas Konda, Jeff Rojek, and Shelli Marsh**, “Nonfatal Injuries to Law Enforcement Officers: A Rise in Assaults,” *American Journal of Preventive Medicine*, 2018, 54 (4), 503–509.
- U.S. Bureau of Labor Statistics**, “Survey of Occupational Injuries and Illnesses,” U.S. Department of Labor 2000. Nonfatal occupational injury and illness incidence rate of 6.3 per 100 full-time-equivalent workers, 1999.
- Weisburd, David, Stephen D Mastrofski, Ann Marie McNally, Rosann Greenspan, and James J Willis**, “Reforming to preserve: Compstat and strategic problem solving in American policing,” *Criminology & Public Policy*, 2003, 2 (3), 421–456.

Willis, James J, Stephen D Mastrofski, and David Weisburd, “Compstat in practice: An in-depth analysis of three cities,” *Police Foundation Reports*, 2003.

Wolfers, Justin, “Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results,” *American Economic Review*, 2006, *96* (5), 1802–1820.

Zimring, Franklin E, *The City That Became Safe: New York’s Lessons for Urban Crime and Its Control*, New York: Oxford University Press, 2012.

A Appendix

A.1 Felony Killing Rate: Event Study

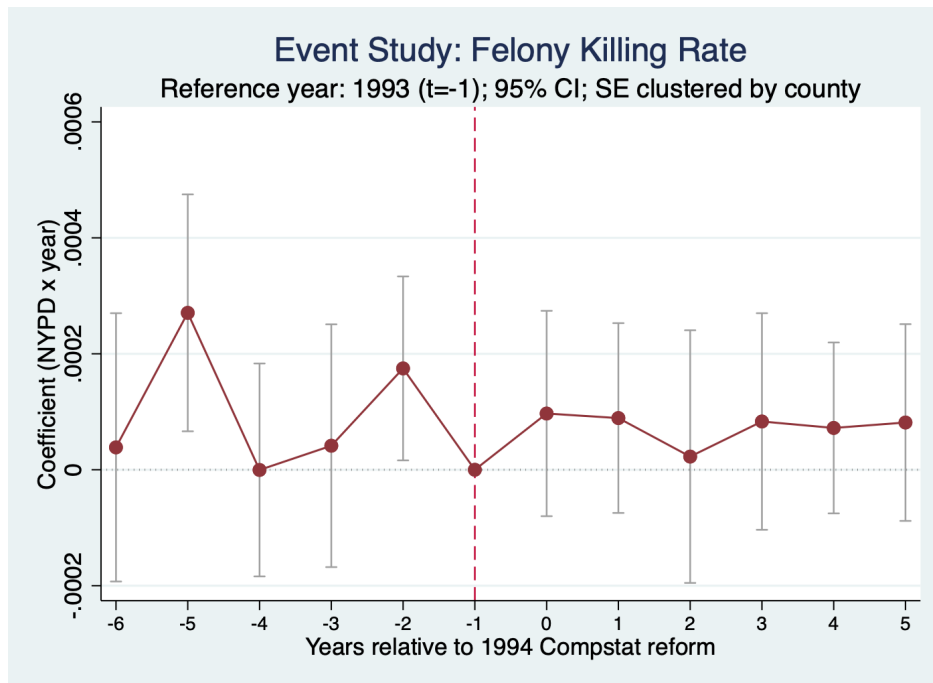


Figure A1: Event Study: Felony Killing Rate (Descriptive)

Notes: Event-study estimates for the felony killing rate. Near-zero baseline produces wide confidence intervals. Reference year: 1993. 95% CI with county-clustered SE. Treated as descriptive only.

B NYPD 1995 Data: Administrative Gap and Imputation Rationale

The April 1995 merger of the NYC Transit Police and Housing Police into the NYPD created two simultaneous anomalies in the LEOKA data: (i) a reporting disruption that caused the NYPD to file zero assault records for May–December 1995, and (ii) a mechanical jump in the officer headcount as Transit and Housing officers were absorbed into the main department. Neither anomaly reflects genuine changes in assault risk. This appendix presents the raw

evidence for both disruptions and explains the imputation procedure used in the baseline analysis.

