Three essays in Economics of crime, corruption and conflict

Elda Celislami

Submitted for the degree of Doctor of Philosophy
Department of Economics
University of Reading
September 2025

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Signature:

Elda Celislami

UNIVERSITY OF READING

Elda Celislami, Doctor of Philosophy

Three essays in Economics of crime, corruption and conflict

Summary

This thesis examines how violence—whether perpetrated by the state, produced by war, or embedded in governance—shapes institutional behaviour, public goods provision, and citizens' perceptions of corruption. It integrates administrative records, geospatial data, and nationally representative survey microdata across three self-contained chapters. The first chapter studies how death-investigation laws affect the visibility of police killings in the United States. Using a county-year dataset (2013–2019) and a multi-outcome stratification framework, it shows that where law enforcement can certify causes of death—especially under sheriff-led systems—reported police killings are roughly 40% lower than expected relative to comparable adjacent counties. Underreporting co-moves with reclassification into "circumstances undetermined" homicides and with greater withholding of homicide statistics. The second chapter investigates post-conflict environmental governance in Kosovo by linking NATO bombing intensity, the siting/opening of post-war landfills, and infant mortality. Difference-in-Differences and Triple-Differences designs show that infants born within 6 km of landfills after opening in heavily bombed municipalities face a 3.3–4.6 percentage point higher risk of dying before age one, representing a 118–164% increase relative to the baseline. These effects are concentrated in externally financed landfills built during the reconstruction period, and are accompanied by higher perceptions of local corruption. The third chapter examines how forms of wartime violence shape long-run perceptions of corruption in Kosovo. Combining municipality-level civilian and armed-group casualties with UNDP Public Pulse surveys (2010–2023), civilian-targeted violence is associated with higher perceived corruption—especially toward local institutions—whereas armed-group casualties are linked to lower perceived corruption. Analyses using individual-level exposure to human rights violations provide more consistent and robust evidence of these long-term effects, underscoring the importance of personal experience in shaping institutional trust. Taken together, the chapters show that institutional design conditions accountability, post-conflict governance is fragile, and the legacies of violence are heterogeneous across perpetrators and institutional tiers.

Acknowledgements

I am deeply grateful to my supervisors, whose guidance, encouragement, and belief in me have shaped this journey. I owe my first and biggest thanks to Stefania Lovo, who made it possible for me to join the PhD in the UK (with funding) and encouraged me from our very first conversation—opening doors, including the double-degree path with Turin. I am equally grateful to Stephen Kastoryano, who trusted me with genuine intellectual freedom on our first paper and helped me see what I could aim for. To Giovanni Mastrobuoni, thank you for showing—more than once—how excellence and kindness can go together.

My thanks also go to my annual assessors, Sam Rawlings and Giovanni Razzu, for reading my work year after year and pointing me toward clearer paths when I could not see them. I am grateful to colleagues and friends at Reading and Turin—Simonetta, Fangya, Carl, Tho, Andrea, Brett—and to the PhD cohorts and the 'Monday Meeting' group for conversations that never quite ended with the Zoom call.

Beyond academia, I am lucky to have a circle that kept me grounded. Ludovica, Simonetta, Giulia, and Erina, thank you for checking in, keeping perspective, and showing up in both the quiet and the chaotic moments. Martina and Cristina, the lunch breaks and post-work beers were small islands of sanity. To my wider group of friends and supporters—near and far, in Florence, Reading, Turin, and beyond—your messages, calls, and reality checks mattered more than you know.

To my family: Mum and Dad, thank you for the kind of support that doesn't fit into a sentence—quiet sacrifices, the standard you set for work and integrity, and the steady belief that turned setbacks into steps forward. To my little brother, who seems to overtake my wisdom more each day: your support is always valued and appreciated. To my grandparents, whose stories and example shaped who I am—this work stands on your shoulders. And to my extended family and family-friends who cheered from the sidelines—your encouragement travelled. Most importantly, to Leonardo—my partner and anchor. Thank you for supporting every decision, asking the right questions at the right time, and holding me together through the breakdowns that accompany any long project. This thesis has many contributors; yours was loyal and reliable support.

Contents

List of Tables

List of Figures

Introduction

Crime, corruption, and conflict are among the most persistent obstacles to sustainable economic development. They undermine state capacity, distort resource allocation, and erode public trust. In the framework of political economy, these phenomena are often intertwined: conflict can create fertile ground for corruption, which can weaken the capacity to prevent and prosecute crime, and crime can perpetuate instability. Across contexts, violence frequently appears as the common thread—sometimes as the manifestation of these failures, other times as the force that cements them.

The interaction between these forces has long been recognised in political thought. Hobbes (1651) described the absence of effective institutions as a 'war of all against all' in which life becomes 'nasty, brutish, and short'. Hume (1739) emphasised that political authority ultimately rests not solely on coercion but on conventions, fragile agreements that can be eroded when trust breaks down. Machiavelli (1532) observed that a ruler must be 'a fox to recognise traps, and a lion to frighten wolves', capturing the delicate balance between guile and force in maintaining stability. These perspectives remain highly relevant to contemporary debates on governance, legitimacy, and the institutional consequences of instability.

This thesis examines the long-run institutional, social, and developmental consequences of violence across three self-contained but thematically connected studies. Each study focuses on a distinct empirical setting, one in the United States, two in post-war Kosovo, yet all explore how violence, whether state-perpetrated, conflict-driven, or embedded in governance structures, influences institutional behaviour and citizen perceptions.

Taken individually, these studies make distinct contributions. The first demonstrates that administrative arrangements—often overlooked in political economy—can shape the visibility of crime and the credibility of official statistics, using a new multi-outcome stratification approach that can be applied to other contexts. The second combines geospatial analysis with microdata on health outcomes to identify how war damage interacts with governance failures to affect mortality and perceptions of corruption. The third integrates conflict casualty records with nationally representative survey data to distinguish the effects of different forms of violence on institutional trust. Together, they illustrate the breadth of data sources and empirical strategies—administrative records, survey microdata, geospatial information, and novel econometric designs—used to address the economics of crime, corruption, and conflict.

The U.S. case was chosen because of the unusually large degree of institutional variation across counties in the governance of death investigations, which provides a natural opportunity to study how administrative structures affect accountability in the context of police killings. The Kosovo cases reflect both the relevance of post-conflict governance challenges in the Balkans and my own familiarity with the region's history and institutions. In 1997, Albania experienced an institutional crisis marked by widespread corruption, the collapse of public trust, and violent unrest. My father, then serving as a police officer, worked in an environment where resisting misconduct could carry personal risk. At the same time, regional instability intensified, as many young Albanians crossed into Kosovo to join the Kosovo Liberation Army. This combination of domestic fragility and cross-border conflict illustrated how institutional weaknesses, when combined with security shocks, can produce self-reinforcing cycles of instability—dynamics that remain at the heart of the research presented here.

The first chapter examines how institutional design influences the visibility of state-perpetrated violence in the United States. Responsibility for determining the cause of death varies substantially across counties: some use independent medical examiners with professional qualifications, while others use elected coroners or sheriffs, who may have direct links to law enforcement agencies potentially implicated in killings. Using a novel dataset combining county-level death investigation arrangements, administrative crime statistics, and information on police killings from the Mapping Police Violence database, the analysis applies a multi-outcome stratification framework to estimate the effect of institutional arrangements on classification and reporting. The results show that when elected coroners or sheriffs serve as lead death investigators, police killings are underreported relative to counties with independent medical examiners, with the effect strongest under sheriff-led systems, about 40% lower reporting. These jurisdictions also reclassify more deaths into ambiguous categories and withhold more crime-related statistics. These jurisdictions also reclassify more deaths into ambiguous categories and withhold more crime-related statistics. This mechanism of 'strategic bureaucratic opacity' illustrates how rules intended to ensure accountability can instead be structured to shield state actors from scrutiny.

While the U.S. case highlights how institutional design can obscure violence in a high-income democracy, the post-conflict Balkans reveal a different dimension of the same challenge: how societies emerging from war confront the enduring presence of violence in their institutions and collective memory. As Kadare (1990, 1993), chronicling the region's turbulent past, writes in *Broken April*: 'The past was so full of blood that it soaked into the ground and flowed in the rivers; and still it was not washed away'. In *The Palace of Dreams*, he notes: 'In the Balkans, the past is never dead; it is not even past'. Such imagery speaks directly to the long shadows cast by crime, corruption, and conflict, reminding us that the consequences of instability extend far beyond the period of open crisis.

The second chapter examines the intersection of conflict and corruption in post-war reconstruction, focusing on waste management infrastructure in Kosovo. The 1998–1999 war inflicted severe damage on infrastructure and governance capacity, and in the aftermath, landfills—critical for public health—were often constructed in heavily bombed municipalities under weak institutional oversight. The study investigates how wartime destruction interacts with post-war infrastructure to shape both health outcomes and governance. It combines municipality-level bombing intensity, geocoded landfill openings, and infant mortality from the Multiple Indicator Cluster Surveys, and estimates Difference-in-Differences and Triple-Differences models, complemented by the staggered adoption estimator of Callaway and Sant'Anna. The results indicate that infants born within six kilometres of a landfill after its opening in heavily bombed municipalities face a 3.3–4.6% point higher probability of dying before age one ($\approx$33–46 per 1,000 live births), a 118–164% increase relative to the baseline mortality rate of 28 per 1,000. These effects are concentrated in externally financed landfills built during the reconstruction period, and are accompanied by higher perceptions of local corruption. The results highlight how conflict-driven governance weaknesses and corruption can translate into persistent environmental and health risks.

The third chapter examines how different forms of wartime violence shape long-term perceptions of corruption in post-conflict Kosovo. It combines municipality-level civilian and armed forces casualty data from the Kosovo Memory Book with individual-level survey responses from the UNDP Public Pulse (2010–2023), measuring perceptions of corruption through indices constructed via simple averages, Principal Component Analysis, and Factor Analysis. Two complementary research designs are employed: (i) cross-municipality models relating perceived corruption to civilian and armed-forces casualties per 1990 population; and (ii) within-municipality models using individual-level reports of human-rights violations (HRV), available from 2018 onward, estimated with municipality and year fixed effects. The casualty results show that civilian victimization is strongly associated with higher perceived corruption—especially for central institutions—while armed-forces casualties are associated with lower perceived corruption locally and are small or imprecise for central institutions. HRV-based estimates, which leverage within-municipality variation, support a stronger causal interpretation and indicate that sustained exposure during and after the war increases perceived corruption, particularly toward local institutions. Robustness checks (alternative exposure thresholds, index constructions, placebo tests, and outlier exclusions) yield consistent conclusions. Heterogeneity analyses—especially by ethnicity—highlight meaningful differences. Overall, the legacies of violence depend not only on intensity but also on its nature and timing, with individual experiences reinforcing long-run mistrust in public institutions.

Across these settings, three common insights emerge. First, institutional design matters for accountability: the rules that govern administrative processes can either enable or constrain misconduct. Second, governance in post-conflict environments is fragile; without safeguards, it risks perpetuating the weaknesses that contributed to instability. Third, the social and political legacies of violence are heterogeneous, depending on the nature of the violence and the actors involved.

By integrating evidence from both a high-income democracy and a post-conflict developing economy, the thesis shows that violence—whether as a direct act or as a structural condition—is a persistent and transformative force in institutional life. While the empirical settings here are the United States and post-war Kosovo, the mechanisms identified are not unique to these contexts. Current conflicts in Ukraine, Gaza, Sudan, and elsewhere will leave behind legacies of institutional fragility, contested governance, and fractured public trust. Understanding how these legacies take shape, and how they can be mitigated, is essential to avoiding the governance failures, public health crises, and persistent distrust that have hindered recovery in other post-conflict settings.

Chapter 1

Strategic Bureaucratic Opacity: Evidence from Death Investigation Laws and Police Killings

Note on Co-Authorship. Chapter 1 is based on a manuscript co-authored with my supervisors Stephen Kastoryano and Giovanni Mastrobuoni. My contributions included formulating the research questions, conducting the literature review, undertaking the major data collection, merging and cleaning, and developing the initial empirical strategy and first set of results. The co-authors have granted their consent for the inclusion of this work in the thesis.

1.1 Introduction

Ideally, state institutions always operate within their legal mandates and in view of socially optimal outcomes. In reality, of course, this is not the case. Institutional mechanisms must, therefore, guard against potential misconduct and ensure accountability where it occurs. Ensuring accountability is particularly relevant when it comes to law enforcement, given their entrusted power, and perhaps most symbolic when it comes to police killings (Sherman, 1978). Failure to keep police accountable can affect trust in the state and have far-reaching social consequences, as evidenced by the Black Lives Matter movement. Unfortunately, monitoring law enforcement and police killings is not without difficulty. Legislative action often depends on information reported by police officers and their superiors who, veiled behind the 'blue wall of silence', have an interest in avoiding lengthy investigations and any financial compensation to relatives of wrongfully killed victims (Zimring, 2017).

As it stands, relevant institutions fail at the very first step: accounting for the scale of the problem. Over the years 2013-2019, official records from the FBI's Supplementary Homicide Reports only disclosed half of the total police killings documented in open-source registries. It remains unclear which failures of monitoring mechanisms allow such widespread opacity in reporting by local police departments – reporting that is not subject to any state or federal auditing – and whether this opacity is strategic (Cook and Fortunato, 2023).

Of particular concern are certain pressures and historical laws embedded in the death investigation system. The National Association of Medical Examiners has documented that about a fifth of coroners and medical examiners[1] had been put under pressure by public officials to change the cause or manner of death (Luzi et al., 2013).[2] In addition, 23 states include some counties in which law enforcement is legally permitted to directly certify the cause of death under certain conditions. A subset of these counties also supports systems with notably adverse incentives in which the sheriff is appointed as the de facto lead death investigator.[3]

This study evaluates the causal effect of different laws and systems surrounding death investigations on the underreporting of police killings, implicitly testing for a conflict of interest in reporting. The study further offers empirical evidence of specific methods used by law enforcement to lower officially reported police killings by drawing on various data sources and a novel multi-outcome stratification potential outcome framework with more general applicability. The paper also investigates the drivers, moderating factors, and repercussions of our measured underreporting effect.

The analysis uses a constructed dataset for the years 2013-2019 which includes county-specific information on death investigation systems. Our primary outcome variable in the

[1]Besides law enforcement's ability to certify the cause of death, state and county death investigation systems also differ along other important characteristics. In general, death investigation systems are separated into two categories: medical examiners and coroners. Medical examiners are always appointed, while over two-thirds of coroners are elected officials. Medical examiners also require a medical degree in a vast majority of states and are often trained as forensic pathologists. The requirements for coroner duties vary substantially. In some states, sufficient credentials to examine a deceased person and certify the cause of death are to be 18 years of age, hold a high school degree, and follow a 40-hour course. We return to this discussion later.

[2]As a result, the National Association of Medical Examiners has petitioned for "whistleblower protection" whenever death investigators uncover abuse or criminal activity in officer-involved shootings (Melinek et al., 2013)

[3]News outlets such as newrepublic.com, npr.org, theguardian.com, among others, have also raised red flags concerning the pressures faced by death investigators in the US, notably after the killing of George Floyd. These have so far not resulted in substantive changes in policy.

dataset is the difference between two variables. The first is the 'true' count of police killings from Mapping Police Violence (MPV)[4], an open source registry that opened in the aftermath of the Black Lives Matter movement. The second is the official count reported by law enforcement agencies in the Supplementary Homicide Report (SHR) of the FBI's Uniform Crime Reporting (UCR) program.

Our main result indicates that counties with laws permitting law enforcement to certify the cause of death display 40% more underreported police killings than their comparable adjacent counties. This effect is larger for counties where the sheriff is the lead investigator. We also show that the effect is not solely driven by non-firearm deaths at the hands of police. Even in the case of deaths by firearm which are recorded as involving police in one of several official documents, there is an effect of certification laws on what is deemed a 'justifiable homicide' and thereby falling into the SHR's police killing category.

Our empirical identification approach takes advantage, in a first step, of variation in death investigation laws across and within state borders rooted in movements to prioritize more scientific autopsies rather than to reduce illegitimate interactions with police. It also exploits the fact that changes in legislation are drafted centrally for each state, but apply to all counties even if differences in laws may be responding to the need of a handful of counties. These developments, which we elaborate on, lead to a strategy of comparing the within cluster difference of adjacent counties, some of which lie across state borders, with different laws and systems. Within this context, our causal effects of interest are long-run differences in the reporting of police killings between neighboring counties with comparable potential for permissive cultures of police violence and corruption.[5]

Beyond uncovering differences in underreporting depending on death certification laws, we also explore different possible mechanisms through which sensitive homicide data can be hidden from the public eye. Our results indicate that counties with permissive certification laws report neither higher total homicides, nor higher law enforcement homicides. They do, however, exhibit 35% more deaths categorized as 'circumstances undetermined', the residual SHR homicide category.

Co-movements in the effects of underreporting, 'circumstances undetermined' homicides, and data withholding, are suggestive of cover-up mechanisms. Looking at average

[4]As discussed later, we do not require the 'true' count to be free from error, as long as on average the 'true' measure is unbiased or the bias is unrelated to the death investigation system.

