Police Officer Identification, Misconduct, and Public Trust

Thomas Simon
University of Mannheim
thomas.simon@uni-mannheim.de

Felix W. Vetter
MIT Sloan School of Management
fwvetter@mit.edu

Gerrit von Zedlitz
University of Mannheim
gerrit.zedlitz-neukirch@uni-mannheim.de

January 2026

Abstract

Police misconduct is difficult to deter when individual officers cannot be identified. We study a reform that introduced identification tags for police officers operating in riot gear, where individual attribution is limited. We estimate the reform's effects on excessive use of force and civilian crime using administrative data and exploiting the staggered implementation across German states. We find that excessive use of force declines by roughly two thirds following the reform, with no evidence of an increase in civilian crime. Survey evidence further indicates higher trust in the police, particularly among groups more likely to be exposed to riot-gear deployments.

We thank David Abrams, Jannis Bischof, Lara Berger (Discussant), Patricia Breuer (Discussant), Dirk Simons, Konrad Stahl, Andrew Sutherland, and seminar participants at the University of Mannheim, the EAA Doctoral Colloquium, the Mannheim Alumni Homecoming Conference, and the University of Neuchâtel for valuable comments and suggestions. We further thank the German Federal Criminal Police Office and the German State Criminal Police Offices for providing data, and GESIS as well as FragDenStaat for support with our data requests. Financial support from the authors' respective institutions, the Deutsche Forschungsgemeinschaft (Project-ID 403041268 – TRR 266), the University of Mannheim's Graduate School of Economic and Social Sciences, and the Head Family Fund for research in Economics, Finance, and Accounting at MIT Sloan is gratefully acknowledged. All errors are our own. Conflicts of interest: none.

Introduction

Police violence remains a persistent concern despite formal legal constraints (Grosjean et al. 2023; Holz et al. 2023). A growing body of evidence shows that police misconduct, including the excessive use of force, imposes substantial social costs, erodes trust in the police, and undermines the functioning of public institutions (Bor et al. 2018; Weitzer and Tuch 2004; Ang 2021; Ang et al. 2025).

Policymakers frequently call for stronger police oversight aimed at increasing accountability. As in standard models of deterrence, the effectiveness of such oversight depends on the likelihood that misconduct is detected and sanctioned. In practice, accountability hinges on the ability to identify individual officers responsible for alleged misconduct. When officers cannot be identified, oversight loses bite and misconduct becomes difficult to deter.

Although complaints are frequently filed, most allegations of police misconduct never reach court. On of the main obstacle is the difficulty of identifying individual officers during public prosecutions (Amnesty 2006; UN 2020). Misconduct often occurs in dynamic and chaotic situations in which interactions unfold rapidly (Dube et al. 2024), making it difficult for claimants and witnesses to identify individual officers who lack clearly distinguishable identification characteristics at the time of the incident. When individual officers cannot be identified by claimants or witnesses, prosecutors must rely more heavily on internal police investigations to establish responsibility. These investigations are widely criticized for shielding officers from accountability through a culture of silence within police forces (Chin et al. 1997; Ravenell 2021).

Given the institutional and political challenges of reforming internal police oversight, policymakers and civil society organizations have increasingly advocated mechanisms that enable

external identification and attribution of responsibility, aimed at strengthening accountability. One prominent approach is to require police officers to wear clearly visible identification tags (Amnesty 2006; UN 2020; ACLU 2021). Such reforms have recently gained traction internationally, reflecting growing concern over opaque policing practices.[1]

This paper studies a police reform adopted by several German states that requires police officers to wear visible identification tags when operating in riot gear. Riot gear is used in closed-unit deployments and obscures individual identity through helmets, face coverings, and uniform appearance, making it difficult to attribute responsibility for police misconduct. Prior to the reform, officers in such deployments could be identified only at the unit level. After implementation, officers wear a five-digit alphanumeric code that allows attribution at the individual level while preserving anonymity outside official investigations. The reform aims to strengthen accountability by facilitating external identification and reducing reliance on internal police investigations (Fung et al. 2007).

The reform has generated opposing predictions. Identification requirements may deter misconduct by increasing accountability, but they may also discourage proactive policing. If officers respond to heightened scrutiny by reducing engagement or avoiding intervention, this could lead to increased civilian crime. Critics, including police unions, therefore warn that civilian crime could increase, reflecting concerns related to depolicing and reduced enforcement activity (MacDonald 2019; Owens 2019; Prendergast 2021).

[1] Recent debates in the United States illustrate these concerns. Following the deployment of masked law enforcement personnel at large-scale protests, several U.S. states, including California, New York, New Jersey, and Pennsylvania, have introduced legislation requiring officers to display visible identification or limiting the concealment of identifying information during public deployments.

To study the impact of police officer identification, we exploit the staggered introduction of this reform across German states. The reform followed a broader shift toward more liberal governments during the 2010s, but implementation timing varied substantially due to state election cycles and administrative delays rather than contemporaneous policing incidents. Importantly, no other major policing reform was introduced during this period. In addition, the reform applies only to officers operating in riot gear and closed units, creating a sharp and well-defined change in individual identifiability in precisely those settings where accountability frictions are most severe. This institutional structure allows us to isolate the effects of individual officer identification while holding constant broader policing institutions and legal standards.

Empirically, a central challenge is that deployments of officers in riot gear are not directly observed and do not occur at random. We address this challenge by combining administrative data from German State Criminal Police Offices on daily district-level incidents of excessive use of force and civilian crime with information on professional soccer matches. Professional soccer matches provide settings with predictable deployments of officers in riot gear and occur at dates determined months in advance, independently of contemporaneous policing incidents or public scrutiny.

We find that the introduction of police officer identification tags leads to a substantial reduction in excessive use of force. Using within-district variation around the staggered implementation of the reform, we show that recorded incidents of excessive use of force decline sharply following the introduction of individual identification. Our baseline estimate suggests a reduction of about two thirds relative to pre-reform levels. These estimates isolate behavioral responses to increased identifiability while holding constant location-specific and time-specific

factors. Throughout the paper, we capture police misconduct using recorded incidents of excessive use of force, a legally defined and consequential subset of police misconduct.

We conduct a wide range of robustness checks to assess the validity of our identification strategy. The estimated decline in excessive use of force is robust to restricting attention to narrow time windows around reform implementation, alleviating concerns about confounding long-run trends. Results are also robust to alternative control groups and to specifications that account for district-specific time-varying shocks. Finally, placebo specifications that assign matchdays to neighboring districts within the same state yield no corresponding effects, suggesting that the estimated decline is not driven by contemporaneous state-level changes in reporting practices, policing resources, or other state-specific trends.