Figure A2 displays monthly data for 1992–1996 at the level of individual months. Panel A shows the monthly assault count: NYPD reports approximately 300–500 assaults per month through April 1995, then abruptly reports *zero* for May through December 1995 before resuming normal reporting in 1996. No plausible behavioral mechanism could produce eight consecutive months of zero assaults followed by an immediate return to prior levels. This pattern is an unmistakable reporting artifact. Panel B shows sworn officer counts: the NYPD denominator rises sharply in mid-1995 as approximately 7,000 Transit and Housing Police officers were administratively transferred. Panel C plots the resulting monthly assault rate (per 100 officer-years): the combination of a zero numerator and an inflated denominator produces the spurious trough visible in 1995. The control group average (dashed gray) shows no corresponding disruption, confirming the anomaly is NYPD-specific.

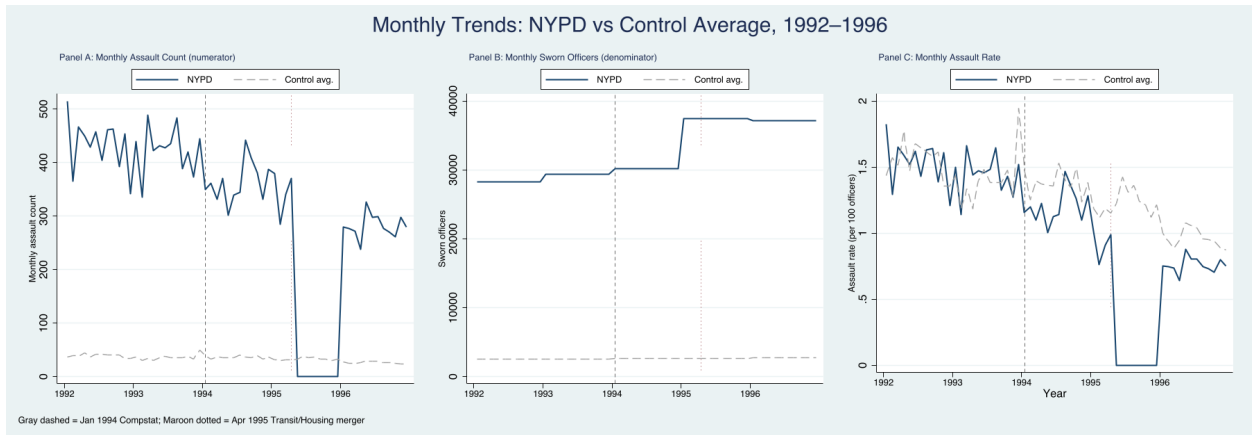


Figure A2: NYPD 1995 Administrative Gap: Monthly Diagnostic (1992–1996)

Notes: Monthly LEOKA data, NYPD vs. unweighted control average (31 agencies). Panel A: monthly assault count (numerator). Panel B: monthly sworn officers (denominator). Panel C: monthly assault rate ($= \text{count/officers} \times 12 \times 100$, annualized). Gray dashed vertical: January 1994 Compstat reform. Maroon dotted vertical: April 1995 Transit/Housing merger. NYPD assault count drops to exactly zero for May–December 1995 (administrative reporting disruption); officer headcount jumps as Transit and Housing officers are absorbed.

Figure A3 presents the same decomposition at annual resolution for the full 1988–1999 window, using the *raw, unimputed* LEOKA data. Panel A shows the assault count numerator: NYPD’s count drops sharply to approximately 1,377 in 1995 (the Jan–Apr partial-year total), before recovering to over 3,700 in 1996. The control group mean is flat across this window, confirming the drop is NYPD-specific. Panel B shows the officer denominator: NYPD’s headcount jumps from approximately 29,000 to 37,450 in 1995, again isolated to NYPD. Panel C shows the resulting raw assault rate: the combination produces the artificial trough to approximately 3.7 per 100 officer-years in 1995, far below any plausible actual level and inconsistent with the smooth post-reform trajectory visible in both adjacent years. This figure documents why the raw 1995 data point cannot be used as-is in the analysis.