[5]Accounting for adjacency cluster fixed effects, we show that the adjacent counties in our analysis sample are balanced along population, urbanization, crime, income, racial-ethnic distributions, and political voting behavior, which are known to be correlated to underlying views on police violence. We further engage in a close discussion of potential selection on unobserved variables and estimate various robustness and placebo checks. Last, we engage in a dedicated analysis of the potential effects of other differences in death investigation systems.

changes across homicide categories may not, however, answer all policy relevant questions. For example, at the extensive margin, the average effects cannot reveal the share of counties which underreport police killings by re-classifying them into 'circumstances undetermined' homicides rather than using other cover-up methods. It is also unclear whether the stated average effects on 'circumstances undetermined' homicides are driven by counties in which law enforcement agencies would only underreport police killings when facing lenient certification laws. They may also be attributable to counties in which law enforcement agencies would always underreport police killings, but would use different cover-up strategies depending on the permissiveness of certification laws. To answer these types of questions about which population substrata drive our average effects at the extensive margin, we develop a multi-outcome stratification potential outcomes framework in the spirit of the principal stratification literature (Frangakis and Rubin, 2002) and with parallels to recent extensions of the LATE framework (Comey et al., 2022).

Among other results, one of the main findings from the proposed framework is that 85%-95% of counties which would only underreport police killings when subject to lenient certification laws would also simultaneously increase reported 'circumstances undetermined' homicides. We further find that among counties which always underreport police killings, an equal share avoid scrutiny by always substituting these killings into the 'circumstances undetermined' category as those which always exploit other avenues. The relevance of these results likely extends beyond the question of whether law enforcement can certify the cause of death or not since both anecdotal and more formal research (Luzi et al., 2013; Melinek et al., 2013) maintain that few death investigators are shielded from police and political pressure. In terms of research, the multi-outcome stratification framework can also be used as an additional layer to reject any alternative mechanism that does not involve the police forces directly.

The last part of the paper summarizes various additional results which we elaborate on in the Appendix. These topics include: i) the effects of certification laws on data sharing by law enforcement agencies, ii) heterogenous effects by race, underreporting of police killings for Hispanic people on the US-Mexico border, iii) assessing potential effects of other differences in death investigation systems (medical-examiner vs coroner, physician vs non-physician, elected vs appointed, high school degree required vs not), iv) descriptive correlations with monitoring and threat measures (body-worn cameras, clearance rates, assaults, gun laws), and v) moderating effects of the Black Lives Matter (BLM) movement.

Two studies related to ours are noteworthy. In a comparison of California counties over 2000-2018, Prados et al. (2022) find higher underreporting of police killings in sheriff-coroner counties. Cook and Fortunato (2023) find law enforcement data withholding and underreporting of police killings are lower in states with high legislative capacity - proxied by a state-level variable which weighs various measures of legislative term limits, financial resources and staff resources.

Our paper contributes to several strands of literature more generally. First, it adds to the literature on incentive structures and police corruption (Becker and Stigler, 1974; Sherman, 1978; Baicker and Jacobson, 2007; Banerjee et al., 2021; Owens and Ba, 2021). More specifically, with regard to the question of how to monitor the monitors (Rahman, 2012; Cheng and Long, 2018; Mastrorocco et al., 2020; Devi and Fryer Jr, 2020), our article emphasizes the potential conflict of interest for officers reporting police violence, underlining the social cost of poorly designed death investigation laws when it comes to monitoring police for their killings. We also add to the literature describing patterns of fatal police violence (Edwards et al., 2018; GBD et al., 2021; Schwartz and Jahn, 2020) by identifying which mechanisms enable law enforcement to avoid accountability. In particular, we explain the role of related, and sometimes co-opted, institutions in aiding cover-up cultures. These come on top of known internal mechanisms within police departments and unions (Rad et al., 2023). Finally, we also contribute to the literature on the collection and dissemination of government data (Guriev and Treisman, 2019; Martinez, 2022).

The remainder of the paper proceeds as follows. In the next section, we provide some background on the current understanding of police violence and its consequences and on death investigations and their historical contexts. In Section 3 we describe our data. Section 4 outlines our empirical identification strategy and our estimating equation. Section 5 contains the results and also presents our new potential outcomes effect decomposition framework. Section 6 concludes.

1.2 Background

1.2.1 Underreporting of police killings

Patterns and trends in police killings have until recently been difficult to assess due to data reporting issues. Data collected by the FBI through law enforcement agencies and in the National Vital Statistics System (NVSS) through state vital registers are widely acknowledged to undercount the true number of deaths involving police in the US (GBD et al., 2021; Edwards et al., 2018, 2020; Nix et al., 2017; Fryer Jr, 2019).

The fragmented understanding of officer-involved fatality statistics came under focus in the mid-2010s following several high-profile police killings.[6] In response to these events and the ensuing Black Lives Matter protests, several news and non-governmental open-source outlets such as The Counted from *The Guardian*, the Police Shooting Database from *The Washington Post*, Fatal Encounters and Mapping Police Violence began collecting and publishing information about fatal police violence across the US.

Analysis of these new data sources showed that officer-involved fatalities are highest in dense urban areas. However, when considered as a share of total homicides, police accounted for more than 10% of all homicides in smaller rural and non-core areas in contrast to 7% of all homicides in large central metropolitan areas.[7] All in all, between 2012 and 2018, more than 1 in 12 of all homicides of adult men came at the hands of police (Edwards et al., 2018; GBD et al., 2021; Schwartz and Jahn, 2020). In terms of race-ethnicity, underreporting of police killings falls between 50% and 60% for Black, White and Hispanic people, with less underreporting for other minorities (Feldman et al., 2017; GBD et al., 2021).

Beyond the direct consequence of lost lives, there exist many other repercussions on a society facing high levels of police violence (Owens and Ba, 2021). The threat of potentially lethal police violence on unarmed citizens within a neighborhood has been found to directly increase mental health issues (Bor et al., 2018; Tyler et al., 2014) and negatively affect students' school attendance and schooling outcomes (Ang, 2021). These immediate effects have also been shown to corrode the implicit social contract between the state and citizens in the form of lower cooperation with law enforcement (Ang et al., 2021; Zaiour and Mikdash, 2023) and reduced willingness to report violent crime (Desmond et al., 2016; Zaiour and Mikdash, 2023). Notably, however, police killings, in particular of Black and Hispanic victims, have been shown to spur voter turnout (Ang and Tebes, 2020) and citizen support for civilian review boards and criminal justice reforms (Olzak, 2021; Morris and Shoub, 2023).

Several more general problems were also emphasized during the Black Lives Matter movement in relation to officer-involved fatalities. One highlighted issue was the prevalence of systematic and institutionalized permissive cultures of police violence that lack accountability. Proponents of reform argue that police are rarely held accountable for

[6]These included, but were not limited to, the shootings of Michael Brown, Trayvon Martin, Oscar Grant, Charleena Lyles, Stephon Clark, Eric Garner, Sandra Bland, Renee Davis, Philando Castille, Laquan McDonald, and Tamir Rice.

[7]For example, in mountain states (Montana, Idaho, Wyoming, Nevada, Utah, Colorado, Arizona, and New Mexico), police were responsible for about 17% of all homicides between 2012-2018, while in mid-Atlantic states (New Jersey, New York, and Pennsylvania), police accounted for 5% of all homicides.

their misconduct. This lack of accountability is aided by formal union agreements that impede internal audits and the creation of civilian review boards (Rad et al., 2023), but also by the existing informal cover-up culture in police departments, known as the 'blue wall of silence' (Benoît and Dubra, 2004).[8]

Parallel to these developments, a new and more complete official source of violent death records, the National Violent Death Reporting System (NVDRS), was created. The NVDRS, initiated in 6 states and gradually expanded to all 50 in 2018, draws from both police and medical records, but also uses broader criteria of violence for inclusion in the data (Paulozzi et al., 2004). Conner et al. (2019) confirms two important facts with regard to the NVDRS. First, it shows that open source registries such as MPV and Fatal Encounters contain almost all recorded instances of death involving police when a firearm was used (98.3% overlap in the case of MPV). By the same token, it showed that some form of official statement recorded police involvement in almost all firearm police killings recorded in the open-source registries.

Given this traceability, one may ask why law enforcement fall far short of reporting their police killings in the SHR if these already appear in the MPV dataset via news or other outlets and, in the case of firearm killings, which constitute the majority of police killings, are largely traceable, albeit with some effort, in official government records. The usual reason alluded to is a failure to report entirely. As we later discuss, close to 85% of law enforcement agencies report SHR police killings, so a failure to share data due to capacity or other constraints cannot fully explain the problem. There may instead be strategic reasons to try to hide police killings by, for instance, reclassifying them. A first reason to reclassify them may be to avoid investigations and the subsequent administrative leave (see newrepublic.com). As we further elaborate on in our results, there appear to be important incentives for exploiting the blurred line of what falls into the category of a 'justifiable homicide' in the SHR, even for firearm-involved police killings. Strategic reclassification may also be used to impede investigations by discarding evidence, making it harder to build a case against an officer.[9] Last, it may be that not all unreported police killings show up in the MPV data. As long as police departments know that a

[8]Addressing culture is not without complications. Internal monitoring mechanisms must rely on officers breaking norms of police solidarity to investigate their own colleagues. Besides social shunning, these perceived betrayals can result in safety threats when officers require support (Rahman, 2012). External monitoring mechanisms, such as external review boards or added media scrutiny, can avoid these conflicts of interest but, by raising the expected penalty of an officer's errors, may in theory reduce the effectiveness of police (Prendergast, 2003; Gavazza and Lizzeri, 2007). There exists some empirical evidence to support this view, showing that various periods of increased oversight on police led to de-policing and higher crime in the short run (Shi, 2008; Devi and Fryer Jr, 2020; Campbell, 2022).

[9]This was the case in the George Floyd murder until additional independent death investigators were summoned, newrepublic.com.

share of unreported police killings can go undetected, there is an incentive to strategically misreport the ones involving wrongdoing.

Some questionable reclassification statistics have been noted when it comes to non-firearm officer-involved deaths. Indeed, as shown in Feldman et al. (2017) when considering the Guardian's Counted open-source database, 22% of taser-related deaths were reported as an 'accident' in the NVSS, and, consequently, in the NVDRS, 15% stated 'circulatory/respiratory diseases' as the underlying cause of death, and 11% stated 'mental/behavioral disorders'. Conner et al. (2019) further show that 42% of taser-related deaths reported by the Counted, the Washington Post, Fatal Encounters, and Mapping Police Violence could not be discovered in the NVDRS.

1.2.2 Death Investigation Systems in the US

Death investigators in the US play a key role in assessing violent deaths involving homicide and suicide (Hanzlick, 2007). These death investigation systems are divided into two main streams: the coroner system and the medical examiner system.

In general, medical examiners are appointed physicians, often specialized as forensic pathologists.[10] Coroners are most often elected officials without medical degrees or extensive forensic training (Hanzlick, 2007).[11] A common lower bar for eligibility as a coroner is to be older than 18-21 years of age, a county resident, and hold a high school degree. Coroners are also required to go through an initial 40-hour training session during their first year, with 17 of 28 states requiring an additional 8-hour training session every subsequent year (Ruiz et al. (2018), CDC training).

Historically, the medical examiner system evolved out of the coroner system, a relic of British colonialism, as outlined in the *Model Postmortem Examination Act* published in 1954. The stated purpose of the act was to promote systems capable of competently overseeing postmortem investiga-

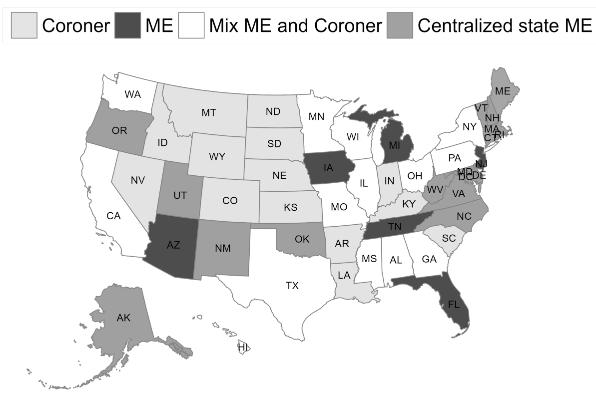


Figure 1.1: Death investigation systems in US.

[10]Some exceptions apply. Wisconsin and West Virginia do not require medical examiners to be physicians (Hanzlick, 2007).

[11]Only 4 states, Kansas, Louisiana, Minnesota, and Ohio, require coroners to hold a medical degree. In addition, in Nebraska, the office of coroner is jointly held with the county attorney, so all coroners are lawyers, while 15 of 39 counties in the state of Washington have coroner duties performed by the county's prosecuting attorney.

tions, in particular for death cases where criminal liability may be involved (Hanzlick, 2014; Leflar, 1955). The call to address the medical incompetence of coroners was followed by a wave of states moving from coroner to medical examiner systems or mixed systems between the early 1960s and mid 1990s.[12] Since 1996, no additional state has moved away from a coroner system and county-specific transitions are rare (Hanzlick, 2007, 2014).[13]

At present, coroners and medical examiners are, depending on the state, distributed into one of four systems as shown in Figure 1.1.[14] In 2019, of the 3,143 counties in the US, only 37% were served exclusively by a medical examiner, but these cover 63% of the total population.[15]

Feldman et al. (2017) and GBD et al. (2021) allude to the question of death investigation systems, mentioning that they find no significant differences in underreporting depending on whether the county had a coroner or medical examiner system. In an analysis limited to California over 2000-2018, Prados et al. (2022) compare sheriff-coroner counties to all others and find higher underreporting of police killings in the former. None of these articles adopts causal analysis methods. The study by Cook and Fortunato (2023) focuses on law enforcement data sharing and underreporting of police killings outcomes close to ours but highlights institutional variation at the state level. Drawing on a conditional on observables independence assumption, they show that lower state legislative capacity - as proxied by a composite measure of legislative term limits, financial resources and staff resources - leads to higher underreporting on both outcomes.

1.2.3 Certifying the Cause of Death

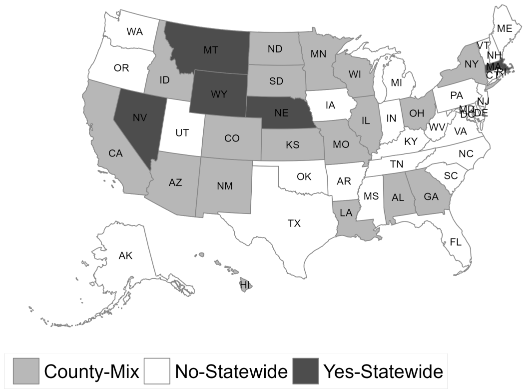
In general, coroners and medical examiners are placed in a vulnerable position with respect to law enforcement. The autopsies they perform must draw from background information and discussions with law enforcement, but the final assessments of the cause of death should, in principle, remain independent of external influence. In practice, also due to

[12]The chronology of these changes is presented in Hanzlick (2007).

[13]Hanzlick (2007) found that only 8 counties out of 3,143 shifted from a coroner system to a medical examiner system between 1996 and 2007.

[14]16 states (and DC) have a medical examiner system in which death investigations are centralized in one location. 6 states have a decentralised county- or district-based medical examiner system. 14 states have a decentralised county-, district-, or parish-based coroner system, and 14 states have a decentralised county-based system with a mixture of coroner and medical examiner offices. 25 of the 28 states with coroners also have an appointed state medical examiner who, upon request, may perform autopsies and provide expert advice, but the coroner still ultimately establishes the cause of death (CDC; Fernandez, 2019).

[15]Numbers derived from our data.



r0.6

Figure 1.2: Can law enforcement certify cause of death?

the repeated interactions, this separation is difficult to maintain. In 2013, the National Association of Medical Examiners conducted a survey on its members' independence in death investigations. It found that over 20% of forensic pathologists working for a medical examiner or coroner reported they were pressured by elected officials or appointees to change their findings. A quarter of those who resisted pressure - threats, termination, intimidation, legal action - suffered consequences for their actions. In addition, approximately 10% of respondents had been asked to sign death certificates or autopsy reports which were inconsistent with the findings of the autopsy (Melinek et al., 2013; Luzi et al., 2013).

The negative repercussions of this conflict of interest are likely exacerbated in legal frameworks which grant stronger authority to law enforcement in the death certification process. Currently, 20 states include a subset of counties in which sheriffs hold the dual post of sheriff-coroner. Of these, some states (California, Nevada, and North Dakota) designate the sheriff as the de facto coroner under some conditions. In addition, Arizona, Massachusetts, Montana, Nebraska, Nevada, New Mexico, and Wyoming allow law enforcement to complete death certificates when the examiner is not present or is unavailable (Ruiz et al., 2018).[16] As shown in Figure 1.2, this results in 18 states in which a subset of counties allow law enforcement to certify the cause of death and 5 states in which all counties allow law enforcement to certify the cause of death. Allowing law enforcement to classify and/or certify the cause of death gives rise to concerning conflicts of interest when it comes to officer-involved fatalities. These conflicts of interest constitute the main focus of our analysis.