One potential concern is that our estimates reflect changes in reporting rather than changes in underlying officer behavior. Several features of the institutional setting mitigate this concern. Excessive use of force is a serious criminal offense under German law, and incidents are typically reported following formal complaints or medical documentation, implying relatively high baseline reporting rates. Moreover, false or malicious reporting is itself punishable, and our empirical design controls for district-specific changes in reporting practices over time.[2]

We next study civilian crime around the reform. There is a concern that increased oversight may discourage proactive policing. If officers respond to heightened scrutiny by reducing engagement or avoiding intervention, potentially leading to increased civilian crime. We address this concern by examining civilian crime outcomes, including simple assault, aggravated assault,

[2] By making individual officers identifiable, the reform could, in principle, increase reporting and attribution of excessive use of force, as victims, witnesses, and prosecutors may find it easier to file and pursue complaints once responsibility can be assigned to a specific officer. Such a reporting channel would bias our estimates upward, leading to an increase in recorded excessive use of force absent any behavioral response, rather than the decline we document.

and property damage.[3] Across all categories, we find no evidence of an increase in civilian crime following the introduction of officer identification tags. If anything, civilian crime declines slightly, providing no indication that the reform led to reduced police engagement.

Finally, we examine how the reform affects public trust in the police. Using individual-level data from the German Victimization Survey, we find that trust in the police increases in states that adopt officer identification tags. The increase is concentrated among groups more likely to interact with officers operating in riot gear, consistent with improved perceptions of accountability.

Together, our findings show that modest identification requirements can substantially increase police accountability without reducing public safety.

I. Institutional Background

Police misconduct is subject to formal legal constraints in many jurisdictions, including criminal sanctions for serious violations. In practice, however, these legal frameworks often do not translate into effective enforcement. Across many societies, allegations of police misconduct rarely lead to legal consequences, as investigations are frequently discontinued and prosecution rates remain substantially lower than for comparable civilian offenses (UNHCR 2020; ACLU 2021). Excessive use of force by police officers constitutes one prominent and particularly consequential example of such misconduct.[4]

[3] These crime categories are less dependent on proactive police detection than many other offenses (e.g., weapon possession). Assaults typically involve victims who seek medical attention, which can trigger reporting through medical staff, or lead victims to file complaints independently of police presence. Property damage is frequently reported for insurance purposes, making reporting relatively stable even when policing intensity changes.

[4] In Germany, only about 3% of such cases proceed to court. Unlike comparable civilian offenses, incidents of police misconduct are frequently dismissed prior to judicial review. The most common cause for such dismissals is a lack of individual officer attribution (Abdul-Rahman et al. 2023).

In Germany, which we study in this paper, these general challenges are reflected in the institutional structure of police misconduct investigations. Public prosecutors are responsible for investigating allegations of police misconduct and deciding whether to bring cases to court. To proceed, prosecutors must be able to attribute alleged misconduct to identifiable individual officers. In practice, this requirement poses substantial challenges, as incidents of excessive use of force often occur in dynamic settings involving multiple officers, making it difficult to reconstruct events and assign individual responsibility. Moreover, because public prosecutors' offices are frequently overburdened, investigations are commonly delegated to neighboring police departments, raising concerns about both independence and rigor (Abdul-Rahman et al. 2023).

The most common reason for discontinuing proceedings in cases of alleged police misconduct is the inability to identify the officers responsible. This problem is particularly pronounced in deployments involving riot gear and closed units, where officers wear helmets, face coverings, and standardized protective equipment that obscure individual identity and hinder external identification.[5] Although fellow officers may be aware of who was involved, they often do not testify, and neighboring departments may refrain from actively seeking such testimony. This reflects a well-documented culture of silence within police forces that discourages officers from naming colleagues and can lead to coordinated efforts to shield individuals from accountability (Armacost 2003; Muehlheusser and Roider 2008; Rivera 2024).

The police officer identification laws we study are adopted at different times across German states, generating substantial variation in adoption timing. Between 2012 and 2019,[6] eleven of

[5] Abdul-Rahman et al. (2023, pp. 72–73) report that approximately 86% of proceedings against police officers deployed at mass events are discontinued. Among these cases, at least 46% are discontinued due to the inability to identify the individual officers involved.

[6] Adoption dates of police officer identification laws by German state are as follows: Berlin (January 1, 2012); Brandenburg (January 1, 2013); Schleswig-Holstein (October 1, 2013); Bremen (July 1, 2014); Rhineland-Palatinate

Germany's sixteen states enact laws requiring police officers to wear identification tags when operating in riot gear or closed units. Two states adopt such tags in subsequent years. The identification tags apply only to police officers operating in closed units and wearing riot gear, and do not affect routine patrol policing. The reforms remain politically and legally contested.[7]

Under the reform, officers are identified by a five-digit alphanumeric code displayed on the uniform, rather than by name. Before the reform, officers in closed units could be identified only at the unit level, but not individually, using a four-digit unit code. The new system therefore enables attribution at the individual officer level, with codes linked to identities through an internal police registry. In deployments involving riot gear and collective action, this change addresses a central identification constraint. Survey evidence shows that individuals reporting excessive use of force in such settings are typically unable to identify the specific officer involved, even when incidents are recorded on video (Abdul-Rahman et al. 2023).

Across states, the timing of adoption of police officer identification laws primarily follows differences in state election cycles and is generally not triggered by state-specific policing events or accompanied by other major policing reforms. In Germany, most governments are formed by multi-party coalitions, and major policy changes are typically negotiated in coalition agreements at the beginning of a legislative term. Such agreements are rarely revised afterward, as amendments can threaten coalition stability. Following a broader shift toward more liberal governments during

(August 31, 2014); Hesse (February 1, 2015); North Rhine-Westphalia (February 28, 2017, abolished on October 24, 2017); Thuringia (May 1, 2017); Mecklenburg-Western Pomerania (January 1, 2018); Saxony-Anhalt (April 28, 2018); Hamburg (November 1, 2019); Saxony (May 1, 2024); and Baden-Wuerttemberg (July 31, 2024).