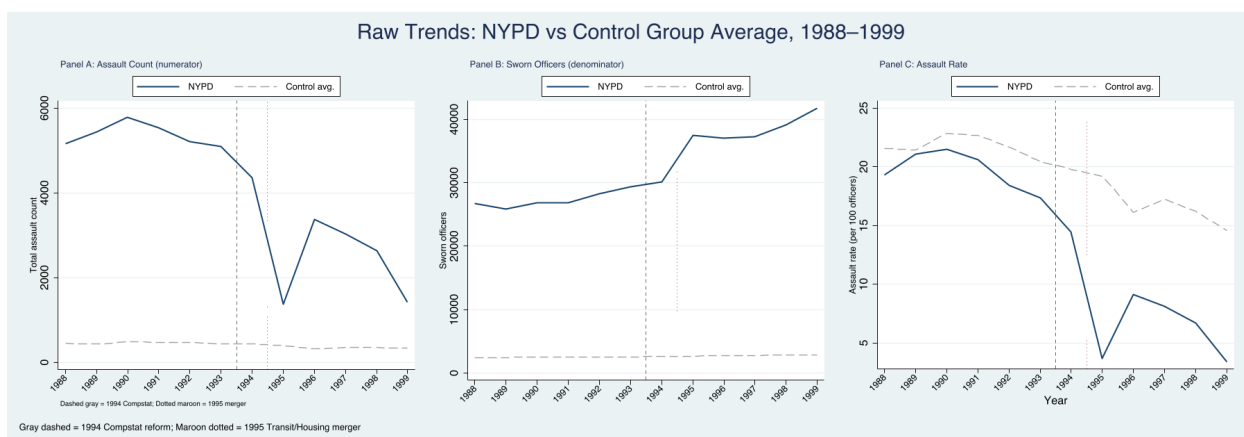


Figure A3: Raw Annual Trends: NYPD vs. Control Average, 1988–1999 (Unimputed)

Notes: Annual LEOKA data, unimputed, NYPD (solid navy) vs. unweighted control average (31 agencies, dashed gray). Panel A: total assault count. Panel B: sworn officer headcount. Panel C: assault rate per 100 officer-years. Gray dashed vertical: 1994 Compstat reform. Maroon dotted vertical: 1995 Transit/Housing merger. The artificial 1995 trough in Panel C reflects both a near-zero numerator (reporting gap) and an inflated denominator (merger absorption), neither of which reflect genuine changes in officer assault risk. The baseline analysis replaces the 1995 assault count with an imputed value of 4,119 based on NYPD’s historical January–April seasonal share (see Appendix C).

Imputation procedure. The baseline analysis replaces the raw 1995 NYPD assault count (1,377, Jan–Apr only) with an imputed full-year total of 4,119. The imputation uses NYPD’s own 1992–1994 average January–April seasonal share: assaults in Jan–Apr constituted 33.4% of the annual total in those three years. Scaling up: $1,377/0.334 = 4,119$. This procedure assumes the 1995 seasonal pattern was similar to the prior three years, a conservative assumption given that assault rates were declining smoothly through 1994. The officer denominator retains the post-merger headcount of 37,450, yielding an imputed 1995 assault rate of $4,119/37,450 \times 100 = 11.0$ per 100 officer-years. This is consistent with the smooth post-reform trajectory: the unimputed 1994 rate is approximately 15.9 and the 1996 rate is approximately 9.0, making 11.0 for 1995 geometrically plausible. Appendix C below demonstrates that all main results are robust to dropping 1995 entirely.