[16]The categorization is based on a search algorithm of the Lexis Nexus database and the CDC.

1.3 Data

1.3.1 Officer-involved Fatalities

Our analysis uses a new county-level panel covering the years 2013-2019,[17] constructed from various data sources. Our main outcome of interest is the difference at the county level between a proxy for the 'true' number of officer-involved fatalities and the officially reported number. Our variable for true homicides is taken from Mapping Police Violence (MPV), a non-governmental open-source database on lethal police violence. MPV sources its data from other non-governmental websites (Fatal Encounters, The Washington Post), publicly accessible media sources, and publicly available official data sources from local and state agencies.[18]

Mapping Police Violence compiles information on police-involved deaths in the United States using a combination of media reports, public records, and crowdsourced verification[19]. While MPV is widely used in academic and policy research due to its transparency and national coverage, it does not constitute an administrative census of all police-related fatalities. Because inclusion relies on public reporting and media documentation, MPV may undercount deaths in areas with limited local media presence, lower public visibility, or weaker information dissemination. Moreover, coverage is likely to be higher for incidents that are highly salient or clearly attributable to law enforcement, while ambiguous cases—such as deaths occurring in custody, during medical emergencies, or following delayed complications—may be less consistently captured.

These features imply that MPV may be subject to non-classical measurement error that varies across jurisdictions. In particular, if institutional opacity affects not only official classification but also the public visibility of police-involved deaths, MPV may underrepresent incidents precisely in settings where administrative records are least transparent. As a result, differences between MPV counts and official death records should not be interpreted as evidence that MPV captures the true incidence of police killings, but rather as informative about gaps between public reporting and official classification. In this sense, MPV serves as a complementary benchmark that helps distinguish between deaths that become publicly observable and those that remain administratively obscured. This distinction is central to our analysis, as the outcome of interest is defined by discrepancies between public reporting and official classification rather than by the absolute level

[17]We do not include the years past 2019 due to the generalized changes induced by the COVID epidemic.
[18]The MPV data is widely agreed to be the most complete source of true police homicide data (GBD et al., 2021).
[19]Mapping Police Violence Methodology, Mapping Police Violence Resources

of police violence.

Counter to this true law enforcement homicide data, we consider officer-involved fatalities in the FBI's Supplementary Homicide Report (sourced from the Murder Accountability Project). The SHR measure should include all 'justifiable homicides' of a felon by a peace officer in the line of duty. In general, 'justifiable homicides' include all police killings in which there is an imminent threat to the officer or other people involved.[20] There are alternative official measures of police killings, including the data from the previously mentioned National Vital Statistics System (NVSS) data and the NVDRS. Our decision to use the SHR data is primarily grounded in a desire to speak directly to law enforcement agencies' incentives to withhold or manipulate potentially incriminating data.[21] We present the distributions of the MPV and SHR variables and their difference, our main outcome of interest, in Appendix A.1.

Figure 1.3 maps our outcome of interest.[22] It shows the number of years each county displayed underreporting in the SHR data relative to the MPV data. Many counties show no difference between both police homicide measures in most years. This is mainly because 83.5% of county-year observations display no true police killings at all. Of the 13% of county-years displaying underreporting, 88.3%, show underreporting of 1-2 homicides.[23] However, underreporting is relatively widespread since 43.9% of counties underreport police killings in at least one of the seven years in our data. We provide further descriptive statistics of these outcomes for the US population and different death investigation systems in Table A.2 of Appendix A.5. In general, the underreporting of police killings is lower in coroner counties than in other death investigation systems, mainly due to smaller populations.

Figure 1.4 maps the average report of homicides categorized as 'circumstances undetermined' in the SHR over 2013-2019 by county. As we will show, this homicide category is closely interlinked to underreported police killings. It represents a large share of the total, with 24% of homicides falling under this category among counties that report their SHR

[20] In practice, the line of what falls under this definition is murky. Death by 'Taser', 'Physical restraint', 'Beaten', 'Asphyxiated', among others, may or may not be included. Even in the case of a death by firearm, the distinction is not always clear (Zimring, 2017). Ultimately, this blurry qualification of 'justifiable homicides' constitutes one of the margins which offers room to avoid official reporting, and is integral to the analysis. In addition, both our measures of officer-involved fatalities include deaths in holding cells but exclude fatalities of inmates in correctional institutions. Also, we only observe 53 convicted officers out of 7642 police killings between 2013-2019, so the classification of these cases is unlikely to greatly affect the SHR police killing measure.

[21] In addition, when it comes to the NVDRS, many of the states in our analysis sample only started collecting NVDRS data starting in 2016 or 2018.

[22] We also present separate maps in Figures A.1-A.2 of Appendix A.2 for MPV and SHR police killings.

[23] In a minority of cases, 0.7% of total observations, we find the difference between MPV and SHR data to be negative. In these cases, we impute a value of 0 to the difference.

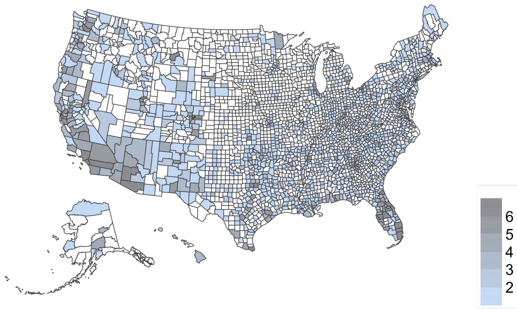


Figure 1.3: Number of years with underreported police killings 2013-2019

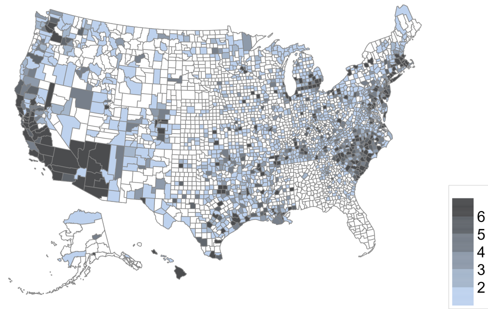


Figure 1.4: Number of years with circ.-undetermined homicides 2013-2019

data to the FBI.

In addition to measures of homicide, we also have yearly information on whether each law enforcement agency shared crime reports with the UCR and NIBRS databases of the FBI. These data are used for a secondary analysis in our results and are described in Appendix A.14.1.

1.3.2 Death Investigation Systems

Our main treatment variable of interest denominating the death investigation system was constructed from multiple sources. We first divide death investigation systems into three mutually exclusive categories, Medical Examiner, Coroner, and Sheriff-Coroner, drawing information from the Center for Disease Control and Hanzlick (2007). For states which allow mixed systems or sheriff-coroners, we searched county by county, consulting county web pages, personal online profiles, LinkedIn profiles, and health center records. We also cross-checked the designations of county systems with records from the 2018 Census of Medical Examiner and Coroner Offices[24] and compared our aggregate numbers by state with those presented in Hanzlick (2007). Appendix Figures A.3-A.6 describe the general distribution of county death investigation systems. Most high population density counties opted to transfer to medical examiner systems during the 1960s-1980s wave of change, but medical examiner counties are also spread throughout more rural areas.

Figure 1.5 displays the distribution of counties that appoint a sheriff as coroner. The sheriff-coroner system is most prevalent in Nevada, California, and Montana, but is also spread across other areas. Figure 1.6 shows the distribution of our main treatment variable of interest. It includes sheriff-coroner counties as well as other counties that allow law enforcement to certify a death certificate as defined in Ruiz et al. (2018). Comparing this

[24]This 2018 census was of limited use beyond cross-checking our definitions of death investigation systems as it contains a large share of missing variable values.

Figure to Figure 1.1, we notice that the additional counties that allow law enforcement to certify the cause of death are spread across both medical examiner and coroner states and counties.

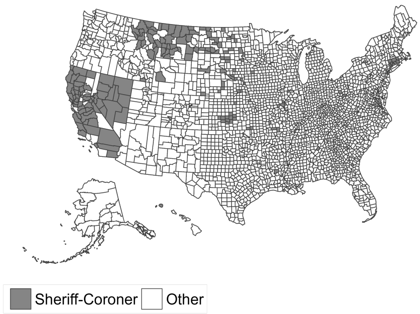


Figure 1.5: Sheriff-Coroner System

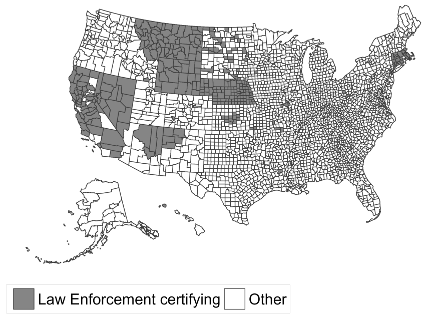


Figure 1.6: Law Enforc. can certify

Finally, we draw additional information from census data concerning county populations, race and gender distributions, and county-level GDP from the Bureau of Economic Analysis, as well as presidential election voting share and indicators for the county urbanization level according to the CDC classification. Table A.2 in Appendix A.5 offers additional descriptive statistics of outcomes and covariates depending on the death investigation system. Of note, Black people are underrepresented in counties permitting law enforcement to certify the cause of death, while Native people are slightly overrepresented. Other than these differences, law enforcement certifying counties display largely similar average characteristics compared to all other US counties.

1.4 Empirical Model

1.4.1 A sketch of the decision model

It is worth outlining a sketch of our underlying decision model to justify our empirical choices and specification, while Appendix A.6 shows a more formal derivation of our empirical specification. Our decision model is one in which, when considering a death at the hands of police, law enforcement (sheriff or other) makes a Beckerian-style cost-benefit analysis. With each death, they consider potential legal, reputational, or pecuniary repercussions, as well as the leniency of certification laws, to decide whether or not to classify - or pressure independent death investigators to classify - the police killing as such in the SHR.

When egregious and unwarranted, some police killings would always go underreported. Others would always be deemed justifiable and reported accurately in the SHR. Our sample selection and estimator intend to capture the change in the remaining decisions to report a police killing as a result of different constraints set by certification laws. Given the setup of this model, the decision by police to go from, say, 0 to 1 underreported killing due to a difference in certification laws results from the same economic decision as that of moving from 3 to 4 underreported police killings. The higher number of underreported killings in the latter case plausibly results from a higher number of police killings in some particular counties. Restated, we assume the extensive and intensive decisions to underreport are not jointly determined. As we elaborate on in the next sections and in Appendix A.6, this decision framework lends itself well to an analysis using closely comparable counties with a Poisson estimator.[25]

1.4.2 Causal Analysis Sample

We conduct our study at the level of death investigation systems, which is the county level. The main causal problem we need to address is that counties with different death investigation systems and laws are unequally distributed across the US, and may be different along many dimensions. In addition, police killings can be underreported due to administrative negligence, controversy in the cause of death, or deliberate acts of misclassification and cover up. Since all of these causes are monotonically increasing in the number of homicides, counties with larger populations and higher urban density will mechanically be more likely to display higher levels of underreporting.

Our principal way to address this issue is by taking a fixed-effects adjacent treated-control county evaluation approach while additionally controlling for total homicides and other potential confounders. More specifically, in the main analysis, we include only the subsample of counties which allow law enforcement to certify the cause of death, including sheriff-coroner counties, and which neighbor at least one county which does not allow such privileges to law enforcement.[26] Similarly, we include as controls the counties which neighbor at least one county which allows law enforcement to certify the cause of death.[27]

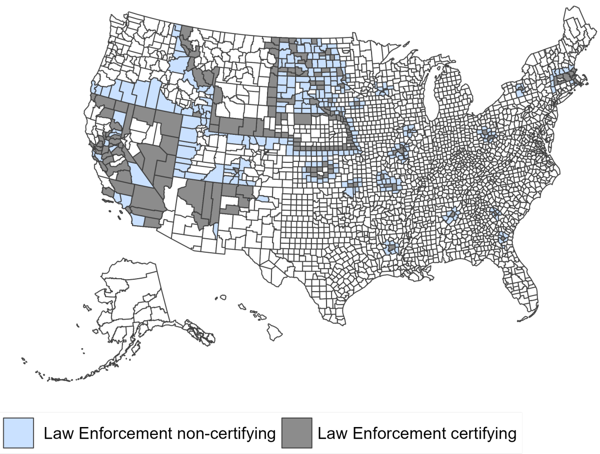
Figure 1.7 displays the adjacent counties included in our main analysis evaluating the effect of law enforcement death certification laws.[28]

[25]Our analysis further estimates a linear model of the number of police killings as well as a linear probability model of any police killing for robustness.

[26]A similar strategy has previously been used, among others, in Dube et al. (2010) when looking at minimum wage effects.

[27]We provide robustness checks throughout in which we condition the inclusion of adjacent counties on similar categories of urbanization. We also offer robustness checks allowing for second-degree adjacency counties to be included.

[28]Following the descriptions in Ruiz et al. (2018), our inclusion criteria for counties as allowing law enforcement to certify the cause of death in Arizona and New Mexico is that the tribal lands cover at least 25% of county territory and that the county includes at least 15% of Native American people. To avoid



r0.6

Figure 1.7: Law enforcement can certify death certificate vs adjacent control county analysis sample

This sample covers 19% of the total US population. Table A.2 in Appendix A.5 further presents some descriptive statistics of the relevant treated and control subsamples compared to the full population. In general, we find that outcome and covariate means of law enforcement certifying counties in our analysis sample are representative of law enforcement certifying counties in general. We also find our adjacent control group outcomes and covariates to be representative of US counties as a whole, although the group slightly over-represents coroner counties and notably under-represents Black people. We return to both points in our analysis.

1.4.3 Adjacent County Poisson Fixed Effect Model

Our main set of results focuses on the difference between 'true' and reported officer-involved fatalities. Let $\mathrm{Y}_{it} = \mathrm{Y}_{it}^{MPV} - \mathrm{Y}_{it}^{FBI}$ denote the observed underreporting of police killings for county i in year t. We take as our baseline model an adjacent county fixed effect Poisson specification to account for the non-negative and right skewed outcome values. More precisely, denoting by J_i the union of a treated county i and its adjacent control counties, our model is given by

$$\mathrm{Y}_{it} = \exp\left(\alpha_{J_i} + \lambda_t + \delta\,\mathrm{D}_i + \gamma'\,\mathrm{X}_{it}\right) u_{it} \tag{1.1}$$

In this model, D_i is our treatment variable of interest. In our main results, it takes value 1 for counties which allow law enforcement to certify the cause of death, and 0 otherwise. X_{it} includes covariates which control for population, urbanisation, political voting behavior,

partial contamination in the control group we do not include surrounding counties in these states which have smaller tribal lands.

local economic conditions, and the racial distribution within each county. It also includes the inverse hyperbolic sine transformation of total reported UCR homicides (H_{it}).[29] α_{J_i} is a fixed effect common for all counties within a cluster J, each formed by a single treated county and all of its bordering controls. It captures additional common unobservables within adjacent county clusters (police culture, violence, etc.). λ_t includes a set of year fixed effects. County fixed effects cannot be included since death investigation systems are fixed at the county level over time.

δ, the parameter of interest in our model, represents the percentage difference in the average underreported police killings between counties permitting law enforcement to certify the cause of death and adjacent counties that do not allow law enforcement to do so. As previously described, since death investigation systems have rarely changed over the past 30 years, our estimator exploits differences within adjacent counties. As noted in Chen and Roth (2024), δ is scale invariant and point identified in this model. In addition, with heterogenous treatment effects δ_i, any two effects $\delta_i = \delta_k$, with $i \neq k$ will contribute equally to the estimate of δ from the Poisson estimator.[30]

1.4.4 Discussion of Causal Effect and Identification

Some assumptions are necessary to allow for a causal interpretation. To discuss these assumptions, we must first clarify the purpose of our evaluation. From a policy perspective, we view the measured effect as a close proxy for the change in underreporting of police killings that would arise if lenient certification laws were removed and the sheriff-coroner system abolished.

More practically, we evaluate differences in the underreporting of police killings for a subset of neighboring counties. The fundamental assumption in our evaluation is that these neighboring counties have similar potential for pernicious cover-up cultures but different death investigation laws and systems which, we argue, are independent of the underreporting outcome. [31] We base this independence assumption on both theoretical and empirical

[29]The transformation $log(H_{it} + (H_{it}^2 + 1)^{1/2})$ is defined for $H_{it} = 0$ and is otherwise approximately equal to $log(2H_{it})$. In theory, this variable could be a bad control. We provide substantial evidence in our results that it is not endogenous to our treatment in a discernible way and that it only serves to improve the precision of our estimates.

[30]This is convenient given the underlying decision model outlined previously in which, for say $\tau = 1$, the extensive margin decision to hide a police killing (go from 0 to 1 underreported police killings due to lenient certification laws) is the same as the decision to hide a killing at the intensive margin (1 to 2, 2 to 3, etc.).

[31]Police departments in counties which permit (under some circumstances) law enforcement to certify the cause of death have, of course, no obligation to make use of their entitlement. It may even be that if one experimentally randomized a set of non-certifying law enforcement counties to more permissive certification laws, little effect would be noticed in the short run. In the long run, however, differences in legislation may lead to pernicious and collaborative executive, and possibly even judicial, norms. We view the results

grounds.