[7] Political opposition to identification tags has included legal challenges by police unions and staff councils. In Berlin, for example, the police union and the general staff council filed a lawsuit against the introduction of identification tags. Legal disputes culminated in a 2019 ruling by the Federal Administrative Court, following a case brought by police officers from Brandenburg who argued that the reform infringed on their personal rights. The court dismissed the claim, emphasizing that the pseudonymized nature of the identification tags constitutes a key safeguard for officers' personal rights.

the 2010s, several states committed to introducing identification tags. Because state elections are staggered and only a limited number of states hold elections in any given year, opportunities for reform arise gradually, leading to slow and uneven adoption across states.

A further feature of the reform process is a substantial lag between legal enactment and practical implementation. These delays largely reflect state-specific mediation procedures and the timing of administrative court rulings. In many cases, implementation occurs more than a year after adoption, and implementation timing varies considerably across states. Together, variation arising from staggered election cycles and implementation lags reduces concerns that the reform coincides systematically with contemporaneous changes in policing strategies or resources.

II. Data

A. State Police Data

Our main analysis draws on proprietary micro-level crime data obtained from twelve of Germany's sixteen state police agencies, covering the period from 2009 to 2020.[8] The dataset contains daily district-level counts of recorded cases of excessive use of force, defined as incidents of assaults by officers while on duty under Section 340 of the German Criminal Code, as well as measures of civilian crime. Specifically, we observe simple assault under Section 223, aggravated assault under Sections 224 through 226, and property damage under Section 303 at the district-day level. The data structure allows us to track both excessive use of force and civilian violence at a high temporal and geographic resolution.

B. Additional Data

[8] We do not have micro-level crime data from Bavaria, Bremen, Lower Saxony and North Rhine-Westphalia. Due to internal reporting system changes, Rhineland-Palatinate provided data for the years 2014 to 2020 only.

Our research design exploits variation in violence around professional soccer matchdays, which are large public events that typically involve substantial deployments of police forces in riot gear. Information on match dates is obtained from Weltfußball. We restrict the analysis to the period prior to the COVID-19 pandemic, as from March 2020 onward professional soccer games in Germany were played without spectators or under severe attendance restrictions. During this period, matches no longer involved large fan gatherings or the regular deployment of police forces, making them unsuitable for our analysis.

We complement the main administrative data with two additional sources. First, we use information from the Central Coordination Office for Football Matches (Zentrale Informationsstelle Sporteinsätze), which collects nationwide statistics on incidents linked to professional soccer events. Unlike the district-day structure of our primary dataset, this source provides game-level data for all 992 matches played in the top three German leagues during the 2018/2019 season.[9] Second, we use data from the German Victimization Survey, conducted by the German Federal Police in 2012 and 2017. The survey provides representative individual-level information on perceptions of personal safety and trust in the police, allowing us to examine broader attitudinal responses to changes in police officer identifiability.

C. Descriptive Statistics

Table 1 presents summary statistics based on the district-day level data used in the main analysis. Our primary outcome variable, Excessive Use of Force, captures recorded incidents of excessive use of force. Across all district-day observations, the average number of excessive use of force cases is 0.04. This figure rises substantially on professional soccer matchdays, reaching

[9] Data is only available starting from the 2018/2019 season. Later seasons are affected by the COVID-19 pandemic.

0.1684. To contextualize this increase, we compute a non-matchday benchmark by averaging the number of incidents exactly one week after a matchday in the same district, provided that day is free of games. This benchmark yields a mean of 0.1305 excessive use of force cases, suggesting elevated levels of excessive use of force specifically on matchdays.

Compared to excessive use of force, civilian offenses are more frequent. On average, we observe 5.02 cases of simple assault, 1.88 cases of aggravated assault, and 8.37 cases of property damage per district and day. These offenses are also more common on matchdays.

Professional soccer matchdays account for 2.64 percent of all district-day observations in the dataset. Among these matchdays, 19.33 percent involve matches in the first league, 33.25 percent in the second league, and 41.29 percent in the third league.

III. Empirical Analysis

A. Excessive Use of Force

Research Design

To examine the effect of police officer identification on excessive use of force, we exploit the staggered introduction of identification mandates across German states. Our empirical strategy focuses on professional soccer matchdays, which provide a setting with predictable deployments of police officers in riot gear and closed units—the environments in which the identification reform applies and individual identifiability changes discretely. Professional soccer match schedules are determined months in advance and are orthogonal to contemporaneous policing incidents, reforms, or public scrutiny, making them well suited for causal analysis (Rivera and Ba 2025).

Our identification strategy exploits the fact that the reform affects officer identifiability only in specific policing settings—namely, deployments involving riot gear—while leaving

routine patrol policing unchanged. We therefore compare changes in excessive use of force on matchdays, which proxy for such deployments, to changes on non-matchdays within the same district.[10] This difference-in-differences structure allows us to net out district-specific trends and time-varying shocks that affect policing more broadly. Focusing exclusively on matchdays would not allow us to distinguish reform effects from contemporaneous changes in reporting practices, policing intensity, or public scrutiny that affect all days equally.

Empirically, a central challenge is that deployments of officers in riot gear are not directly observed and do not occur at random. We address this challenge in two steps. In the first step, we isolate matchday-related excessive use of force by exploiting within-district variation between matchdays and non-matchdays before and after the reform. This step identifies how the excess use of force associated with matchdays changes once officers become individually identifiable. In the second step, we expand the analysis to include never-treated districts as additional controls in a triple-difference framework, further accounting for time-varying shocks that may differentially affect policing behavior across locations.

In the first step, we estimate the following Poisson regression:

$$\mathrm{E}[Y_{dt}] = exp\big[\beta_1(Matchday_{dt}) + \beta_2(Post_{st}) + \beta_3(Matchday_{dt} \times Post_{st}) + \alpha_d + \gamma_y + \delta_m + \zeta_w + \eta_{d\times y} + \theta_{d\times m} + \kappa_{d\times w}\big], \quad (1)$$

where Y_{dt} is the number of excessive use of force incidents in district d on day t, $Matchday_{dt}$ is an indicator equal to one if a professional soccer match takes place in district d on day t, and

[10] We validate this measure with intraday district-level administrative data for a very limited number of states and years. Relative to comparable days in the same district without professional soccer games, we adopt an event study design and show that most abnormal incidents of excessive use of force and civilian crimes occur in the two hours before and the two hours after the match. During the game itself, the level of incidents is comparable to that on non-matchdays.

$Post_{st}$ equals one after the implementation of the reform in state s. The coefficient of interest, β_3, captures the change in matchday-related excessive use of force following the introduction of individual identification.