Two additional denominator notes. First, LEOKA records officer headcount as a single annual figure replicated identically across all 12 months of that year, rather than as a true monthly count. Consequently, any year-over-year change in headcount appears as a discrete step at January 1 in the monthly series. The 1993–1994 transition shows such a step: the NYPD headcount rises from 29,327 to 30,135, an increase of 808 officers (+2.8%). This increment is actually *smaller* than the preceding two years (1991–1992: +1,393, +5.2%; 1992–1993: +1,078, +3.8%), placing it within the normal range of annual staffing growth rather than indicating a structural reporting change or hiring surge. The large discrete jump occurs at the 1994–1995 transition (+7,315 officers, +24.3%), which reflects the Transit/Housing merger.

Second, there is a denominator inconsistency embedded in the January–April 1995 period. The 1995 LEOKA annual headcount (37,450) reflects end-of-year staffing after the April merger, and LEOKA applies this figure to all twelve months of 1995, including January through April when the Transit and Housing Police had not yet been administratively absorbed and were filing their own separate LEOKA reports. As a result, the January–April 1995 assault rate uses a denominator approximately 24% larger than the officer pool

actually covered by the assault numerator. Scaling the pre-merger headcount (30,135) into the imputation instead yields an estimated 1995 assault rate of approximately 13.7 per 100 officer-years, closer to the 1994 rate of 14.4 and implying a more modest 1995 effect than the 11.0 figure in the baseline. This denominator inflation slightly overstates the event-study coefficient at $t = +1$ (1995). Readers should therefore weight the post-1995 evidence more heavily: from 1996 onward, both the numerator and denominator consistently reflect the fully merged NYPD, and the event-study path from $t = +2$ through $t = +5$ is free of this inconsistency. The donut specification in Appendix C provides the cleanest test precisely because it bypasses 1995 entirely and begins the post-reform comparison from 1996.

C Robustness: Dropping 1995

The 1995 NYPD LEOKA data suffer from an administrative reporting gap caused by the April 1995 merger of the NYC Transit and Housing Police into the NYPD. Monthly LEOKA records confirm that NYPD filed complete monthly assault reports only for January through April 1995; May through December 1995 reports zero assaults, reflecting a disruption to the reporting pipeline during the administrative integration rather than a genuine cessation of officer assaults. Our baseline analysis imputes the full 1995 annual assault count by scaling the observed January–April total (1,377 assaults) upward using NYPD’s own 1992–1994 average January–April seasonal share (33.4%), yielding an imputed annual total of 4,119 assaults for 1995.

Although we prefer to retain 1995 with this principled imputation, which recovers the counterfactual assault level that NYPD would have reported absent the administrative disruption, we report here a robustness check that drops 1995 entirely from both the TWFE event study and the synthetic control analysis. This provides a clean sensitivity test: if the main results are sensitive to the 1995 imputation, they should change when 1995 is removed.

Figure A4 presents the event-study estimates with 1995 excluded. Pre-reform coefficients

($k = -6$ to $k = -2$) remain flat and statistically indistinguishable from zero. Post-reform coefficients turn negative beginning in 1994 (now $k = 0$) and grow steadily through 1999, with the gap at $k = +2$ (1996) through $k = +5$ (1999) mirroring the pattern in the baseline Figure 3. Note that in this specification, $k = +1$ corresponds to 1996 (not 1995), so the event-study path skips the transition year entirely.