A potential concern with the adjacent-county design is that differences in death investigation systems may affect not only the classification of officer-involved fatalities, but also law enforcement behavior itself. If the institutional features governing death certification are well known locally, officers may face different perceived legal or reputational costs of lethal force across counties, which could induce behavioral responses in addition to differences in reporting. The identification strategy does not require police behavior to be identical across adjacent counties. Rather, it relies on the weaker assumption that, absent differences in death investigation institutions, the underlying determinants of fatal encounters—such as crime conditions, legal standards governing use of force, prosecutorial environments, and policing practices—vary smoothly across space. Adjacent counties within the same state share state law, judicial institutions, training standards, and oversight frameworks, making sharp discontinuities in these determinants at county borders unlikely. Accordingly, the estimated coefficient δ should be interpreted as capturing the equilibrium effect of death investigation institutions on observed underreporting of police killings. This effect may operate through multiple channels, including differences in classification and reporting incentives, as well as potential behavioral responses to institutional opacity. From a policy perspective, both channels reflect the consequences of more permissive death certification regimes, as institutional arrangements that reduce transparency may simultaneously affect how police-involved deaths are recorded and the incentives surrounding their use. This interpretation is consistent with the underlying decision framework outlined above, in which law enforcement responds to institutional constraints when deciding both how police-involved deaths are classified and, potentially, how such encounters are undertaken.

As detailed in Section 1.2.2, major changes in state death investigation systems to address the competency of examiners took place between the 1960s-1980s and are largely traceable. The specific history of changes in state death certification laws is less clear. Some of the differences in state laws, such as removing the sheriff-coroner system, came about with the general death investigation system remodeling (Hanzlick, 2007). These decisions were squarely rooted in a desire to address the medical competence in forensic work as opposed to addressing possible misaligned incentives in reporting police killings when it came to sheriff-coroners. They also offer us a first dimension of quasi-randomization.

Other concessions allowing law enforcement to certify the cause of death arose due

in this study as representing these long-term differences, enabled by more lenient law enforcement death certification laws.

to state-specific considerations, such as granting more independence in tribal territories (Ruiz et al., 2018). There is also reason to believe that laws permitting law enforcement to certify the cause of death if no coroner or medical examiner is present were put in place to address delays in rural areas with fewer resources. This argument is in line with the discussions of capacity considerations in Cook and Fortunato (2023). These concessions offer a second dimension of quasi-randomization. In any given state, the certification laws targeted a handful of counties, but were enacted centrally and therefore affected all counties within the state. For these situations, we assume that neighboring counties divided by a state line are similar along observable and unobservable measures relevant to our study, besides the fact that one county allows law enforcement to certify death certificates due to centrally enacted state laws, while the neighboring county does not give such privileges to law enforcement.[32]

Empirically, our analysis also provides a wide set of balance, robustness, and placebo checks which fail to reject our claim of independence. Speaking indirectly about cultural views on policing, we show in our balance table in Appendix A.5 that demographic and political variables - population, urbanization, racial-ethnic distribution, crime, income - known to be correlated with views on policing culture, including those on voting behavior - are balanced across the treatment and control groups of our analysis in our fixed effect specification.

1.5 Results:

1.5.1 Law Enforcement Certifying Cause of Death

Table 1.1 presents our main results concerning the effect of allowing law enforcement to certify the cause of death, including sheriff-coroner counties, on police killings.[33] The first column shows the estimate of the underreporting of police killings while only including a covariate for total homicides reported in that county-year in the SHR (scaled in arcsinh).[34] The second column adds our full set of control variables accounting for possible differences in urbanization levels, political voting behavior, racial distribution, demographics, and

[32]To our knowledge, there were no important changes to legal structures around death investigation systems during our analysis years which were designed with a view to allow more permissive police protection in cases of police killings. The Criminal Justice and Forensic Science Reform Act in 2014 was supposed to introduce important changes to the death investigation system, but stalled in the US Senate.

[33]All data and estimation code are available at www.skastoryano.com.

[34]The table also displays the total number of county-year observations in the adjacent county sample N_{tot}, the effective observations in estimation N_{eff} which will exclude in Fixed Effects (concentrated) Poisson counties with outcomes which do not change over time, as well as the number of treated year-counties in the effective sample N_{treat} and the number of counties $N_{counties}$.

economic conditions in adjacent counties but does not include any adjacency cluster-specific fixed effects. Both coefficients are positive, showing a $(exp(0.225)-1)*100 \approx 25\%$ to 32% higher underreported police killings in counties that allow law enforcement to certify the cause of death relative to their controls.

Table 1.1: Effects of Authorising Law Enforcement to Certify Cause of Death

Dep. Var.:	Diff. L.E. homicides					
	(1)	(2)	(3)	(4)	(5)	(6)
LE certify	0.252*	0.315***	0.402***	0.074**	0.689***	0.244
	0.225	0.274	0.338		0.524	0.218
	(0.121)	(0.106)	(0.129)	(0.033)	(0.193)	(0.219)
	[0.064]	[0.010]	[0.009]	[0.028]	[0.007]	[0.321]
Covariates	No	Yes	Yes	Yes	Yes	Yes
FE	No	No	Yes	Yes	Yes	Yes
Spec.	Pois.	Pois.	Pois.	Lin.	Pois.	Pois.
N_{tot}	3289	3289	3289	3289	2128	1056
N_{eff}	3289	3289	2561	3289	1666	761
N_{treat}	1224	1224	874	1224	497	279
$N_{counties}$	470	470	366	470	238	109
$\mu_{outc.}$	0.223	0.223	0.286	0.158	0.315	0.248

Note: Table displays transformed coefficient with $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$, followed by estimated coefficient, its standard error, and its p-value. The dependent variable for all columns is *MPV police killings - FBI-SHR police killings* over 2013-2019. Column 5 shows effects on sheriff-coroner counties. Column 6 shows effects on law enforcement certifying counties which are not sheriff-coroners. In columns 3, 5-6 we use a Poisson specification with year and adjacency county cluster fixed effects, and standard errors clustered at the county level. Column 4 uses a within Fixed Effects estimation on treated-control adjacency counties and standard errors clustered at the county level. $\mu_{outc.}$ of column 4 presents the outcome mean for control counties. Set of controls described in Appendix A.5.

This difference in underreporting increases to 40% in our preferred specification, column 3, which includes adjacent county cluster fixed effects. It represents 0.11 more underreported police killings per year, noting that the county mean of 0.29.[35] The effect is of a similar magnitude in column 4 when specifying the adjacent county cluster fixed effect model with a linear specification. We find that counties that allow law enforcement to certify death have 0.07 more underreported police killings than their adjacent controls.[36]

In Table A.4 of Appendix A.7 we provide several additional robustness checks and additional specifications to address potential confounding and qualify our measured effect: (1) we estimate a specification which includes adjacent county-cluster fixed effects and state fixed effects, dropping any state with no variation in our law enforcement certification variable of interest, (2) we estimate a specification which includes adjacent county cluster fixed effects and only includes bordering state counties (3) we estimate our preferred

[35]Note that the mean in all tables represents the effective estimation mean which drops Fixed Effects clusters with no variation in outcomes and weights clusters. As a result, these means will differ from those presented in the summary statistics presented in Appendix A.5.

[36]If we include the few cases in which there are more SHR police killings reported than MPV police killings, the result is virtually the same at 0.076.

specification but only matching treated and control counties with the same urbanization level, (4) we estimate a model only considering as 'treated' sheriff-coroner counties and counties with large tribal territories in Arizona and New Mexico allowing 'controls' to include counties in states which allow law enforcement to certify the cause of death,[37] (5) We estimate our main specification conditioning the analysis on counties that have positive reported SHR homicides, (6) we estimate a specification which includes a covariate indicating whether the death investigator was a physician,[38] (7) we estimate a specification which drops the potential bad control *total homicide* variable, (8) we add to our main specification a gun law covariate capturing the stringency of gun and ammunition laws by state,[39] (9) we provide a specification excluding off-duty officer police killings from the MPV police killings, of which there are only 33 out of the 3,146, since these may not always be officially reported in the SHR as police killings, (10) we account for different possible inclusion criteria by looking at the underreporting outcome of *MPV police killings by firearm - SHR police killings*, (11) we produce estimates when weighting each county-year observation in the sample by the propensity of entering the sample, which gives an approximate extrapolation to the effect on the entirety of the US. Finally, for validation purposes, (12-15) present estimates from pooled OLS and Poisson models and nearest neighbor matching OLS and Poisson effects on the full sample of US counties. Effects estimated on all counties are within the same range or larger than those in our main table. The estimates from the robustness checks in specifications 1-5 also range from 39%-95%. The stated result of 40% from our preferred specification therefore falls at the lower band of this range.

In order to further assess the robustness of our results, we also engage in a placebo evaluation exercise emulating the approach of Bertrand et al. (2004). More specifically, using the specification of column 3 in Table 1.1, we estimate the effect of county level death certification laws in 999 placebo adjacency samples. When constructing the law enforcement certifying designation by county, we ensure that the relative distribution of treatments follows the true distribution in the population.[40] Our results indicate that

[37]A map of these analysis counties is presented in Figure A.11 of Appendix A.3.

[38]This is potentially a bad control, as it is obviously correlated with not being a sheriff-coroner. Our subsequent evaluations of other characteristics of death investigation systems on different adjacent samples, speaking more directly to these potential confounding effects, show, somewhat surprisingly, no effects of medical qualifications.

[39]The gun law variable is a yearly measure ranging from 0-11 based on Giffords law center scorecard, where 11 corresponds to a score of A and 0 to a score of F. The yearly state score weighs the many laws pertaining to ammunition and gun distribution, possession, and right-to-carry within a state. We again do not include this covariate in most of our specifications since it could, potentially, be a bad control.

[40]We randomly select 5 states and assume death certification can be completed by police in all counties therein. Then we select 18 other states and ensure the distribution of death certification laws is randomly distributed holding the relative share of those same laws equal to that in the population.

only 1.2% of placebo estimates are larger than the true estimate of the effect of allowing law enforcement to certify the cause of death. We present the CDF of placebo estimates relative to the true estimate in Figure A.15 of Appendix A.7.

The specification of column 9 in Table A.4 of Appendix A.7 also allows us, using a back-of-the-envelope calculation, to extrapolate the approximate long-term reduction in underreported police killings were the law enforcement certification laws abolished throughout the US. Abolishing these laws would produce a reduction of 328 underreported killings every 10 years. This number may be just a minor representation of the potential consequences given the additional social consequences of abolishing death certification laws inhibiting police accountability.

The final two columns of Table 1.1 separate effects for sheriff-coroner counties and counties in which law enforcement can certify the cause of death but which are not sheriff-coroner counties.[41] We find positive effects on both subsamples but only find significant effects when looking at the subsample of sheriff-coroner counties. The results also suggest that our effects are predominantly driven by sheriff-coroner counties.

Figure A.16 of Appendix A.7 presents effects when excluding each of the 23 states with at least one county that allows law enforcement to certify the cause of death, including sheriff-coroner counties. This allows us to inspect how much leverage each state has on our main point estimate. The effects removing each state remain for the most part close to our main effect. In terms of leverage, we find that when the sheriff-coroner counties and their matched adjacent counties in California are removed, the effect reduces by 56% while they increase by 28% when excluding relevant adjacent counties in Massachusetts. However, these effects are not significantly different from our main effect.

1.5.2 Strategic Re-classification

Table 1.2 delves further into the mechanisms that are used to cover up police killings. Columns 1 and 2 present the separate parts of our underreporting outcome, namely the reported SHR and true MPV officer-involved fatalities, respectively. The results speak to whether underreporting effects are mainly driven by relatively fewer reported killings or relatively higher true police killings in counties permitting law enforcement to certify the cause of death. The latter effect may be indicative that permissive laws aiding the cover-up of police killings contribute to a culture of more police killings. Although point estimates

[41] For the sheriff-coroner analysis of column 5 and the analysis of other law enforcement certifying counties in column 6, the adjacent control counties exclude any law enforcement certifying county.We present maps of included counties for these analysis samples in Figures A.12-A.13 of Appendix A.3.

are lower for reported SHR police killings and higher for MPV, neither is significant at conventional levels.

Figure A.17 in Appendix A.8 delves deeper into the question of violent policing culture looking at sub-circumstances of SHR police killings (column 1 of Table 1.2). We show that police killings in law enforcement certifying counties occur less often in response to attacks by victims and more often when a victim was fleeing or under undisclosed reasons. Appendix A.8 further elaborates on these results.

Returning to the results of Table 1.2, it is difficult to hide a homicide entirely, even if, in principle, a law enforcement officer might tamper with evidence and report a suicide[42] or a natural death.[43] However, only a minority of killings are not by police gunshot ($\sim 6\%$) and therefore could potentially be reclassified as natural deaths or suicides.

Table 1.2: Officer-involved Fatalities and Re-classification Effects

	(1)	(2)	(3)	(4)	(5)	(6)
LE certify	-0.092	0.106	0.347***	-0.014	0.066	0.028**
	-0.097	0.101	0.298	-0.014	0.064	0.028
	(0.105)	(0.094)	(0.072)	(0.085)	(0.039)	(0.011)
	[0.355]	[0.280]	[0.000]	[0.866]	[0.101]	[0.013]
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes	Yes	Yes
Spec.	Pois.	Pois.	Pois.	Pois.	Pois.	Pois.
N_{tot}	3289	3289	3289	3289	3289	319
N_{eff}	2561	2561	2666	3191	3191	315
N_{treat}	874	874	895	1140	1140	226
$N_{counties}$	366	366	381	456	456	93
$\mu_{outc.}$	0.339	0.618	2.223	5.597	5.321	44.298

Note: Table displays transformed coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by estimated coefficient, its standard error, and its p-value. Dependent variable: (1) SHR police killings, (2) MPV police killings, (3) SHR: circumstances undetermined, (4) SHR: Total homicides, (5) SHR: Total homicides - police killings, (6) SHR: Total homicides - police killings, conditional on positive MPV police killings. Poisson estimation with county adjacency year and adjacency county cluster fixed effects, and standard errors clustered at the county level. Set of controls described in Appendix A.5.

An easier way to hide these homicides, especially when they are by gunshot, is to reclassify them in an alternative homicide category. Column 3 of Table 1.2 investigates whether counties that allow law enforcement agents to certify the cause of death show a higher reporting of homicides categorized as 'circumstances undetermined', the catch-all residual circumstance of homicide category in the SHR. We find that law enforcement-certifying counties have 34.7% more 'circumstances undetermined' homicides than their adjacent controls, supporting a theory of strategic reclassification.[44] As discussed in Ap-

[42]Unfortunately, for privacy reasons, suicide data at the county-year level are difficult to access.

[43]In the case of George Floyd, the defense attempted to call upon 'excited delirium', a - non-testable, unproven in humans - cause of death in which an excessive amount of stress on certain animals causes sudden death.

[44]Although not presented in the table, we find this effect to be fully driven by sheriff-coroners counties.

pendix A.9, additional placebo checks do not offer compelling evidence for shifts in other SHR homicide categories.

Column 4 additionally shows that total UCR homicides in law enforcement certifying counties are no different from those in control counties. This result indicates that the higher reported 'circumstances undetermined' homicides in law enforcement certifying counties are not simply due to higher overall homicides in those counties.[45]

Columns 5 and 6 further consider whether law enforcement certifying counties appear to have higher homicides in the remainder of their total UCR homicides after excluding the SHR police killings from column 1. Column 5 presents differences in remainder homicide reporting for all adjacent counties, while column 6 shows these differences conditioning on counties which actually show positive MPV true killings. We find higher non-police killing homicides reported in both columns, but these are only significant in the specification conditioning on counties actually observed to have true police killings. Taken together, the effects in Table 1.2 offer substantial evidence that there are higher homicides specifically in the 'circumstances undetermined' category for counties which allow law enforcement to certify the cause of death.

1.5.3 Decomposing Joint Effects

The evidence for re-classification as a cover-up mechanism seems convincing given the average effects on underreported police killings combined with those on 'circumstances undetermined' homicides. However, several questions remain given these observed effects. First, we would like to know, at the extensive margin, whether the joint effects are due to many counties or limited to only a few. In addition, we are interested in whether the average effects on circumstances undetermined homicides are driven by counties in which law enforcement agencies would only underreport police killings when facing lenient certification laws. They may also be due to counties in which a large share of law enforcement agencies would always underreport police killings but would use different cover-up strategies depending on the certification laws. To address these questions, among others, regarding which population groups contribute to our average effects, we propose a novel multi-outcome potential outcomes framework inspired by the principal stratification literature Frangakis and Rubin (2002) which also draws parallels with the discussion of supercompliers in Comey et al. (2022).

[45]This result also offers empirical evidence that total UCR reported homicides are not a bad control in our main specification.