The specification includes district fixed effects α_d and a rich set of time fixed effects, including year, month, and day-of-week indicators, which absorb time-invariant district characteristics and common temporal shocks. In our most saturated specification, we additionally include interactions of districts with year, month, and day-of-week fixed effects to flexibly control for district-specific time patterns. We also control for public holidays following Andres et al. (2023). Standard errors are clustered at the district level.

In the second step, we extend the analysis by including districts that host professional soccer matches but were never subject to identification mandates during the sample period. We implement a stacked triple-difference design in which each treated district is matched to all never-treated districts and estimate an augmented version of the baseline specification that includes interactions between treatment status, matchdays, and the post-reform period. This approach compares changes in matchday-related excessive use of force in treated districts before and after the reform to corresponding changes in never-treated districts over the same period. Never-treated districts are those districts that host soccer matches but do not introduce identification tags.

Across specifications, the estimated coefficient on the matchday-by-post interaction captures how the excess use of force associated with riot-gear deployments changes once officers become individually identifiable. By differencing out non-matchday outcomes within the same district and leveraging staggered reform timing across states, the design isolates behavioral responses to increased identifiability while holding constant broader changes in policing institutions, reporting practices, and enforcement intensity.

Primary Results

We begin by documenting the baseline relationship between professional soccer matchdays and excessive use of force. Across all specifications reported in Table 2, matchdays are associated with substantially higher levels of excessive use of force relative to non-matchdays within the same district. Conditioning on district, year, month, and day-of-week fixed effects, we estimate that matchdays are associated with a 42.94 percent increase in recorded incidents of excessive use of force prior to the reform. This pattern reflects the heightened policing intensity and frequent deployments of officers in riot gear on matchdays.

The central result concerns how this matchday-related excess use of force changes following the introduction of police officer identification tags. The coefficient on the interaction between matchdays and the post-reform period captures this effect. Across specifications, the interaction term is negative and statistically significant, indicating that the excess use of force observed on matchdays declines with the introduction of identification tags. In our most saturated specification, the matchday effect falls from 42.94 percent before the reform to 13.2 percent afterward, corresponding to a 69.3 percent reduction in matchday-related excessive use of force.

Importantly, this decline reflects a change in the matchday differential rather than a general reduction in excessive use of force across all policing contexts. By comparing matchdays to non-matchdays within the same district, the design differences out time-varying shocks and district-specific trends that affect policing more broadly. The estimated effect therefore isolates how officer behavior changes specifically in settings where the identification reform alters individual accountability.

To assess the economic magnitude of the effect, we benchmark the estimates against the average number of excessive use of force incidents observed on non-matchdays exactly one week after a matchday in the same district, which is 0.1305 incidents. Applying the estimated pre-reform matchday increase and the post-reform reduction implies a decline of approximately 39 incidents of excessive use of force per year, assuming 992 professional soccer matches per season in Germany. This reduction is sizable relative to the low baseline incidence of excessive use of force and underscores the substantive importance of the reform.

A potential concern is that the estimated decline reflects changes in reporting rather than changes in underlying officer behavior. As discussed in the Introduction, several features of the institutional setting mitigate this concern, including the criminal nature of excessive use of force and the role of formal complaints and medical documentation in generating reports. Moreover, any improvement in reporting attributable to increased identifiability would bias the estimates upward. The observed decline in excessive use of force is therefore consistent with a behavioral response to increased accountability.

B. Robustness

We first restrict the estimation window to one year before and after the staggered introduction of police officer identification tags, rather than the two-year window used in the baseline analysis. This narrower window reduces the influence of longer-term changes in policing practices that may coincide with reform timing. Results remain similar to the baseline estimates (Table 3, Columns I and II). In the most saturated specification, the estimated matchday effect on excessive use of force declines from 35.83 percent before the reform to 15.08 percent afterward, indicating a substantial reduction following the introduction of identification tags.

Second, we expand the sample to include districts that host professional soccer matches but were never subject to identification mandates during the sample period. Using a stacked triple-difference design, we compare changes in matchday-related excessive use of force in treated districts before and after the reform to corresponding changes in never-treated districts over the same period. The coefficient on the triple interaction term is negative and statistically significant (Table 3, Columns III and IV), indicating that the decline in matchday-related excessive use of force is specific to districts exposed to the reform and does not reflect broader trends affecting all districts hosting soccer matches.

Finally, we conduct a placebo test to examine whether the estimated effects could be driven by contemporaneous state-level changes unrelated to the reform. We assign matchdays from treated districts to neighboring districts within the same state that do not host professional soccer matches and treat these dates as pseudo-matchdays. Because these placebo districts share the same state-level reporting institutions, policing resources, and oversight environment, any coincident state-level shocks would be expected to affect both treated and placebo districts similarly. We find no statistically significant effect of pseudo-matchdays on excessive use of force, and no change in this relationship following the reform (Table 3, Columns V and VI), supporting the interpretation that the baseline estimates are not driven by state-level confounders.

C. Civilian Crimes

We next examine whether the introduction of police officer identification tags has unintended consequences for civilian violence. A common argument raised by critics of increased police oversight is that heightened scrutiny may discourage proactive policing. In particular, officers may become more reluctant to intervene in volatile situations or reduce proactive de-escalation efforts in order to avoid scrutiny (Cheng and Long 2022; Prendergast 2021). If so,

declines in excessive use of force could reflect disengagement rather than improved behavior, potentially leading to higher levels of civilian violence.

To examine this argument, we replicate our main identification strategy using district-day data from state police records and re-estimate equation (1) with civilian crime outcomes. Specifically, we consider simple assault, aggravated assault, and property damage, which are the most common offenses committed by fans at professional soccer matches. As in the main analysis, the coefficients of interest capture changes in matchday-related civilian crime following the introduction of officer identification tags. Results from this analysis are reported in Panel A of Table 4.

Consistent with Andres et al. (2023), professional soccer matchdays are associated with higher levels of civilian violence. In contrast to the decline in excessive use of force following the reform, we find no evidence that matchday-related civilian crime increases after the introduction of officer identification tags. The estimated coefficients for simple and aggravated assault are close to zero and statistically insignificant, while property damage declines following the reform. Event-study estimates shown in Figure 1, Panels B through D, confirm the absence of any increase in civilian violence around the time of implementation.