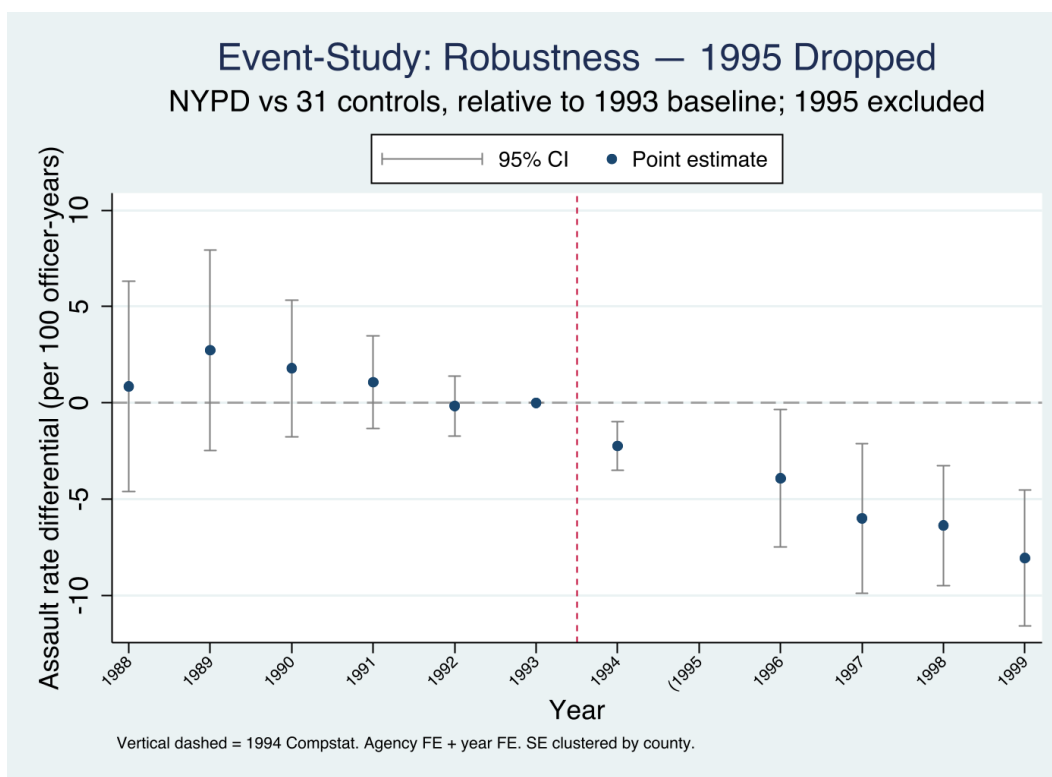


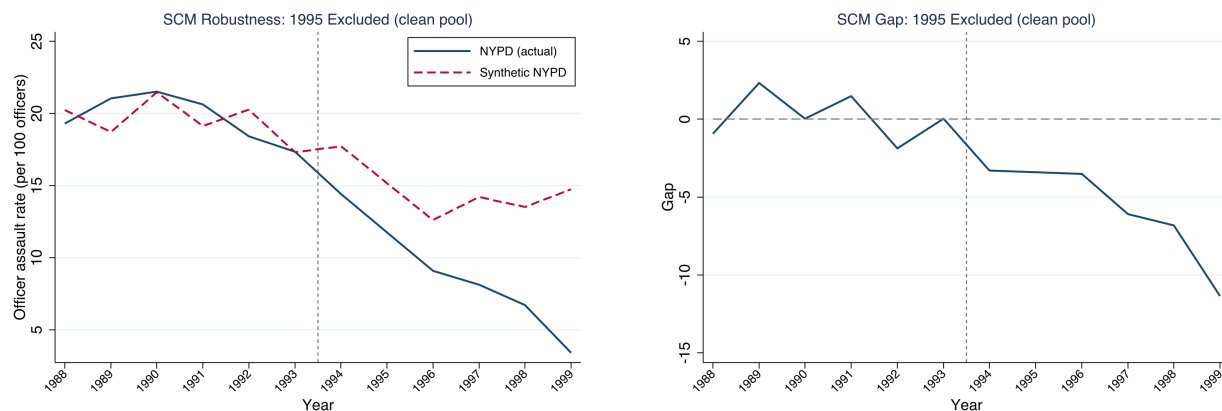
Figure A4: Event Study: 1995 Dropped

Notes: TWFE event-study estimates with 1995 excluded from the panel. Reference year: 1993.