Identification of Types

For simplicity of presentation, we take a one-period model and ignore the time t subscript. We also leave the conditioning on exogenous covariates X_{it} implicit since identification is nonparametric. Take D_i as the treatment status for agent i, $i = 1, \ldots, N$. In our case, $\mathrm{D}_i = 1$ for counties in our adjacent sample in which law enforcement can certify the cause of death, and $\mathrm{D}_i = 0$ otherwise. In addition, we define the potential outcomes for two variables Y_i^d and W_i^d, $d = 0, 1$. Here, we consider only the case in which Y_i^d and W_i^d are both binary. In our study, $\mathrm{Y}_i^d = \mathbb{1}[\mathrm{Y}_{i,mpv}^d - \mathrm{Y}_{i,shr}^d > 0]$ takes value 1 if county i underreports police killings when exposed to $\mathrm{D}_i = d$, and 0 otherwise. Similarly, take $\underline{\mathrm{W}}_i^d$ as the ratio of reported 'circumstances undetermined' homicides to total reported UCR homicides when exposed to $\mathrm{D}_i = d$. Then $\mathrm{W}_i^d = \mathbb{1}[\underline{\mathrm{W}}_i^d > median(\underline{\mathrm{W}}_i^0)]$ takes value 1 if county i reports $\underline{\mathrm{W}}_i^d$ when exposed to $\mathrm{D}_i = d$ above the sample median when exposed to $\mathrm{D}_i = 0$, and 0 if $\underline{\mathrm{W}}_i^d$ is equal or lower than the median ratio under $\mathrm{D}_i = 0$. Because the median is 0 in the case of 'circumstances undetermined' homicides we refer henceforth to W_i^d as 'positive undetermined homicides'.[46] Referring back to Figures 1.3 and 1.4, we see clear overlap in the number of years in which each county underreported police killings (Y) and positive reported undetermined homicides (W).

With this notation in hand, we can define probabilities representing the possible combinations of potential outcomes. These are defined, omitting for convenience the i subscript, by the joint probability $\Pr(\mathrm{W}^1 = w, \mathrm{W}^0 = w^*, \mathrm{Y}^1 = y, \mathrm{Y}^0 = y^*)$ for each of w, w^*, y, y^* equal to 0 or 1. For example, we are particularly interested in the probability that a county synchronously displays underreported police killings and positive undetermined homicides only when subject to laws permitting law enforcement to certify the cause of death. The probability that a county belongs to this group is $\Pr(\mathrm{W}^1 = 1, \mathrm{W}^0 = 0, \mathrm{Y}^1 = 1, \mathrm{Y}^0 = 0)$. Borrowing from the principal stratification and LATE nomenclature, this probability can be understood as that of the W-outcome complier and Y-outcome complier county type, or principal strata, which we write more succinctly as $\Pr(c^W c^Y)$. We can similarly define outcome always-taker, outcome never-taker, and outcome defier county types for both outcomes of interest, resulting in a total of 16 probabilities representing each combination of county types. Each of these types is described in Table A.5 of Appendix A.10.1.

Our goal is to identify a maximal amount of these 16 probabilities given a set of plausible behavioral assumptions about law enforcement's reaction to changes in death

[46] We keep our more developed definition of W since in later robustness checks on other SHR categories the median values are not 0.

certification laws. Our first two assumptions draw parallels to those of the principal stratification literature Frangakis and Rubin (2002) including the extensions of the LATE literature Comey et al. (2022), with the difference being that we are not only considering classical instrument-treatment-outcome settings.

Assumption A.I: Unconfoundedness:

$$(\mathrm{Y}_i^1, \mathrm{Y}_i^0, \mathrm{W}_i^1, \mathrm{W}_i^0) \perp\!\!\!\perp \mathrm{D}_i$$

Assumption A.II: Outcome Monotonicity: *For all i*

$$(i) \quad \Pr(\mathrm{Y}_i^1 \geq \mathrm{Y}_i^0) = 1$$

$$(ii) \quad \Pr(\mathrm{W}_i^1 \geq \mathrm{W}_i^0) = 1$$

Assumptions A.I is a classical unconfoundedness assumption. It requires that in our adjacent analysis sample all potential outcomes of interest are random with respect to whether law enforcement is permitted to certify the cause of death. A.II draws parallels with Comey et al. (2022) by proposing two monotonicity assumptions on the outcomes of interest.[47] The first excludes the existence of counties in which law enforcement agencies would underreport police killings when not able to certify the cause of death but would not underreport police killings when facing more permissive laws for certifying causes of death. The second part, A.IIii, adds the same monotonicity assumption to the positive undetermined homicide outcome.[48]

Assumption A.II fixes any probability with a defier type to equal zero, reducing 16 unknown probabilities to 9. As shown in Appendix A.10.1, under assumptions A.I and A.II we can already identify a subset of joint probabilities ($\Pr(a^W a^Y)$, $\Pr(n^W n^Y)$) and the shares of each county type depending on the outcome ($\Pr(n^Y)$, $\Pr(c^Y)$, $\Pr(a^Y)$, $\Pr(n^W)$, $\Pr(c^W)$, $\Pr(a^W)$).

Identifying the remaining joint probabilities requires an additional assumption.[49] We elicit the following two assumptions consistent with incentives in our setting and which we further justify in our empirical results,

[47]Because Y_i and W_i are binary variables, we can always transform them by $1 - X$ such that the stated formulation in A.II holds.

[48]Strictly speaking, because 90% of counties comprise of multiple law enforcement agencies, the assumptions must only hold on aggregate at the county level.

[49]One such assumption, as presented in the identification framework of Comey et al. (2022), would be an exclusion restriction on the influence of death certification laws. Unfortunately, their exclusion restriction assumption is not plausible in our setting. It would impose that reporting positive undetermined homicides depends only on whether there were underreported police killings and not on whether the county was facing lenient death certification laws. This is unlikely to hold since police departments may use different cover-up methods depending on the leniency of death certification laws.

Assumption A.III: Cross-Monotonicity Restriction: *For all i*

$$(i) \quad \Pr(\mathrm{W}_i^1 = 1, \mathrm{W}_i^0 = 0, \mathrm{Y}_i^1 = 0, \mathrm{Y}_i^0 = 0) \equiv \Pr(c^W n^Y) = 0$$
$$(ii) \quad \Pr(\mathrm{W}_i^1 = 0, \mathrm{W}_i^0 = 0, \mathrm{Y}_i^1 = 1, \mathrm{Y}_i^0 = 0) \equiv \Pr(n^W c^Y) = 0$$

Assumption A.IIIi and A.IIIii are tailored to our setting but other combinations of 0-probability types can also be used to achieve identification.[50] Assumption A.IIIi states that there are no counties in which, on aggregate, law enforcement agencies would not respond to permissive certification laws by increasing the underreporting of police killings, but would respond to these laws by increasing reported positive undetermined homicides. This seems plausible as long as the main reason for increasing declared positive undetermined homicides in response to more permissive certification laws is to hide certain police killings.

Assumption A.IIIii states that there are no counties in which, on aggregate, law enforcement agencies would respond to permissive certification laws by increasing the underreporting of police killings, but would not respond to these laws by increasing reported positive undetermined homicides. This assumption effectively claims that the most prevalent cover-up strategy for underreporting law enforcement agencies in complier counties is to reclassify police killings as 'circumstances undetermined' homicides.[51] If this is the case, then underreported police killings specifically induced by the more permissive laws will always show up on aggregate at the county level as positive undetermined homicides.

Assumptions A.I-A.III allow us to identify all remaining county-type probabilities ($\Pr(c^W c^Y)$, $\Pr(a^W c^Y)$, $\Pr(a^W n^Y)$, $\Pr(c^W a^Y)$, $\Pr(n^W a^Y)$). We present the simple proof in Appendix A.10.1. We also show that if either of Assumptions A.IIIi and A.IIIii do not hold, we can still infer potentially informative upper and lower bounds on the above type probabilities. From the joint probabilities in equations A.2 and the share probabilities in equations A.1, we can also apply Bayes' rule to obtain all conditional probabilities of interest such as $\Pr(\mathrm{W}^1 = 1, \mathrm{W}^0 = 0 | \, \mathrm{Y}^1 = 1 \, \mathrm{Y}^0 = 0) = \Pr(c^W c^Y) / \Pr(c^Y)$.

Type Probability Estimates

A first point of interest in our results is knowing which substrata drive our main underreporting and 'circumstances undetermined' effects at the extensive margin. Table 1.3 therefore presents results for all probabilities that include complier counties, those induced

[50] Five probability types can be identified under any combination setting one of $\Pr(c^W n^Y)$, $\Pr(a^W n^Y)$ or $\Pr(a^W c^Y)$ to 0 with setting any one of $\Pr(n^W c^Y)$, $\Pr(n^W a^Y)$ or $\Pr(c^W a^Y)$ to 0.

[51] In addition, as elaborated on in Section 1.5.4, we must also assume that not all underreporting complier law enforcement agencies which reclassify police killings opt to withhold sharing their SHR homicide data to the FBI.

to change outcome(s) when facing more lenient certification laws. Each column estimates a probability change using a linear model with adjacency cluster fixed effects.[52]

Column 1 presents the average effect of lenient law enforcement certification laws on our constructed police killing underreporting binary outcome Y, $\Pr(\text{Y} = 1 | \text{D} = 1) - \Pr(\text{Y} = 1 | \text{D} = 0)$. The extensive margin estimate indicates a 3.9 percentage point higher underreporting of police killings in counties with lenient certification laws. Taken in combination with the estimate of column 4 in Table 1.1, the effects show that allowing law enforcement to certify death increases the underreporting of police killings by 0.074, with the effect coming from 3.9% of counties. This translates to an average 0.074/0.039 = 1.89 underreported police killings per year for counties induced to underreport due to more lenient police certification laws.

Table 1.3: Underreporting and Undetermined Causes of Death

Dep. Var.: Effect	Y = 1 $\Pr(c^Y)$	Y = 0 & W = 0 $\Pr(c^W c^Y)$	Y = 1 & W = 0 $\Pr(a^W c^Y)$	Y = 0 & W = 1 $\Pr(c^W a^Y)$
	(1)	(2)	(3)	(4)
LE certify	0.039***	0.037**	0.002	0.005
	(0.012)	(0.021)	(0.021)	(0.007)
	[0.001]	[0.040]	[0.461]	[0.245]
Covariates	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes
Spec.	Lin.	Lin.	Lin.	Lin.
N_{tot}	3289	3289	3289	3289
N_{eff}	3289	3289	3289	3289
N_{treat}	1224	1224	1224	1224
$N_{counties}$	470	470	470	470

Note: Table displays coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by standard error, and p-value. Dependent variable: (1) $\mathbb{1}$*(MPV police killings - FBI-SHR police killings > 0)*, (2) $\mathbb{1}$*(MPV police killings - FBI-SHR police killings > 0)** $\mathbb{1}$*(SHR: circumstances undetermined > 0)*, (3) $\mathbb{1}$*(MPV police killings - FBI-SHR police killings = 0)**$\mathbb{1}$*(SHR: circumstances undetermined > 0)*, (4) $\mathbb{1}$*(MPV police killings - FBI-SHR police killings > 0)**$\mathbb{1}$*(SHR: circumstances undetermined = 0)*. For aesthetic reasons, we present in columns 2-4 results with '-Le cert' to display positive estimates. Panel linear model with year and adjacency county cluster fixed effects. Set of controls described in Appendix A.5.

We can further decompose this total effect into two joint probability effects. The first joint probability effect of $\Pr(\text{W} = 0, \text{Y} = 0 | \text{D} = 0) - \Pr(\text{W} = 0, \text{Y} = 0 | \text{D} = 1) = \Pr(c^W c^Y)$ is presented in column 2 of Table 1.3 and is equal to 3.7 percentage points. Column 3 produces an estimate of $\Pr(a^W c^Y) = \Pr(\text{W} = 1, \text{Y} = 0 | \text{D} = 0) - \Pr(\text{W} = 1, \text{Y} = 0 | \text{D} = 1)$, which is the second joint probability effect contributing to the effect of column 1. These two effects show that among counties with at least one law enforcement agency induced to underreport police killings due to permissive certification laws, 95% $(\Pr(c^W | c^Y) =$

[52]We further discuss data related issues and present additional results in Appendix A.10.1 which generate probabilities from logit specifications.

$\Pr(c^W c^Y)/\Pr(c^Y)$) of these counties would also have at least one law enforcement agency induced to report 'circumstances undetermined' homicides due to the permissive certification laws. Column 4 estimates $\Pr(c^W a^Y) = \Pr(\text{W} = 0, \text{Y} = 1 | \text{D} = 0) - \Pr(\text{W} = 0, \text{Y} = 1 | \text{D} = 1)$ which is not significantly different from 0, meaning that we don't find signs of counties which always underreport police killings but shift to re-classifying them as 'circumstances undetermined' homicides due to permissive certification laws.

Appendix A.10.2 further engages in a placebo exercise inspecting whether the joint effect observed in columns 1-3 is special relative to the other 29 SHR crime circumstances categories. Among all other SHR categories, only 'other manslaughter' and 'other non-specified' display joint significant effects with our underreporting outcome, albeit notably smaller than for the 'circumstances undetermined' category.[53] Appendix A.10.3 also explains how the zero effects of columns 3 and 4 offer an implicit test towards the validity of Assumption III.

Beyond their intrinsic value, these synchronous effects reinforce our initial results on law enforcement certification laws. Any alternative explanation for our measured increases in the underreporting of police killings due to law enforcement certification laws would need to simultaneously explain the increase in 'circumstance undetermined' homicides.[54]

In addition to effects on compliers, there is interest in knowing the joint outcomes for always- and never-taker groups. These remaining type probabilities are presented in Appendix A.11. Some results are worth highlighting. We find that around 83% of counties never underreport police killings, and about 9% of counties always underreport them. We further show that, conditional on underreporting police killings, a county will almost surely report positive undetermined homicides. Similarly, conditional on reporting positive undetermined homicides, a county will almost surely underreport police killings.

1.5.4 Additional results

We present several additional results with dedicated discussion in Online Appendices A.14-A.18, the main lines of which are summarized here.

[53]These results may indicate that underreported police killings are also reclassified into these 'other etc.' categories. Although we cannot exclude this possibility, some of the 'other etc.' categories are associated to a set of conditions (e.g. felony vs non-felony), which is not the case for 'circumstances undetermined'. To be sure, we ran our joint estimate combining the 'other manslaughter' and 'other non-specified' categories with circumstances undetermined and obtained almost the same result (0.042 (0.010)) as that presented in column 2 of Table 1.3. We also estimated our results in column 3 of Table 1.2 with the combined measure and still found significant positive effects.

[54]One such explanation might be to postulate that higher media scrutiny in sheriff-coroner counties results in higher reported MPV true police killings, and as a result more measured underreporting in sheriff-coroner counties. This explanation, however, cannot explain the simultaneous change in undetermined circumstances homicides.

Strategic Withholding of Data

Besides reclassifying police killings into other homicide categories, an additional or alternative strategy would be for law enforcement agencies to avoid sharing homicide information altogether with the FBI. Table A.9 in Appendix A.14.2 assesses differences in data sharing to the UCR and NIBRS programs drawing from recently released data sharing information for all 24,620 law enforcement agencies in the US. While in theory lenient certification laws could lead to increases or decreases in data sharing, we find substantially higher data sharing among law enforcement agencies in counties which permit law enforcement to certify the cause of death.

These participation effects are entirely driven by counties which permit law enforcement to certify the cause of death but are not sheriff-coroner counties. Considering these results in hand with those of Tables 1.2 and 1.3, we surmise that the main cover-up method in sheriff-coroner counties is to reclassify police killings into the 'circumstances undetermined' homicide category. In addition, law enforcement agencies in sheriff-coroner counties do so without additionally resorting to hiding their homicide data from the public. This less cautious behavior is consistent with the theory that sheriff-coroners, united with police departments behind the 'blue wall of silence', are more likely to stand behind the reclassified cause of death in the event of an external inquiry.

Death Investigation Systems and Race

Table A.10 of Appendix A.15 considers whether the main underreporting effects in Table 1.1 are different by race and ethnicity. We find that the effects, while positive for all race-ethnic groupings, are largest and only significant for White people. Although surprising, the absence of significant effects for minorities does not, of course, preclude underreporting for minority police killings altogether. It may simply imply that permissive certification laws do not induce further underreporting. In the case of Black people, the results may also simply reflect the low share of Black people and the lower share of underreported police killings for Black people in the analysis sample relative to the US population as a whole.

Death Investigation Systems and Borders

One consequence of institutions which can be exploited to circumvent accountability is that they will affect those with fewer rights. In particular, families of illegal immigrants killed by police may be less likely to bring a civil case against law enforcement. As such, law enforcement may be more likely to underreport unwarranted killings of illegal immigrants. To explore this question, somewhat informally, we use our adjacent sample approach comparing the underreporting of police killings in counties along the US-Mexico border, to their nearest inland neighboring counties. Assuming Hispanic people along Mexico border counties are more easily identifiable as illegal immigrants by police, and more subject to unwarranted deaths at the hands of police, then we would expect higher underreporting of Hispanics along the immediate US-Mexico border counties. Table A.11 of Appendix A.16 appears to support this in the data since counties on the US-Mexico border display higher underreporting of police killings for Hispanic people than their controls more inland, but display no differences for other racial-ethnic groups.