While these results provide no indication of reduced police engagement following the reform, one might still be concerned that the observed patterns reflect changes in crime reporting or detection rather than actual crime. If civilian offenses were primarily recorded through proactive police detection, reduced engagement could mechanically lower reported incidents even in the absence of behavioral change among civilians. This concern is mitigated by the nature of the crimes we examine, which are relatively less dependent on proactive police detection. In cases of serious physical harm, victims typically seek medical treatment, and such treatment often triggers

reporting by medical staff independently of police activity. Property damage is also frequently reported by third parties, such as venue owners, particularly for insurance purposes, making reporting or detection less sensitive to changes in police activity.

To further assess whether the reform affects civilian violence, we complement the district-day analysis with a second proprietary dataset from the Central Coordination Office for Football Matches in Germany. This dataset compiles nationwide statistics on incidents and casualties associated with professional soccer events and provides game-level information for all 992 matches played in the top three professional leagues during the 2018/2019 season. Compared to the district-day data from state police records, these data capture incidents directly in the context of soccer matches and reduce unrelated noise. In addition, the data draw not only on police reports but also on information from private security personnel and emergency medical services, making them less sensitive to changes in police reporting or detection following the introduction of identification tags.

To analyze the game-level data, we estimate a Poisson regression that exploits variation across match locations and timing. The specification compares the same match pairing—i.e., the same two teams playing against each other—under different policing regimes, depending on whether the match takes place in a state that had implemented police officer identification tags at the time of the game. By conditioning on match-pair fixed effects, the design holds constant team characteristics, rivalry intensity, and fan composition that are specific to a given pairing. Additional fixed effects for year, month, and day account for common temporal variation across matches.

The identifying assumption underlying this specification is that, absent the reform, the level of civilian violence associated with a given match pairing would be similar across locations, conditional on time controls. A potential threat to this assumption would arise if the composition

of fans systematically differed depending on whether a match was held in a treated state. This concern is mitigated by the structure of German professional soccer, where organized supporter groups—including those more prone to violence—routinely travel to away games. Because attendance is driven primarily by allegiance to the clubs rather than by local policing policies, treatment variation is plausibly orthogonal to fan composition, conditional on the match pairing.

Panel B of Table 4 reports the results from the game-level analysis. Consistent with the district-day findings, we observe no statistically significant differences in overall civilian crime or assault levels depending on whether matches are played in states that had implemented police officer identification tags. Property damage is lower at treated games, and the total number of casualties does not differ significantly across treated and untreated matches. Taken together, these results provide no evidence that the introduction of police officer identification tags reduces police engagement in ways that increase civilian violence; if anything, estimates point to lower levels of civilian violence.

D. Trust in the Police

In the final part of our analysis, we examine whether the introduction of police officer identification tags is associated with changes in public trust in the police. We use data from the German Victimization Survey, a large-scale repeated cross-sectional survey conducted by the German Federal Police in 2012 and 2017. In each wave, more than 30,000 individuals report on their interactions with the police as well as their general level of trust. These data allow us to assess whether a reform aimed at strengthening officer accountability is accompanied by changes in public perceptions of the police as an institution.

We use the repeated cross-sectional survey data to estimate a difference-in-differences specification that compares changes in trust across states before and after the introduction of police officer identification tags. The dependent variable is an indicator equal to one if a respondent reports a level of trust in the police above the median response across both survey waves. The treatment indicator equals one for respondents residing in states that implemented identification tags between the 2012 and 2017 survey waves. State fixed effects control for time-invariant differences in trust across states, while survey-wave fixed effects account for nationwide changes in public attitudes toward the police. The specification further includes individual-level controls for gender, age, education, employment status, and urban residence.

Table 5 reports the results of this analysis. We find that the introduction of police officer identification tags is associated with a statistically significant increase in trust in the police among respondents residing in treated states. To examine heterogeneity, we interact the treatment indicator with a dummy for adult men aged 18 to 45 residing in urban areas. The estimated coefficient on this interaction is positive and larger in magnitude than the baseline effect, indicating that trust gains are concentrated among groups more likely to interact with officers operating in riot gear. This pattern is consistent with the institutional context of the reform, which primarily affects policing in settings involving riot-gear deployments rather than routine patrol activities.

As a final step, we refine the treatment definition to better capture exposure to the reform. Specifically, we restrict the sample to men aged 18 to 45 and define treatment as residing in an urban area within a state that implemented police officer identification tags. This refinement is motivated by the fact that the reform applies to officers closed units, which are primarily deployed in urban settings and are more likely to interact with younger men. The specification compares changes in trust between urban and non-urban areas within treated states, as well as relative to

corresponding changes in untreated states, while controlling for state-by-urban fixed effects and individual characteristics.

IV. Conclusion

Cases of excessive use of force by police officers impose substantial societal costs and are difficult to deter when individual officers cannot be identified. This paper studies a reform that introduced identification tags for officers operating in riot gear and exploits staggered implementation across German states to estimate its effects.

We find that the reform leads to a large decline in matchday-related excessive use of force, reflecting a reduction in the excess use of force observed in deployments where officers would otherwise be anonymous. At the same time, we find no evidence that the reform discourages police engagement: civilian crime does not increase, and results based on independent game-level data point in the same direction. In addition, public trust in the police rises following the reform, with larger gains among groups more likely to interact with officers subject to the identification requirement.

More broadly, our findings highlight how changes in individual identifiability can shape police behavior and public perceptions, even when applied narrowly. Although accountability reforms may raise concerns about reduced police engagement, we find no evidence that such concerns materialize in civilian crime outcomes. The results speak to ongoing debates about oversight in law enforcement and suggest that targeted institutional changes can improve accountability without detectable adverse effects on public safety.

References

Abdul-Rahman, L., Espín Grau, H., Klaus, L., & Singelnstein, T. (2023). Gewalt im Amt: übermäßige polizeiliche Gewaltanwendung und ihre Aufarbeitung.

Alexander, K. L., S. Rich, and H. Thacker (2022). The Hidden Billion Dollar Cost of Repeated Police Misconduct. Ed. by T. W. Post. *The Washington Post*.

Andres, L., Fabel, M., and Rainer, H. (2023). How much Violence does Football Hooliganism Cause? *Journal of Public Economics*, 225:104970.

Ang, D. (2021). The Effects of Police Violence on Inner-city Students. *The Quarterly Journal of Economics*, *136*(1), 115-168.