Years on the x-axis are labeled as the actual calendar year; $k = +1$ corresponds to 1996. Agency FE + year FE. SE clustered by county. 95% CI.

The synthetic control results are reported in Figure A5. The synthetic NYPD tracks the actual series closely in the 1988–1993 pre-reform period and diverges after 1994. The average post-1994 gap (excluding 1995) grows from 1996 through 1999. The MSPE ratio for NYPD under this specification is 23.58, and the Fisher permutation p -value is 0.238 (NYPD ranks 5th out of 21 agencies in the corrected donor pool). These results are directionally consistent

with the main specification: the Compstat effect is not an artifact of the 1995 imputed data point.



(a) Actual vs. Synthetic (1995 excluded)

(b) Gap series (1995 excluded)

Figure A5: Synthetic Control: 1995 Dropped

Notes: Synthetic control estimated on 1988–1993 pre-reform period; post-period = 1996–1999. 1995 is entirely excluded. Vertical dashed line: 1994 Compstat reform. Permutation p-value = 0.238 (NYPD MSPE ratio = 23.58, ranks 5th of 21).

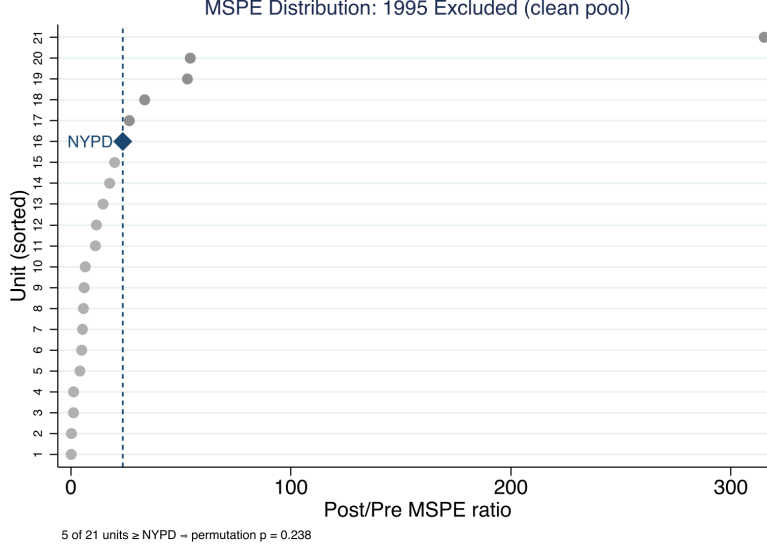


Figure A6: Permutation Inference: 1995 Dropped

Notes: Post/pre MSPE ratio distribution for the 21 agencies in the corrected donor pool, with 1995 excluded. Navy diamond = NYPD (ratio = 23.58, 5th highest). Permutation p -value = 0.238.

D Officer Fatalities: Binary Outcome Analysis

Officer fatality outcomes (felony killing, accidental killing) have near-zero pre-reform baselines, making the linear rate specification uninformative. We instead construct binary indicators equal to one if any officer in the agency was killed in a given year and estimate a logit model with agency and year fixed effects. The continuous felony-kill rate is excluded from Table 2; the positive accidental-kill rate coefficient in earlier specifications is a denominator artifact of the 1995 Transit/Housing merger, which sharply increased NYPD's sworn officer count without a commensurate change in accidental deaths.

Table A1: Appendix Table A2: Effect on Officer Fatalities (Logit)

	(1)	(2)
	Felony Killing	Any Killing
	(0/1 indicator)	(0/1 indicator)
NYPD $\times$ Post-1994	−0.060	+0.081
	(0.061)	(0.056)
Agency FE	Yes	Yes
Year FE	Yes	Yes
Observations	384	384

Logit with agency and year fixed effects. Average marginal effects reported. SE clustered by agency level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Col. (1): felony killing indicator. AME = −0.060 (SE = 0.061, $p = 0.318$). Col. (2): any killing indicator (felony + accidental). AME = +0.081 (SE = 0.056, $p = 0.150$). Both effects are statistically indistinguishable from zero. The positive point estimate in Col. (2) reflects the additional accidental-killing component and is consistent with the 1995 Transit/Housing merger increasing the officer count denominator without proportionally reducing accidental deaths.