Additional Differences in Death Investigation Systems

As discussed previously, different death investigation systems and lead examiners may differ along several dimensions. As statistically described in Table A.3 of Appendix A.5 and represented in maps in Appendix A.3, they may be coroner or medical examiner counties, the death investigator may or may not be required to be a physician, the coroners may be appointed or elected, and they may be required to hold a high-school degree or not.

Table A.12 of Appendix A.17 explores whether these partially overlapping differences in laws and systems lead to differences in the underreporting of police killings. Each analysis restricts the sample to adjacent treated and control counties with different death investigation systems and excludes counties in which law enforcement can certify the cause of death. The results do not indicate that our main results of law enforcement certification effects on underreporting are driven by other underlying differences in death investigator characteristics or systems.

Responses to Monitoring and Threats

Appendix A.18 discusses in depth, but mostly descriptively, the relation between monitoring and threat measures - body-worn cameras, clearance rates, assaults on police, gun laws, Black Lives Matter - law enforcement certification laws, and our the police killing underreporting measure.

Table A.13 of Appendix A.18 shows that citizens in counties permitting law enforcement to certify the cause of death do not seem to engage in more violent and lethal assaults against police. Instead, looking at nationwide correlations, Table A.14 of Appendix A.18

informs us that lenient gun laws are positively associated to the underreporting of police killings. Taken together, these results may underline that fear, rather than true threat, is more relevant in determining unwarranted, and subsequently reclassified, police killings. Table A.13 of Appendix A.18 does not indicate any moderating effects of body-worn cameras, but this may be due to the data source which imprecisely measures their actual usage.

In a last analysis on the question of monitoring, Table A.15 of Appendix A.18 considers whether awareness and concern for issues raised by the BLM movement, as proxied by 'Black Lives Matter' Google search trends by state and their year-to-year changes, influence the underreporting of police killings. We show that states with higher concern for the BLM movement, whether positive or negative, display more underreporting of police killings. We also find that law enforcement certifying counties in states with high searches for 'Black Lives Matter' display more underreporting of police killings.

1.6 Conclusion

This paper adds an additional consideration to discussions surrounding the lack of accountability when it comes to law enforcement's excessive use of lethal force: poorly designed institutional laws can be co-opted to hinder the accountability of state actors. In particular, beyond the magnitude of our estimates, the study highlights patterns of reclassification which, given the known pressures from politicians and police on death investigators (Melinek et al., 2013; Luzi et al., 2013), are likely to also apply more widely to counties which do not allow law enforcement to certify the cause of death. The results underline the need for a more assiduous separation of law enforcement from final assessments of the cause of death, especially in the case of police killings. A first step in this process is to increase the checks and balances for death investigation offices. To date, of the 2,342 death investigation offices in the US, only 108 offices are accredited by the National Association of Medical Examiners (NAME) and only 35 are accredited by the International Association of Coroners and Medical Examiners (IACME). Part of these checks and balances should come in the form of external monitoring, ensuring that death investigators are independent of law enforcement when it comes to sensitive death investigations.

In addition to providing a more reliable picture of police killings than the UCR, new open-source registries compiled by private citizens represent a significant step toward holding law enforcement accountable. However, private citizens can be intimidated or threatened, so relying on individuals rather than institutions to uphold the social contract between citizens and police is unstable. Although some families may seek accountability in courts based on unofficial records such as MPV, police killings of people with fewer protections, such as illegal immigrants or sex workers, may never face review if omitted from official records. Results from our newly developed multi-outcome potential outcomes framework suggest that police exploit certain unclear lines of what constitutes a 'justifiable homicide' when deciding how to report police killings. A second policy recommendation would therefore be to categorize all deaths that, prima facie, appear to be caused by police officers as police killings in official records. Although recent NVDRS data offer a notable improvement when it comes to detailed information on homicides, triangulating officer-involved deaths is not immediate, in particular for non-firearm deaths (Paulozzi et al., 2004).

Chapter 2

The Long-Lasting Effects of Bombing on Environmental Management. Evidence from Kosovo

2.1 Introduction

There were times when the peninsula seemed truly large, with enough space for everyone; for different languages and faiths, for a dozen people, states, kingdoms, and principalities. But times changed and with them the ideas of the local people changed, and the peninsula began to seem quite constricting (Kadare, 2000). The conflict in Kosovo in the late 1990s, intensified by ethnic tensions and political unrest, escalated into a full-scale war by 1998. In 1999, NATO launched Operation Allied Force, a 78-day bombing campaign targeting Serbian military and infrastructure sites across all municipalities in Kosovo. The stated aim was to halt ethnic cleansing by striking air defense systems, communication hubs, police force headquarters, ammunition depots, and strategic transport routes (Grant, 1999; ICTY, 2000; Trako, 2018). However, several airstrikes resulted in collateral damage due to targeting imprecision, as acknowledged in the ICTY Final Report (ICTY, 2000), creating plausibly exogenous cross-municipality variation in bombing intensity[1].

The aftermath of conflict extends well beyond the cessation of violence, often leaving

[1]These factors generated cross-municipality variation in bomb density tied to battlefield conditions and the spatial distribution of military assets, not to pre-existing differences in civilian administration or post-war waste-sector governance.

deep scars on governance, public services, and institutional trust. In many post-war societies, the destruction of infrastructure and the influx of reconstruction funds create conditions ripe for corruption and institutional failure. In such settings, vital services like waste management—crucial for public health and environmental safety—are particularly vulnerable to mismanagement. These challenges are especially pronounced in Kosovo, a territory that emerged from the 1998–1999 war with widespread devastation and an urgent need for reconstruction. This paper examines whether the legacy of conflict violence in Kosovo has contributed to persistent institutional failure in environmental governance, with a particular focus on waste infrastructure. It explores whether municipalities more heavily bombed during the 1999 NATO intervention suffer today from higher levels of corruption in waste management and worse health outcomes. In particular, it assesses whether exposure to wartime violence is associated with the construction of unsafe waste landfills and higher infant mortality in the surrounding areas. It also explores whether these outcomes are driven by the misuse of reconstruction funds in the post-war period, thereby shedding light on how violence-induced institutional degradation shapes the long-term management of environmental externalities.

A growing literature documents the enduring effects of early-life exposure to environmental hazards, linking air pollution (Currie et al., 2014; Currie and Neidell, 2005; Arceo et al., 2016), water quality (He and Perloff, 2016; Greenstone and Hanna, 2014), and proximity to mining (Von der Goltz and Barnwal, 2019) to adverse health outcomes. However, few studies examine how post-conflict institutional breakdowns contribute to environmental mismanagement, or how these failures interact with public health, particularly for vulnerable groups such as infants. This paper contributes to closing this gap by focusing on the consequences of war-induced corruption in waste infrastructure and its implications for infant mortality.

This analysis shows that municipalities more heavily bombed during the war are more likely to host unsafe waste landfills, and that these facilities have severe health consequences. Children born within 6 km of post-war landfills in heavily bombed municipalities experience a 3.3–4.6 percentage point increase in infant mortality ($\approx$33–46 per 1,000 births) relative to a baseline of 28 per 1,000.

In the conflict's aftermath, more than 13,500 people were killed or went missing, and between 1.2 and 1.45 million Kosovar Albanians were displaced (Krieger, 2001; UNHCR, 2000). By November 1999, 848,100 of the 1,108,913 displaced persons—roughly 76%—had returned to Kosovo, and over 87% of them resettled in the same municipalities they had lived in before the war. This high return rate, combined with detailed municipality-level data on wartime violence, presents a rare opportunity to study the long-term effects of conflict on local institutions and public service delivery.

International donors poured aid into Kosovo to rebuild essential infrastructure, but the scale of reconstruction and weak institutional oversight created fertile ground for corruption. Waste management projects—key to restoring public health—were often compromised. In municipalities most affected by bombing, the urgency to rebuild sometimes resulted in substandard waste facilities and misallocation of funds. According to UNODC (2022), 56% of respondents perceive corruption to be increasing, with particularly high distrust in local government and the judiciary.

Kosovo's case also bears resemblance to other instances of mismanaged reconstruction aid. For example, studies on post-earthquake recovery in Sicily highlight how large inflows of EU funds were drained off through corrupt networks, leaving infrastructure projects incomplete or non-functional (Aiello and Pupo, 2012; De Angelis et al., 2020). These parallels underscore a broader problem: when post-conflict reconstruction is channeled through weak institutions, the resulting infrastructure may be both wasteful and hazardous.

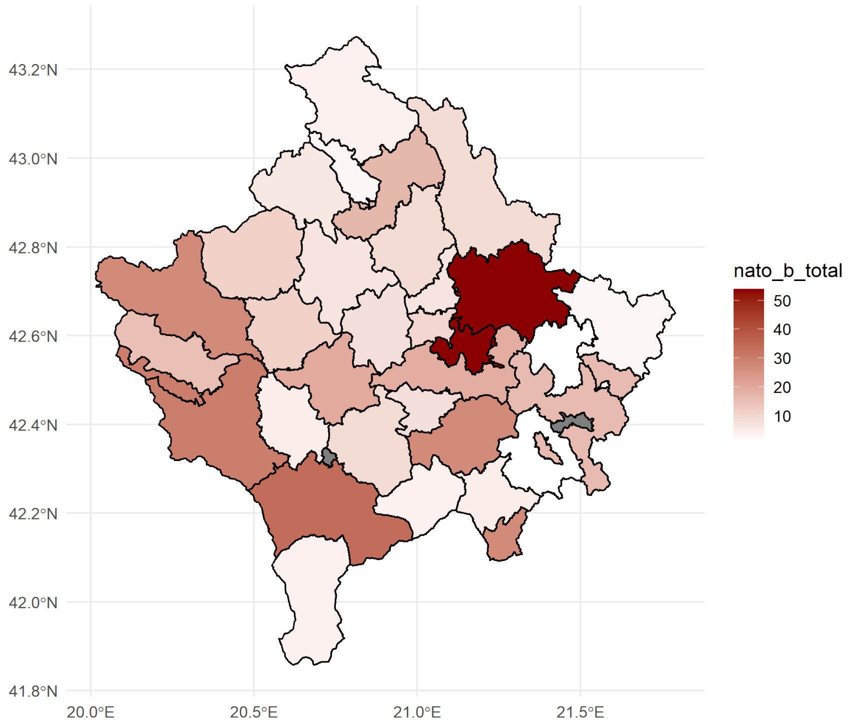
The paper proceeds as follows. Section 2 provides background information on Kosovo, including details about the war and the current waste infrastructure. Section 3 describes the data sources and construction. Section 4 presents the main empirical results, while Section 5 explores the underlying mechanisms. Section 6 discusses robustness checks, and Section 7 concludes.

2.2 Historical Background and Current Situation

2.2.1 Kosovo and the War

Kosovo is a country in the Balkans with partial diplomatic recognition and a history of ethnic diversity and political contestation. The population was predominantly Albanian (about 83% by 1998). Under the 1974 Constitution of the Socialist Federal Republic of Yugoslavia (SFRY), Kosovo held broad autonomy as a socialist autonomous province within Serbia. In 1989, Slobodan Milošević—then President of Serbia—revoked Kosovo's autonomy and imposed direct rule from Belgrade, including the dismissal of many ethnic Albanian state employees. As tensions escalated and repression persisted, an armed movement—the Kosovo Liberation Army (KLA)—emerged by 1997 with the stated aim

Figure 2.1: Kosovo municipalities and bombing intensity



The figure displays a map of Kosovo divided into municipalities, shaded using a red gradient to represent bombing intensity. Darker red indicates higher bombing counts. All municipalities were affected at least once. Pristina experienced the highest number of bombings (54), followed by Prizren (34), while Viti and Novoberde were hit only once. Municipality boundaries correspond to those during the war period; bombing locations with uncertain coordinates were assigned accordingly.

of securing Kosovo's independence.

By March 1998, Serbian and Yugoslav forces launched operations widely described as 'ethnic cleansing', targeting the KLA and Kosovar-Albanian civilians through forced expulsions and widespread abuses. Contemporaneous reporting documented village-by-village expulsions, destruction of identity records, and large-scale displacement. On 30 March 1999, the UNHCR reported approximately 94,000 ethnic Albanians had fled Kosovo since NATO action began on 24 March 1999, with patterns of forced removals described by multiple observers.[2]

In total, approximately 1.5 million ethnic Albanians, representing about three-quarters of the estimated 1998 population, have been forcibly expelled from their homes (U.S. Government, 1999). According to Human Rights Watch, hundreds of thousands of Kosovar-Albanians were forced from their homes, fleeing to neighboring countries like Albania and

[2]Selected contemporaneous excerpts (UNHCR, OSCE, HRW, and press reports) are reproduced in Appendix B.1.

(North[3]) Macedonia. This led to a NATO intervention involving a bombing campaign targeting the Republic of Serbia, including Kosovo. The airstrikes were considered essential to prevent the ongoing and potential massacres and displacements of Kosovar-Albanians by Serbian forces (Cutts, 2000). After a 78-day aerial campaign, on June 9th, 1999, the Republic of Serbia agreed to a peace plan mandating the withdrawal of all Serbian forces from Kosovo, the safe repatriation of refugees and displaced individuals, and the implementation of a UN mission.

Operationally, the 1999 air campaign concentrated on military assets (air defenses, communications, bridges, logistics), producing cross-municipality variation in strike exposure not mechanically tied to pre-war socioeconomic profiles. In the reconstruction period, municipal capacity and oversight over basic services—including solid-waste systems—became pivotal, providing a plausible link from wartime intensity to post-conflict environmental governance. Further details on the political chronology, the international context of the intervention, and contemporaneous accounts of expulsions are provided in Appendix B.1.

2.2.2 Kosovo waste management

Kosovo's waste management system faces persistent capacity and coverage gaps and is repeatedly identified as a priority environmental governance challenge in national strategies and regional assessments (Ministry of Environment and Spatial Planning, 2013; Agjencia për Mbrojtjen e Mjedisit të Kosovës, 2018; European Environment Agency, 2021, 2020). Regional assessments also note ongoing disposal in non-compliant landfills and incomplete municipal planning coverage (European Environment Agency, 2020).

In 2000 the European Union established the European Agency for Reconstruction (EAR) to manage EU-funded reconstruction efforts in Kosovo. One of its focuses was on environmental rehabilitation and industrial clean-up, especially addressing the pollution caused by decades of poor waste management and industrial activities. More on Kosovo's waste management challenges and historic development can be found in Appendix B.2.

External financing: the role of the European Agency for Reconstruction (EAR)

The European Agency for Reconstruction (EAR) served as the EU's main implementing arm for Kosovo's post-war recovery, operating from 2000 to 2008. Initially focused on

[3] The author acknowledges Macedonia's name change to North Macedonia in 2019. However, since most of the historical events discussed occurred prior to this date, and for consistency with the sources cited, the text will use the former name.

Table 2.1: Waste Management Sites in Kosovo

Site	Activity	Surface (ha)	Opening Year	Financer	Municipality	High bombing
Municipal sanitary landfill in Gjilian	Waste landfill	20.5	2003	EAR	4	No
Municipal sanitary landfill in Mitrovice	Waste landfill	7	2001	EAR	11	No
Municipal sanitary landfill in Peje	Waste landfill	4.85	2001	EAR	17	Yes
Municipal sanitary landfill in Mirash-Obliq	Waste landfill	40	2004	EAR	15	No
Municipal sanitary landfill Dragash	Waste landfill	6.161	2003	EAR	5	No
Regional sanitary landfill in Prizren	Waste landfill	14	2003	EAR	20	Yes
Regional sanitary landfill in Podujeve	Waste landfill	8.72	2005	EAR	18	No
Waste transfer station in Ferizaj	Waste transfer station	0.359	2003	EAR	25	Yes
Waste transfer Drenas	Waste transfer station	-	2010	EAR	3	No
Waste transfer Gjakove	Waste transfer station	-	2008	EAR	2	Yes
Tires and conveyor production plant-Suhareke	Waste oils	14	2009	Internal	24	No
Industrial Park in Mitrovice	Industrial landfill	44.5	2014	Internal	11	No
South east part of Cikatrove-Drenas	Waste from clinker of Feronickel	4	2004	Internal	3	No
Landfill of sterile material in Kishnice	Landfill of heavy metals	10.23	2001	Internal	34	Yes
Radioactive materials Trepce-Mitrovice	Storage of radioactive materials	0.04	2005	Internal	27	No
Radioactive materials at Tubeku I Pare, Mitrovice	Storage of radioactive materials	0.03	2005	Internal	11	No
Industrial landfill in Zvecan	Landfill of heavy metals	62.28	2003	Internal	29	No

Notes: The site 'Radioactive materials in Trepce-Mitrovice' primarily refers to Thoriumnitrate, while 'Radioactive materials at Tubeku I Pare, Mitrovice' primarily refers to Strontium, Thorium, and Americium. The 'Site' column lists the official name of each dump site, and the 'Activity' column describes the main purpose of the site. Surface includes the extension of the landfill as per last available report in hectares (ha). The 'Opening Year' indicates the year when the landfill was officially opened. The 'Financer' specifies whether the landfill was financed by the European Agency for Reconstruction (EAR) or other sources ('Other'). The 'mcode' column provides the code of the municipality where the landfill is located, and the 'high bomb' column indicates whether the site is in a municipality with bombing intensity higher than the median. Any missing data for the surface area is denoted with a dash (-).