Ang, D., Bencsik, P., Bruhn, J., & Derenoncourt, E. (2025). Community Engagement with Law Enforcement after High-Profile Acts of Police Violence. *American Economic Review: Insights*, 7(1), 124-140.

Bor, J., Venkataramani, A. S., Williams, D. R., & Tsai, A. C. (2018). Police Killings and their Spillover Effects on the Mental Health of Black Americans: a Population-based, Quasi-experimental Study. *The Lancet*, 392(10144), 302-310.

Cheng, C., & Long, W. (2022). The Effect of Highly Publicized Police Killings on Policing: Evidence from Large US Cities. *Journal of Public Economics*, 206, 104557.

Chin, G. J., & Wells, S. C. (1997). The Blue Wall of Silence as Evidence of Bias and Motive to Lie: A New Approach to Police Perjury. *U. Pitt. L. Rev.*, *59*, 233.

Cohen, A., Neeman, Z., & Auferoth, F. (2024). Judging under Public Pressure. *Review of Economics and Statistics*, 106(1), 151-166

Cordis, A. S. and Warren, P. L. (2014). Sunshine as Disinfectant: The Effect of State Freedom of Information Act Laws on Public Corruption. *Journal of Public Economics*, 115:18–36.

Dube, O., MacArthur, S. J., and Shah, A. K. (2024). A Cognitive View of Policing. The *Quarterly Journal of Economics*.

Enzmann, D. (2015). Anzeigeverhalten und polizeiliche Registrierungspraxis. Viktimisierungsbefragungen in Deutschland, 1, 511-541.

Fung, A., Graham, M., and Weil, D. (2007). Full Disclosure: The Perils and Promise of Transparency. Cambridge University Press

Grosjean, P., Masera, F., & Yousaf, H. (2023). Inflammatory Political Campaigns and Racial Bias in Policing. *The Quarterly Journal of Economics*, 138(1), 413-463.

Haseman, J., K. Zaiets, M. Thorson, C. Procell, G. Petras, and S. J. Sullivan (June 2020). Tracking Protests across the USA in the Wake of George Floyds Death. Ed. by U. TODAY. *USA TODAY* [Online; posted 3-June-2020].

Hener, T. (2022). Noise Pollution and Violent Crime. *Journal of Public Economics*, 215:104748.

Holz, J. E., Rivera, R. G., & Ba, B. A. (2023). Peer Effects in Police Use of Force. *American Economic Journal: Economic Policy*, 15(2), 256-291.

Jones, J. M. (July 2022). Confidence in U.S. Institutions Down; Average at New Low. Ed. by Gallup.com. Gallup.com [Online; posted 5-May-2022].

Knox, D., Lowe, W., & Mummolo, J. (2020). Administrative Records Mask Racially Biased Policing. *American Political Science Review*, 114(3), 619-637.

Muehlheusser, G., & Roider, A. (2008). Black Sheep and Walls of Silence. *Journal of Economic Behavior & Organization*, 65(3-4), 387-408.

Prendergast, C. (2021). "Drive and Wave": The Response to LAPD Police Reforms After Rampart. SSRN Electronic Journal.

Ravenell, T. (2021). Unidentified Police Officials. *Tex. L. Rev.,* 100, 891.

Rivera, R. (2024). Do Peers Matter in the Police Academy?. *American Economic Journal: Applied Economics*.

Rivera, R. G., & Ba, B. A. (2025). The Effect of Police Oversight on Crime and Misconduct Allegations: Evidence from Chicago. *Review of Economics and Statistics*, 1-18.

Rozema, K., & Schanzenbach, M. (2019). Good Cop, Bad Cop: Using Civilian Allegations to Predict Police Misconduct. *American Economic Journal: Economic Policy*, 11(2), 225-268.

Schuck, A. M., & Rosenbaum, D. P. (2005). Global and Neighborhood Attitudes toward the Police: Differentiation by Race, Ethnicity and Type of Contact. *Journal of Quantitative Criminology*, 21, 391-418.

Schwartz, J. C. (2016). "How Governments Pay: Lawsuits, Budgets, and Police Reform". eng. In: *UCLA Law Review* 63.5, p. 1144.

Weitzer, R., & Tuch, S. A. (2004). Reforming the Police: Racial Differences in Public Support for Change. *Criminology*, *42*(2), 391-416.

Williamson, V., Trump, K. S., & Einstein, K. L. (2018). Black Lives Matter: Evidence that Police-caused Deaths Predict Protest Activity. *Perspectives on Politics*, 16(2), 400-415.

Table 1: Descriptive Statistics

	Type	N	Mean	Std. Dev.	Min	p1	p25	p50	p75	p99	Max	Matchday Mean	Matched Non-Matchday Mean	T-Statistic
Crime Types														
Excessive Use of Force	Count	79,133	0.0391	0.3038	0	0	0	0	0	1	17	0.1684	0.1305	1.71*
Civilian Simple Assault	Count	79,133	5.0160	11.9365	0	0	1	2	5	73	305	14.5469	12.0151	3.42***
Civilian Aggravated Assault	Count	79,133	1.8827	4.8143	0	0	0	1	2	27	159	6.1914	4.8198	4.44***
Civilian Property Damage	Count	71,828	8.3711	20.1762	0	0	2	4	7	125	834	20.5106	18.0941	2.16**
Matchday	Indicator	79,133	0.0264	-	0	0	0	0	0	1	1	-	-	-
Matchday Characteristics														
1st League	Indicator	2,090	0.1933	-	0	0	0	0	0	1	1	-	-	-
2nd League	Indicator	2,090	0.3325	-	0	0	0	0	1	1	1	-	-	-
3rd League	Indicator	2,090	0.4129	-	0	0	0	0	1	1	1	-	-	-
National Cup	Indicator	2,090	0.0612	-	0	0	0	0	0	1	1	-	-	-
High-Risk	Indicator	2,090	0.0971	-	0	0	0	0	0	1	1	-	-	-

Notes: This table presents descriptive statistics for all variables used in the analysis. The unit of observation is the district-day, restricted to districts that host professional soccer matches during our sample period. For crime-related outcomes, statistics are reported separately for *Matchdays* and for benchmark days (*Matched Non-Matchdays)*. The latter are defined as days without a match that fall exactly one week after a matchday in the same district. The final column reports t-statistics from tests of mean differences between matchdays and these matched non-matchdays. Matchday characteristics are based on the 2,090 observed matchdays. *High-Risk Matchday* is an indicator for matches involving fan groups from the 20 clubs with the most violent supporters. We indicate significance at the 1%, 5% and 10% levels (two-tailed) with ***, **, and *.