E Donor Weights: Synthetic NYPD

Table A2: Synthetic Control Donor Weights

City	State	Weight
<i>Panel A: Total assault rate</i>		
Los Angeles	CA	0.502
Philadelphia	PA	0.406
Baltimore	MD	0.052
Indianapolis	IN	0.021
El Paso	TX	0.018
<i>Panel B: Injury assault rate</i>		
Seattle	WA	0.610
Austin	TX	0.263
Charlotte	NC	0.127

Weights via nested optimization (`synth`, Abadie et al. 2010) over the corrected 21-city donor pool. Pre-treatment period: 1988–1993. Cities with weight < 0.005 omitted. Panel A weights the total-assault synthetic; Panel B weights the injury-assault synthetic.

F Per-Population Assault Rate: Event Study

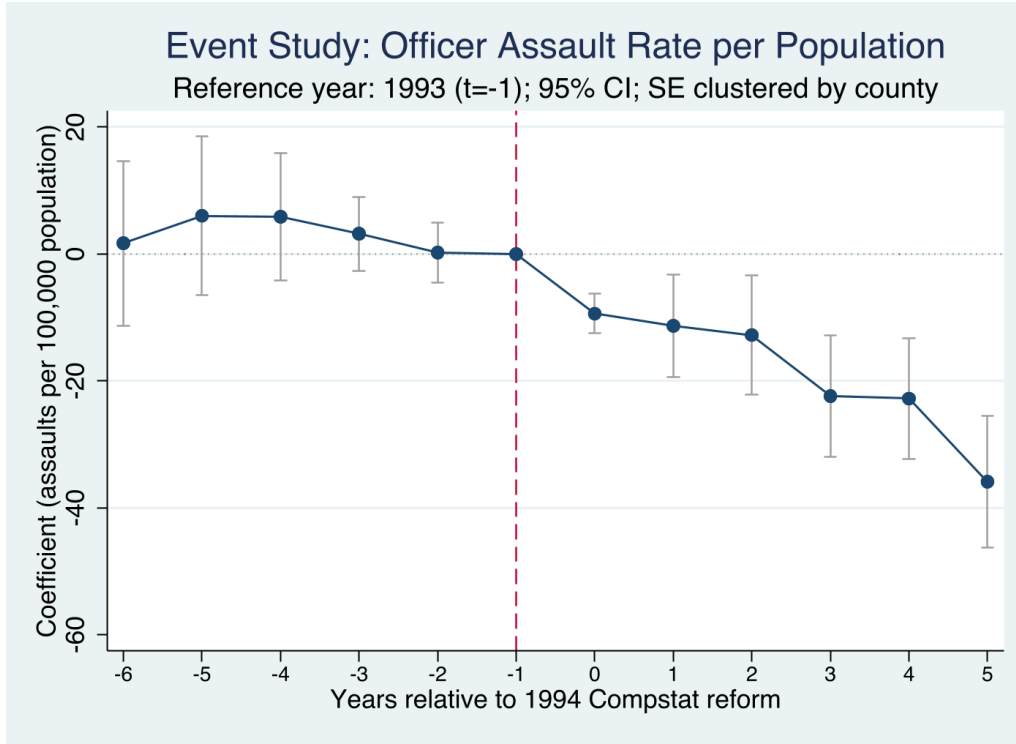


Figure A7: Event Study: Officer Assault Rate per 100,000 Population

Notes: Event-study coefficients using assaults per 100,000 served population as the outcome (population denominator unaffected by the 1995 Transit/Housing merger). Reference year: 1993 ($k = -1$, normalized to zero). Vertical bars: 95% CI with county-clustered SE. Pre-reform coefficients are flat; post-1994 coefficients turn sharply negative, mirroring Figure 3.