Kosovo, its mandate soon expanded to Serbia and Montenegro and to (then) Macedonia. Across the Western Balkans, EAR managed a multi-billion-euro portfolio (European Union, 2001; House of Lords European Union Committee, 2002). In Kosovo, EAR funding underpinned core infrastructure—including the 2003–2007 sanitary landfill investments discussed below—illustrating how large external transfers interacted with constrained municipal capacity. This institutional setting motivates the focus on governance quality: layered accountability between donors and municipalities could loosen local oversight over landfill siting, standards, and operations, with downstream consequences for infant health[4].

Because Kosovo's post-war reconstruction unfolded under UN administration, amid large-scale displacement, damaged infrastructure, and urgent winterisation needs, the EAR was set up with delegated budget-implementation powers and operational autonomy to accelerate delivery. In practice, the Agency operated with decentralised ex-ante financial control and adapted procurement timelines under emergency conditions, which sped up contracting and disbursement relative to standard Commission procedures[5].

This allowed the agency to cut bureaucratic corners and speed up the contracting and disbursement process compared to the EU's usual cumbersome procedures[6]. The repositories from European Union (2006) indicate that the Agency concentrated its efforts on rehabilitating infrastructure and public utilities to restore normality in Kosovo. The primary areas of operation included energy, housing, transport, water, enterprise, agriculture, and health, with a total budget of 262 million euros. These efforts had a particularly notable impact on the housing and energy sectors, with the agency managing the entire process from contracting through to implementation and monitoring. According to European Union News Bulletin Oct (2002), improving water and waste utilities was also a key objective. By 2002, over 35 km of new water pipelines and sewer pipes had been in-

[4]Establishment was linked to the European Council's June 1999 commitment; the Commission's Kosovo Task Force implemented emergency programmes before EAR took over in early 2000. The Agency later broadened geographically and, by the end of 2002, oversaw substantial allocations across its operational centres (House of Lords European Union Committee, 2002). See also European Union (2001) for mandate and scale.

[5]Contemporaneous assessments highlight (i) the objective to enhance speed and decentralise operations (European Commission, 1999), (ii) the Agency's use of restricted tenders and shortened time limits 'for reasons of urgency' (European Court of Auditors, 2001, European Court of Auditors report on EAR/Kosovo, OJ C 355/2001), and (iii) that internal (agency-level) ex-ante financial control enabled quick implementation (European Parliament, 2002, European Parliament, Texts Adopted 10 April 2002). See also Council Regulation (EC) No 2667/2000 establishing the EAR's legal personality and mandate.

[6]Ramboll and Insight (2009); Jusufi (2021) argue that the EAR represented an institutional model of independence from politics. The call by the European Commission (EC) for the integration of aid management with policy/politics with EC services or the European External Action Service won the argument for keeping aid away from policy-making bodies of the EU, leading to the termination of the EAR in 2008

stalled, along with more than 12,000 domestic water meters. Additionally, enhancements were made to 13 pumping stations and reservoirs across the municipalities of Pristina and Mitrovica. The source also mentions that the EU funded the design and construction of six landfills built to EU standards, aiming to replace the existing municipal dumpsites as soon as they were completed, ensuring maximum environmental protection.

Unfortunately, no information could be found regarding the six landfills mentioned in the bulletin. However, it is known that they were operational between 2003 and 2007[7] and focuses on municipal waste landfills. Ministry of Environment and Spatial Planning (2019) report talks about seven sanitary landfills; it might be that one has been added after October 2002. The dataset used for this paper analyses five municipal landfills and two regional landfills, all constructed during the EAR's operating years. For the purpose of this analysis, the regional landfills will also be considered as financed by the EAR, consistent with the 2019 report.

Kosovo waste landfill management today

Although seven sanitary landfills were built or renovated during 2003–2007 with external support, the infrastructure remains in poor condition (Ministry of Environment and Spatial Planning, 2019). Less than 60% of residents receive regular waste collection, with coverage around 40% in rural areas, and virtually all collected waste is landfilled, given the absence of alternative treatment infrastructure[8].

Capacity and operational constraints are binding: according to Ministry of Environment and Spatial Planning (2019), the seven sanitary sites receive less than 40% of municipal waste and most are at or near capacity, requiring urgent closure or expansion. Current plans include closing the Mirash landfill (Pristina region) with KfW[9] support and extending sites in Peja and Prizren under the Integrated Waste Management Strategy 2021–2030 (Kosovo Environmental Protection Agency, 2021).

Systemic challenges documented by national and regional assessments include coordination gaps, budget and staffing constraints, low cost recovery, underinvestment, and weak enforcement (European Environment Agency, 2021). These governance frictions—together

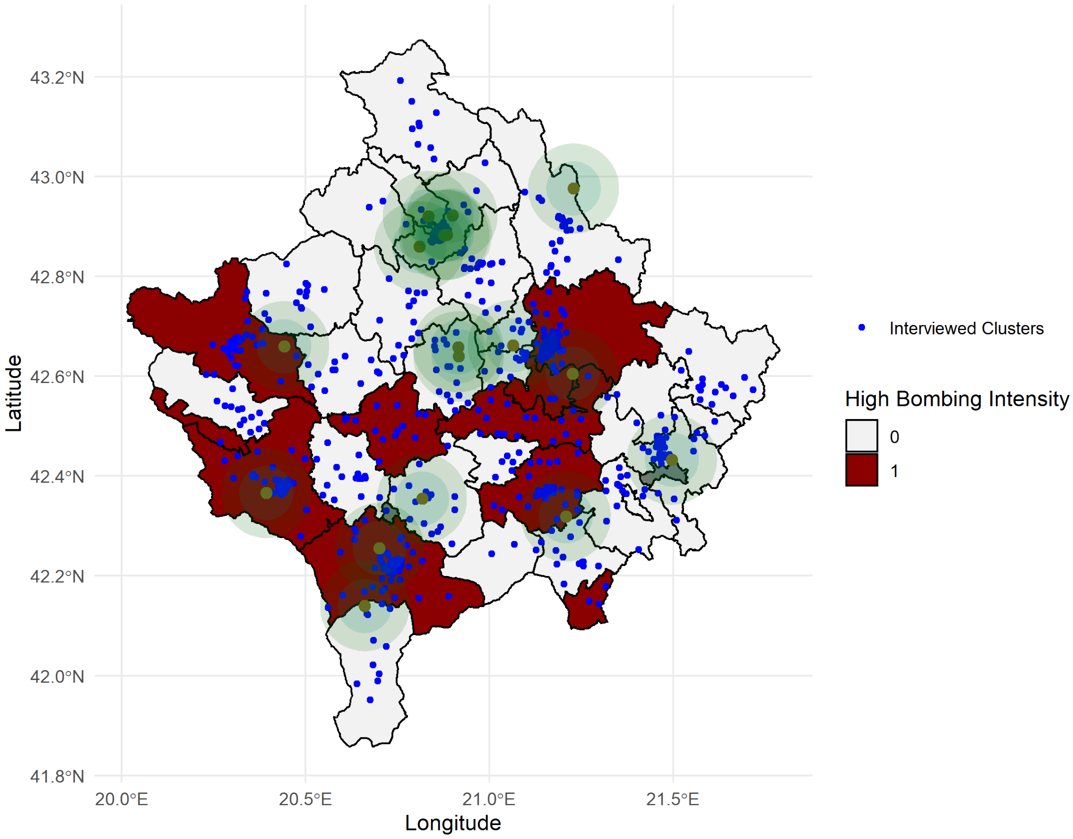
[7] Timeline confirmed also by the Ministry of Environmental and Spatial Planning in the 2018 report. Establishing the agency was seen in retrospect as an efficient way to significantly improve the EU's action in this field. EAR, as an EU agency operating in the field of aid management, showed some level of independence. The evidence found is that, as the aid was politicized, this led to the de facto independence of the EAR. While the EAR had autonomy and independence, the way it was designed was guaranteed. The European Council and the European Commission play an important role in the direct control of how the agency performs its tasks, in order to make sure that it serves EC political goals.

[8] Recycling remains very low (≈15%) and the treatment system is underdeveloped (Agjencia për Mbrojtjen e Mjedisit të Kosovës, 2018).

[9] German state-owned investment and development bank, based in Frankfurt am Main.

with uneven maintenance—are central to the mechanism: lower standard landfill operations raise environmental risks for households living nearby. Further detail, including a case study and financing structure, is provided in Appendix B.2.1.

Figure 2.2: Core spatial data for identification: bombing intensity, landfill sites, MICS clusters, and treatment buffers

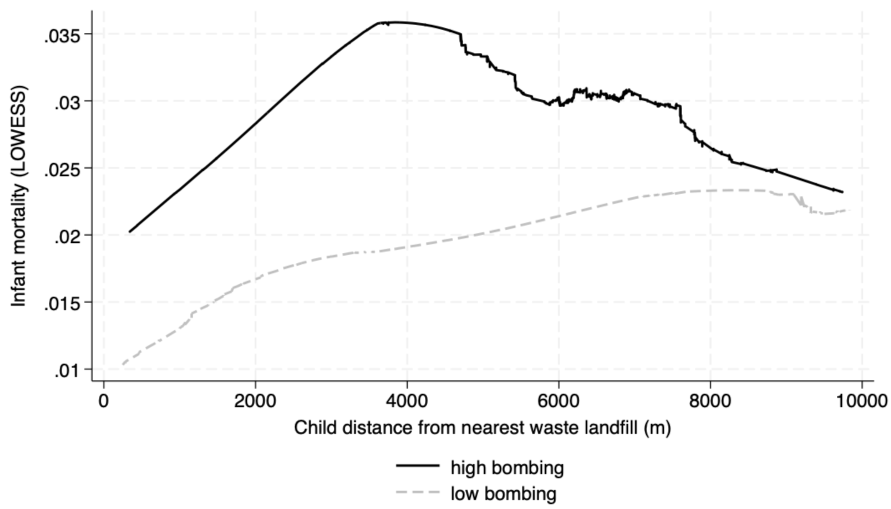


Note: Municipalities are shaded by bombing intensity during the 1999 NATO intervention (dark red marks the municipalities with bombing above-median). Brown points indicate post-war landfill sites, blue points mark MICS survey clusters (household interview locations). Green shaded rings show treatment buffers: the inner ring (0–6km) defines the treated area, and the outer ring (6–10km) defines the comparison area. Together, they form the 10km sample frame. These spatial layers are used to construct the bombing-intensity measure, landfill proximity, and treatment indicators in the empirical analysis.

2.3 Data

This study assembles a novel dataset combining post-conflict environmental and health data with conflict exposure records. The primary source for health outcomes is the Kosovo Multiple Indicator Cluster Survey (MICS) conducted between 2013 and 2014 (MICS5), and 2018-2019 (MICS6). The survey collects information on both households and individuals. This analysis is centered on the women's survey, which specifically includes females in the

Figure 2.3: Infant mortality rate behaviour according to minimum distance from waste landfill



Note: Locally weighted scatterplot smoothing (LOWESS) of infant mortality rates as a function of the minimum distance from waste landfills. The figure shows that infant mortality behavior differs based on bombing intensity. Within a 6 km radius of the landfills, the infant mortality rate increases significantly, particularly in municipalities with high bombing intensity (solid black line). Beyond the 6 km threshold, the infant mortality rate stabilizes and begins to decrease after 8 km. The 6 km radius was chosen for analysis because it marks the point where the trend becomes stationary and ensures a balanced distribution of observations between high- and low-bombed municipalities.

age range of 15 to 65 years old. It collects information on their ethnicity, level of education, and wealth. The attention specifically focuses on infant mortality, which is not a variable directly collected in the survey. However, women are asked if they have had children, if they lived with them, whether any of these children passed away, and, if so, the date of death. Using this information, a reliable measure of infant mortality can be constructed, as it is unlikely that a mother would forget such a significant event. This analysis centers on the infant mortality rate, which is defined by the NICHD as the death of an infant between one day and one year of age. Furthermore, the MICS dataset allows to connect the children and the mothers to their households as well as the level of education and wealth of the mother and the ethnicity. Lastly, the MICS6 survey run in 2019 contains information also on the E.Coli count in water samples, which can be used in the analysis to understand potential mechanisms.

The exact location of five Municipal sanitary landfills, two Regional sanitary landfills, and three waste transfer stations constructed after the conflict has been provided by the Kosovo Environmental Protection Agency. Unfortunately, their list was not exhaustive, and the other landfill information, such as opening year and location, was retrieved by reading a collection of reports, papers, and news articles, as well as double-checking the exact location and coordinates through Google Maps. As a result, a list comprising ten municipal and regional waste landfills, as well as seven industrial and heavy metals landfills, is compiled, giving a total of seventeen waste landfills created after the conflict.

Information on the conflict has been sourced from Human Rights Watch regarding NATO bombing incidents in Kosovo. The data includes the daily frequency of bombings within each municipality but does not provide the exact locations of the strikes. According to this source, the bombing campaign began on March 25th, 1999, with the first strike in Prizren municipality, and ended on June 10th, 1999, with the final bombing in Kamenica. Every municipality experienced at least one bombing event, with the capital municipality of Pristina being the most heavily targeted, receiving 54 bombs over the 77-day conflict. It is important to note that present-day Kosovo consists of 38 municipalities, while at the time of the bombings, there were only 29. Since the exact locations of the bomb strikes are not available, the municipalities created after the war, which were previously part of larger ones, are assumed to have received the same number of bombs as their parent municipalities before the division. Another proxy for violence used in this analysis is the number of casualties per municipality. The data comes from the Kosovo Memory Book, an online resource that provides detailed information, including names, locations of death, ethnicities, and whether the individuals were civilians or members of armed formations[10].

Figure 2.2 illustrates the spatial overlay used to construct the key variables. Municipalities are shaded by 1999 NATO bombing intensity, with dark red denoting above-median exposure. Post-war landfill sites are shown as brown points and MICS survey clusters as blue points. Two distance bands are drawn around each landfill: within 6km (treated area) and 6–10km (comparison area), which together define the 10km sample frame. This structure enables the construction of the bombing-intensity indicator, landfill proximity, and their interaction with the post-opening period for the triple-difference design. The choice of a 6km radius is motivated by the LOWESS pattern in Figure 2.3, which shows that infant mortality remains stationary beyond 6km and declines after 8km. In addition, the 6km cutoff provides a balanced number of observations between high- and low-bombing municipalities.

Following Currie and Walker (2011), municipalities are classified into high- and low-

[10]The casualty data originates from the Kosovo Memory Book, while the bombing data is sourced from Human Rights Watch. These are the only available datasets on the Kosovo conflict. For this analysis that exploits violence, the suggested dataset is usually Armed Conflict Location Event Data (ACLED), but their data collection starts in 2018, while the conflict happened in 1999. The Uppsala Conflict Data Program (UCDP) was also considered for this project, but there is no data on the Kosovo war either.

bombing exposure (above and below the median, respectively), and clusters are categorized as residing within or beyond the 6km landfill radius. To test robustness, alternative specifications employ thresholds of 4km and 8km. The analysis is further restricted to a 10km radius, consistent with Currie and Walker (2011), to minimize bias from unobserved heterogeneity at greater distances.

Figures 2.4a–2.4b plot average infant mortality by event time (years relative to the opening of the nearest landfill) for births within 10 km. Municipalities above the median wartime bombing intensity ('High bomb') are shown with a solid black line; those below the median ('Low bomb') with a grey dashed line. Figure 2.4a reports annual means; Figure 2.4b shows two-year rolling means. Before the opening ($t < 0$), high- and low-bombing municipalities exhibit similar declining trends, supporting the parallel-trends assumption for the Difference-in-Differences and Triple Difference designs. After the opening ($t \geq 0$), trends diverge: mortality keeps falling in low-bombing areas, while it stabilises in high-bombing areas—consistent with the regression estimates of differential post-treatment effects.