Table 2: Mandatory Police Officer Identification and Excessive Use of Force

	Dep. Var.: Excessive Use of Force			
	(I)	(II)	(III)	(IV)
Matchday	0.4160***	0.3338***	0.3524***	0.3572***
	(0.0839)	(0.0928)	(0.0900)	(0.0960)
Post	0.0518	0.0820	-0.0849	0.0639
	(0.0353)	(0.1838)	(0.1834)	(0.3680)
Matchday x Post	-0.2172***	-0.2321***	-0.2385***	-0.2332***
	(0.0535)	(0.0593)	(0.0579)	(0.0553)
Observations	43,096	43,096	43,096	19,341
Pseudo R^2	0.5494	0.5534	0.5544	0.4915
District FE	Yes	Yes	Yes	Yes
Year FE		Yes	Yes	Yes
Month FE		Yes	Yes	Yes
Day of Week FE		Yes	Yes	Yes
Public Holiday FE			Yes	Yes
District x Time FE				Yes

Notes: This table presents Poisson regression estimates with fixed effects. The dependent variable in all specifications is *Excessive Use of Force*, the number of assaults committed by police officers while on duty, measured at the district-day level. The sample is restricted to treated districts, defined as those where police officer identification tags were introduced. The analysis is standardized relative to each district's implementation date and limited to a two-year window before and after the policy change. Column (IV) includes interactions of all time fixed effects (year, month, and the day of the week) with the districts. Standard errors, clustered at the district level, are reported in parentheses. We indicate significance at the 1%, 5% and 10% levels (two-tailed) with ***, **, and *.

Table 3: Specification Robustness

	1 Year Pre and Post: Only Treated Districts		1 Year Pre and Post: Stacked With Control Districts		1 Year Pre and Post: Placebo With Neighboring Districts	
	(I)	(II)	(III)	(IV)	(V)	(VI)
Matchday	0.3092***	0.3060***	0.4259*	0.4975**	0.0684	0.0923
	(0.1095)	(0.1058)	(0.2316)	(0.2184)	(0.2624)	(0.2473)
Post	0.0575	-2.0082***	-0.9006*	-0.6064*	9.8771***	13.5928***
	(0.3353)	(0.7171)	(0.4636)	(0.3522)	(1.0437)	(1.2261)
Matchday x Post	-0.2103***	-0.1915***	0.4169	0.2727	0.1797	0.1491
	(0.0648)	(0.0581)	(0.2796)	(0.2624)	(0.4215)	(0.4422)
Treat x Post			0.0241	-1.4951*		
			(0.0993)	(0.8355)		
Treat x Matchday			-0.1528	-0.1878		
			(0.2565)	(0.2447)		
Treat x Matchday x Post			-0.6113**	-0.4644*		
			(0.2980)	(0.2686)		
Observations	18,891	6,591	468,130	144,712	20,432	7,823
Pseudo R^2	0.5512	0.4603	0.2195	0.1613	0.0876	0.0841
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
District x Time FE		Yes		Yes		Yes

Notes: This table presents Poisson regression estimates with fixed effects. The dependent variable in all specifications is the number of assaults committed by police officers while on duty, measured at the district-day level. Relative to Table 2, Columns (I) and (II) restrict the analysis window to one year before and after the staggered introduction of police officer identification tags. Columns (III) and (IV) include never-treated districts – defined as districts that host professional soccer matches but did not adopt identification tags – as additional controls in a stacked design, where all never-treated districts are matched to each treated district. Columns (V) and (VI) present placebo tests, assigning each treated district's matchdays to its neighboring districts that do not host professional soccer games (*pseudo matchdays*). Standard errors, clustered at the district level, are reported in parentheses. We indicate significance at the 1%, 5% and 10% levels (two-tailed) with ***, **, and *.

Table 4: Civilian Crimes and Casualties

Panel A: District-Day-Level Data from State Police Records

	Civilian Simple Assault		Civilian Aggravated Assault		Civilian Property Damage	
	(I)	(II)	(III)	(IV)	(V)	(VI)
Matchday	0.1160***	0.1190***	0.0683***	0.0711***	0.0495***	0.0491***
	(0.0326)	(0.0305)	(0.0188)	(0.0211)	(0.0154)	(0.0152)
Post	0.0209	-0.0156	0.0178	0.0236	0.0196	0.0335
	(0.0267)	(0.0478)	(0.0128)	(0.0216)	(0.0248)	(0.0289)
Matchday x Post	0.0403	0.0392	0.0024	-0.0008	-0.0326*	-0.0430***
	(0.0410)	(0.0399)	(0.0229)	(0.0245)	(0.0186)	(0.0103)
Observations	45,530	45,524	45,530	45,524	38,225	38,219
Pseudo R^2	0.6500	0.6551	0.7652	0.7683	0.8167	0.8199
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
District x Time FE		Yes		Yes		Yes

Panel B: Game-Level Data from Central Coordination Office for Football Matches

	Civilian Crimes			Casualties
	Overall	Assaults	Property Damage	
TreatedGame	-0.2198	0.0467	-1.8951***	-0.0746
	(0.2459)	(0.1965)	(0.4723)	(0.3163)
Observations	874	598	288	516
Pseudo R^2	0.7070	0.4359	0.7910	0.4898
Match Pairing FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes

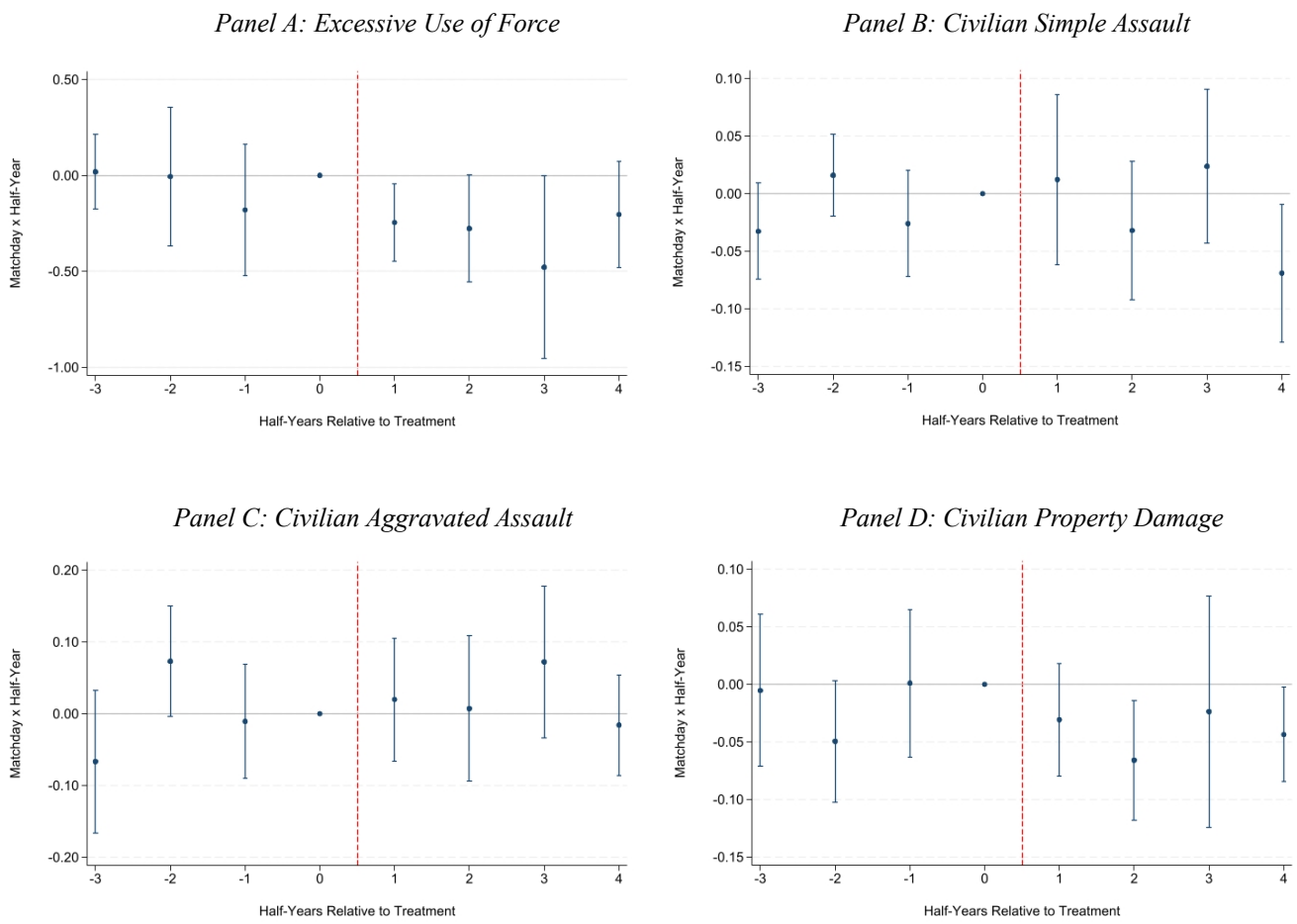
Notes: This table presents Poisson regressions with fixed effects. Panel A uses district-day level data on civilian crimes obtained from state police records. The sample is restricted to treated districts, defined as those where police officer identification tags were introduced. The analysis is standardized relative to each district's implementation date and limited to a two-year window before and after the policy change. Standard errors, reported in parentheses, are clustered at the district level. Panel B uses game-level data from the Central Coordination Office for Football Matches, covering all 992 professional soccer games in the 2018/2019 season in Germany. Each match pairing, i.e., the same two teams playing each other, includes a home and an away game per season. *TreatedGame* is an indicator variable equal to one if a match took place in a state with a police identification mandate. Standard errors are clustered at the match-pairing level. We indicate significance at the 1%, 5% and 10% levels (two-tailed) with ***, **, and *.

Table 5: Police Officer Identification and Trust in the Police

	Treatment: State-Level		Treatment: State-Level		Treatment: Urban	
	(I)	(II)	(III)	(IV)	(V)	(VI)
Treated x Post	0.0177* (0.0100)	0.0203** (0.0102)	0.0143 (0.0107)	0.0171 (0.0111)	0.0446** (0.0182)	0.0437** (0.0190)
Treated x Post x AdultMenUrban			0.0501* (0.0296)	0.0491* (0.0291)		
Observations	41,493	41,302	41,493	41,302	4,300	4,294
R^2	0.0111	0.6551	0.0114	0.0153	0.0140	0.0232
SE	State (bootstrapped)	State (bootstrapped)	State (bootstrapped)	State (bootstrapped)	State x Urban (clustered)	State x Urban (clustered)
Survey Wave FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes		
State x Urban FE					Yes	Yes
Respondent Controls		Yes		Yes		Yes
City Size Control		Yes		Yes		

Notes: This table reports linear regression estimates with fixed effects. The analysis is based on individual-level response data from the German Victimization Survey, conducted by the German Federal Police in 2012 and 2017. The dependent variable is a binary indicator equal to one if a respondent reports trust in the police above the median level. In Columns (I) through (IV), the treatment variable is defined at the state level and indicates whether police officer identification tags were introduced between the 2012 and 2017 survey waves. Columns (II) and (IV) control for respondent characteristics, including gender, age (in 5-year brackets), education, employment status, and whether the respondent resides in an urban area, defined as municipalities with more than 50,000 inhabitants. Columns (III) and (IV) include a triple interaction with *AdultMenUrban*, defined as males aged 18 to 45 in urban areas. In Columns (V) and (VI), treatment is defined at the respondent level and equals one for individuals residing in urban areas within states that introduced identification tags between 2012 and 2017. The sample in these columns is restricted to male respondents aged 18 to 45. Standard errors are clustered at the state-by-urban level and reported in parentheses. For Columns (I) through (IV), standard errors are bootstrapped at the state level. We indicate significance at the 1%, 5% and 10% levels (two-tailed) with ***, **, and *.

Figure 1: Event-time Estimates



Notes: This figure reports Poisson regression estimates based on a sample restricted to treated districts within a two-year window before and after the staggered implementation of police officer identification tags. The dependent variable in each panel corresponds to a different crime type: (A) *Excessive Use of Force*, (B) *Civilian Simple Assault*, (C) *Civilian Aggravated Assault*, and (D) *Civilian Property Damage*. In contrast to our main specification, the *Post*-Indicator is replaced by a series of half-year indicators relative to the implementation date of the identification tags. The half-year period immediately preceding implementation is omitted as the reference category. All specifications include district fixed effects and time fixed effects (i.e., year, month, and day-of-week). Standard errors are clustered at the district level, and 95% confidence intervals are shown.