G Placebo-in-Time: Synthetic Control

A concern with any synthetic control application is that the estimated post-treatment gap could reflect a spurious trend rather than a causal effect. We address this directly with a placebo-in-time test. Restricting the sample to the pre-reform period (1988–1993), we assign a false treatment date of 1991 and re-run the full synthetic control and permutation inference procedure on this pre-treatment window only. If NYPD’s assault rate exhibits no systematic break at the false treatment date, the permutation p-value at 1991 should be large.

The results support causal identification. Under the false 1991 treatment, the permutation p -value is 0.400 (NYPD's post/pre MSPE ratio ranks 8th of 20 units in the pre-reform sample), indicating that NYPD is unremarkable among the donor pool when the wrong treatment date is imposed. This contrasts with the true 1994 treatment, which yields a permutation p -value of 0.238 (5th of 21). The more extreme result at the true date is consistent with 1994 as the genuine source of the break, and with Compstat as the causal mechanism (Appendix Figure A8). The very short three-year training window (1988–1990) inflates the raw MSPE ratio *levels*, so inference relies on the permutation ranks rather than the ratio magnitudes.

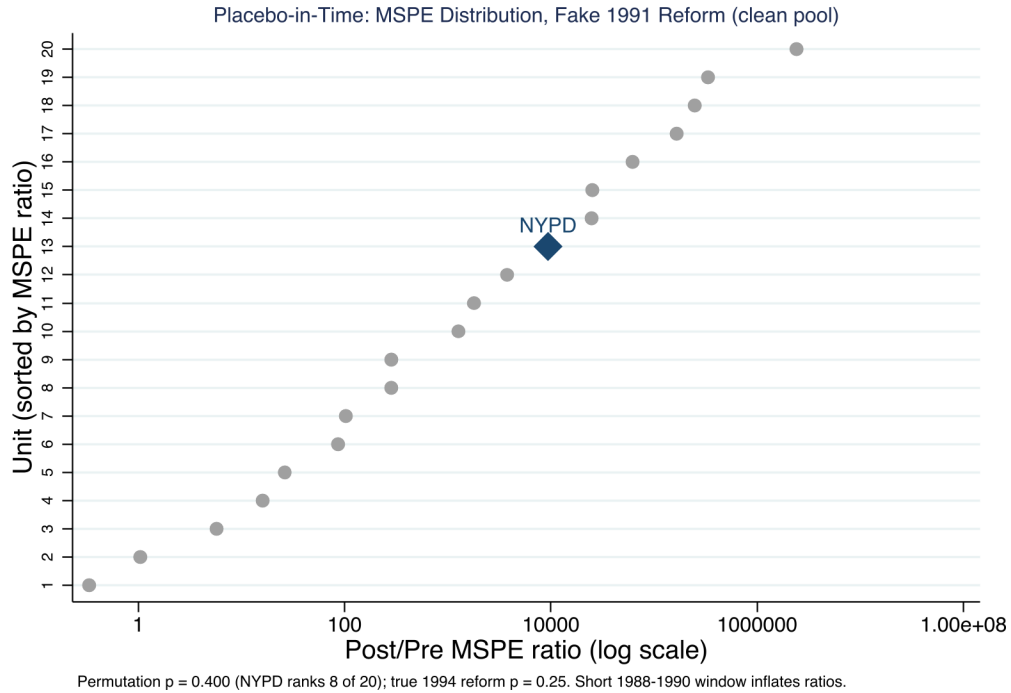


Figure A8: Placebo-in-Time: MSPE Ratio Distribution (Fake Treatment 1991)

Notes: Post/pre MSPE ratios (log scale) for all units under the false 1991 treatment date. Navy diamond = NYPD, which ranks 8th of 20 (permutation p -value = 0.400), indicating no spurious pre-trend; the short 1988–1990 training window inflates ratio levels, so we rely on ranks. For comparison, the true 1994 treatment yields a permutation p -value of 0.238.