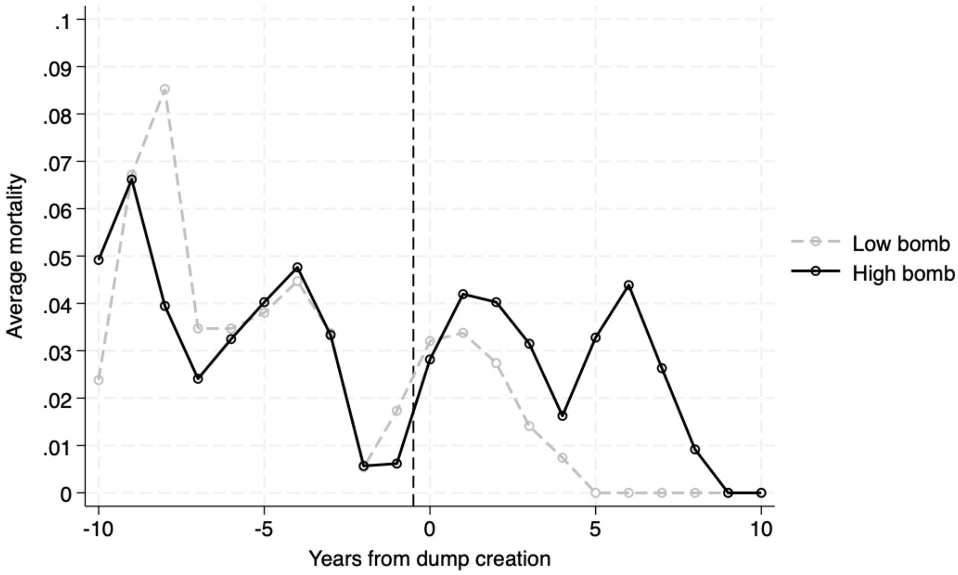
2.4 Descriptive Statistics

The analysis is restricted to births within 10km of a post-war landfill, with children residing within 6km classified as 'close'. This yields 7,160 births in total. To assess group comparability, the sample is further limited to a ±10-year window around each landfill's opening, resulting in 5,373 observations. Table B.2 reports descriptive differences in key outcomes and covariates across three sets of comparisons: (i) high- versus low-bombing municipalities in the full 10km sample (Panel A); (ii) close ($\leq$ 6km) versus far (6–10km) births, pooling bombing intensities (Panel B); and (iii) high- versus low-bombing municipalities within the close group (Panel C).

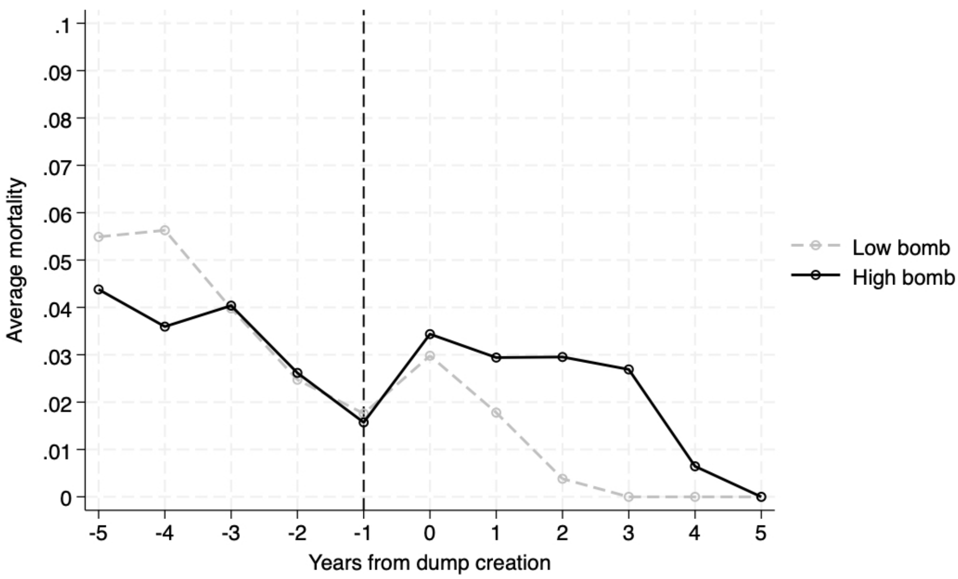
Table B.2 reports differences across three dimensions: bombing intensity (Panel A), landfill proximity (Panel B), and their combination (Panel C).

Panel A compares municipalities by bombing intensity irrespective of landfill proximity. Infant mortality averages are nearly identical across high- and low-bombing areas (0.030 vs. 0.028; difference of 0.001). By contrast, socioeconomic and demographic characteristics vary systematically: high-bombing municipalities are less rural (41.9% vs. 55.6%; difference of –13.8 p.p.) and have a lower share of Albanian households (–4.0 p.p.) and higher shares of Other Ethnicities (+7.0 p.p.). Mothers in high-bombing areas are somewhat more concentrated in secondary education (+3.2 p.p.) and upper-middle wealth

Figure 2.4: Average mortality and rolling window



(a) Panel (a)



(b) Panel (b)

Notes: Panel (a) plots annual mean infant mortality by event time (years relative to the opening of the nearest landfill; vertical dashed line at $t = 0$) for municipalities above ('High bomb', solid black) and below ('Low bomb', grey dashed) the median wartime bombing intensity. Panel (b) shows the corresponding two-year rolling means. The sample includes births within 10 km of a landfill; event time is computed at the child level relative to the opening year of the nearest site. Lines summarize simple averages and are intended for visualization only; statistical inference relies on the DD/DDD estimates reported in the main tables.

(+3.1 p.p.).

Panel B contrasts births close to versus farther from landfills. Infant mortality rates are again very similar (0.028 vs. 0.029; difference of –0.001). In contrast, proximity is associated with substantial differences in composition: close clusters are much less rural (39.6% vs. 67.1%; difference of –27.5 p.p.), have a lower middle-class share (–3.7 p.p.), and display shifts in maternal education (primary –1.3 p.p., secondary –4.9 p.p., upper secondary +3.8 p.p.). The post-opening share of births is also lower in close areas (41.5% vs. 47.2%; difference of –5.7 p.p.), underscoring the importance of flexible time controls.

Panel C combines both dimensions, comparing close births in high- versus low-bombing municipalities. Infant mortality remains similar (0.031 vs. 0.026; difference of 0.005). However, notable differences emerge in group characteristics: high-bombing municipalities with nearby landfills are more rural (54.6% vs. 49.1%), have a lower Albanian share (–4.3 p.p.) and fewer Serbs (–3.5 p.p.), while the share of Other Ethnicities is higher (+7.9 p.p.). Maternal education also differs, with higher rates of primary (+1.8 p.p.) and secondary (+7.3 p.p.) education and lower rates of upper secondary (–3.5 p.p.), alongside a larger proportion of poor households (+4.7 p.p.).

Taken together, these descriptive patterns indicate that bombing intensity and landfill proximity are correlated with socioeconomic and demographic composition, even though infant mortality rates are broadly similar across groups. These imbalances motivate the inclusion of maternal education, wealth, age, rural status, ethnicity, year, and municipality fixed effects in the regression analysis. Complementary evidence on pre-treatment comparability is provided by event-study tests of pre-opening infant mortality trends (Figures 2.4a–2.4b) and by pre-war nightlight balance checks in Appendix B.3.

2.5 Empirical approach

This section outlines the empirical strategy used to estimate the causal effects of bombing on negative externalities, focusing on the impact of bombing intensity and proximity to waste landfills on infant mortality. Although the bombing was initially intended to target strategic military objectives, numerous reports acknowledge significant errors in targeting, resulting in widespread civilian damage. Given that the bombing followed extensive civil attacks across the country, many areas affected by the bombing may have been struck despite not being clear military targets, which introduces some quasi-random variation in exposure. While the bombing itself was not literally random, the key identifying assumption is that it was not systematically correlated with pre-war determinants of infant mortality (such as local economic development, proxied by nightlight intensity). To assess this assumption, pre-existing conditions across municipalities are examined using pre-bombing nightlight data.

2.5.1 Bombing intensity

To ensure the validity of the causal inference, the first step is to check whether bombing exposure is plausibly exogenous with respect to pre-existing socioeconomic conditions. Although strikes targeted military sites, reports suggest imprecision in their location, raising the possibility that civilian areas were affected independently of their socioeconomic characteristics. To investigate this, pre-1998 nightlight intensity — a proxy for local economic development — is compared across municipalities with high versus low bombing intensity. The absence of systematic differences in these pre-treatment measures would support the parallel-trends assumption underlying the Difference-in-Differences (DiD) framework introduced below. The analysis is based on the DMSP-OLS Nighttime Lights data (Goodman et al., 2019), with a focus on the period before and after the 1999 bombing event.

Pre-Treatment Comparison

To assess whether bombing intensity was systematically related to pre-war economic development, a two-sample t-test was conducted on average nightlight intensity during 1995–1997, a widely used proxy for local economic activity. The results indicate no statistically significant difference in mean nightlight luminosity between municipalities later classified as high- versus low-bombing. This finding reinforces the view that variation in bombing intensity was not driven by pre-existing differences in local economic development.

Difference-in-Differences Estimation

To estimate the effect of the bombing on light intensity, an interaction term between a post-1999 indicator and a bombing indicator is constructed. The results indicate that the interaction term is not statistically significant. This suggests that there is no significant change in light intensity in bombed areas relative to non-bombed areas after 1999, after controlling for pre-existing differences. Furthermore, the coefficient for the bombed indicator is significant, indicating that, prior to the bombing, bombed areas had higher light intensity compared to non-bombed areas. However, no significant change after the bombing was found, suggesting that the bombing did not lead to a statistically significant shift in light intensity in the bombed areas. A critical assumption in the DiD methodology is the parallel trends assumption, which requires that the treatment (bombing) and control (non-bombed) groups would have followed similar trends in the absence of the treatment. Given the quasi-random nature of the bombing, it is essential to check this assumption. Pre-treatment trends are examined using lead and lag indicators. Specifically, for 1995–2013, leads and lags are generated to estimate average nightlight intensity as a function of bombing exposure, capturing both anticipatory (lead) and post-event (lag) effects. Additionally, robust standard errors are used to account for potential heteroscedasticity in the data, ensuring more reliable and unbiased estimates.

The regression results show that none of the lead or lag coefficients for the interaction terms are statistically significant for the period before the conflict and up to ten years after the conflict. This suggests that, in the pre-bombing period, there were no significant differences in trends between bombed and non-bombed areas, supporting the parallel trends assumption. To visualize these results, the coefficients for the lead and lag interactions that show the estimated effects of bombing for the years preceding and following the 1999 bombing event are plotted in Figure B.1. The absence of significant differences in pre-treatment trends further supports the validity of the DiD approach; the plot can be found in Appendix B.3.

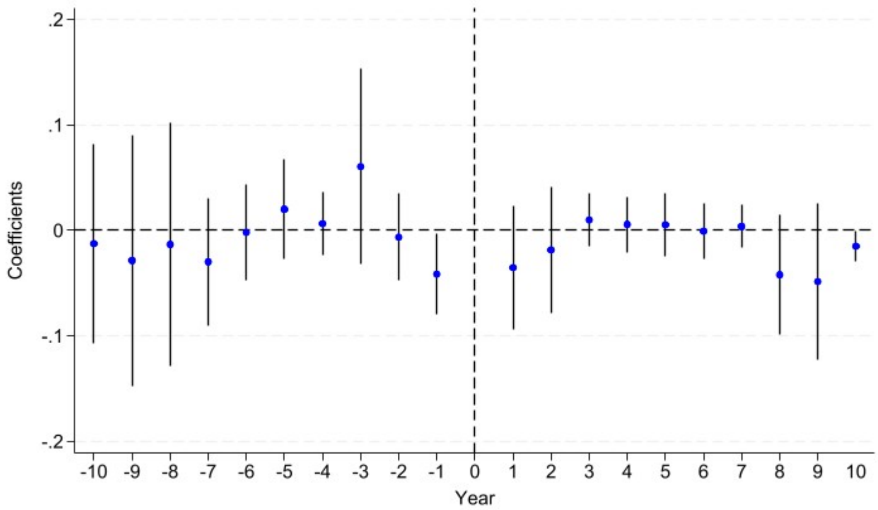
2.5.2 Testing the effect of waste landfill openings on infant mortality

The second layer of analysis examines dynamics around landfill openings using an event–study with relative time indicators (leads and lags), where treatment is defined by proximity to the landfill. In the three years prior to opening, coefficients are small and statistically indistinguishable from zero, providing visual support for parallel trends. When bombing intensity is omitted from the specification, proximity appears negatively associated with infant mortality; this association attenuates to zero once bombing intensity and the full set of covariates are included, consistent with selection into landfill siting or omitted confounding in the uncontrolled regression.

A fixed-effects regression with dummies for each year before and after opening (municipality and year fixed effects; maternal and household controls; clustered standard errors) produces the event–study plotted in Figure 2.5. Estimates remain close to zero both before and after opening, with wide confidence intervals that span zero.

Aggregating post-opening coefficients yields a marginally negative but statistically insignificant average effect: the mean of the first eight post-opening coefficients equals -0.0124 ($p = 0.788$). A comparable average over the last ten pre-opening leads is also insignificant (-0.0178, $p = 0.660$). These findings indicate that landfill openings, considered

Figure 2.5: Time series of infant mortality before and after the opening of a waste landfill



Notes: Points plot coefficients from a regression of infant mortality on indicators for years relative to the landfill opening; the series is normalised at the opening year ($k = 0$, vertical dashed line). Bars denote 95% confidence intervals. Specification includes municipality and year fixed effects and controls for maternal education, wealth, and age; rural status; and ethnicity; standard errors are clustered at the municipality level. Sample includes births within 10 km of post-war landfills, with proximity defined as $\leq$ 6 km ('close') versus 6–10 km ('far'). Pre-opening estimates are indistinguishable from zero, and post-opening estimates are centred near zero, indicating no detectable average effect of landfill opening on infant mortality in this specification.

in isolation, do not produce detectable changes in infant mortality. As shown below, effects emerge when proximity interacts with conflict intensity, consistent with the proposed governance channel.

2.5.3 Main specification

The analysis focuses on children residing within a 6 km radius of waste landfills. The radius choice is supported by distance–response diagnostics: Figure 2.3 shows a flattening of infant mortality around 6 km and a gradual attenuation thereafter. These visual diagnostics are documented in Appendix B.5 and align with the assumption that exposure effects diminish with distance, validating the spatial scope.

The first difference-in-differences (DiD) specification compares infant mortality in municipalities classified as heavily bombed versus those less affected. Treated municipalities are defined as those with above-median bombing intensity; controls are below-median. As an intermediate step, the analysis restricts attention to children born within 6 km of a landfill (the 'close' group) and compares high- versus low-bombing municipalities within this spatially comparable sample.[11] This specification serves as a preliminary DiD check before introducing landfill proximity explicitly in the triple-difference framework. Within this range, those residing in a heavily bombed municipality are considered treated, while those in a less bombed municipality serve as the control group.

To ensure robustness, an event–study framework is employed to validate parallel trends and explore dynamics. Baseline models exclude individual covariates, and standard errors are clustered at the municipality level[12]. All specifications include child birth-year fixed effects, which absorb common time-varying shocks to infant mortality.

A triple Difference in Difference (DDD) model, therefore, allows a clearer identification of the causal effect by jointly considering bombing intensity, landfill proximity, and timing of landfill openings. This approach accounts for variation in bombing exposure across municipalities and the differential exposure of households near versus farther from landfills, providing a more robust basis for causal inference.

Equation 2.1 presents the basic DiD specification. *Postdump* is a binary indicator equal to one if child i was born after the opening of the nearest landfill. The sample is restricted to children residing within 6 km ($i \in \{\text{close}\}$), so the comparison is between heavily and lightly bombed municipalities within the same proximity band. *High Bomb* indicates whether bombing intensity is above the median. Baseline estimates are reported without controls; robustness checks add X_{it}. Controls (X) for maternal education, household wealth, rural residence, maternal age, and the size of the nearest landfill are later introduced in the analysis. While such controls can reduce residual confounding, their inclusion requires stronger assumptions, and in practice, they are usually incorporated via doubly robust estimators that rely only on pre-treatment information. By accounting for these variables, the model seeks to reduce bias and isolate the causal relationship between bombing intensity, landfill proximity, and infant mortality.

$$\begin{aligned}\text{Infant mortality}_{it} &= \beta_0 + \beta_1 \,\text{Postdump}_{it} + \beta_2 \,\text{High bomb}_i \\ &\quad + \beta_3 \,(\text{Postdump}_{it} \times \text{High bomb}_i) + \beta_4 \,X_{it} + \nu_t + \epsilon_{it}\end{aligned} \tag{2.1}$$

To incorporate both close and farther households, the triple difference-in-differences specification in Eq. 2.2 is employed. This design adds an explicit interaction with proxim-

[11] As explained in the data section, the 6 km radius is informed by locally weighted scatterplot smoothing in Figure 2.3: infant mortality is approximately flat beyond 6 km and declines toward 8–10 km.

[12] As a sensitivity check, specifications that add maternal education, household wealth, and rural/urban status (measured at survey time) are reported without causal interpretation; they are used only as precision controls. Because these variables may be affected by conflict and post-war reconstruction, they are not used for identification. Results are materially unchanged

ity (within 6 km versus 6–10 km), allowing the analysis to exploit variation both across bombing intensity and across distance bands. This approach maintains the previous features while introducing an additional interaction term denoted by the variable 'Close'. This variable takes on a value of one if the nearest waste landfill to the child's residence is within 6 km. The analysis frame is restricted to births occurring within 10 km of the nearest landfill, ensuring that treated and comparison observations are drawn from geographically comparable areas.

$$\begin{aligned}\text{Infant mortality}_{it} = {} & \beta_0 + \beta_1\,\text{Postdump}_{it} + \beta_2\,\text{High bomb}_i + \beta_3\,\text{Close}_i \\ & + \beta_4\,(\text{Postdump}_{it} \times \text{High bomb}_i) + \beta_5\,(\text{Postdump}_{it} \times \text{Close}_i) \\ & + \beta_6\,(\text{High bomb}_i \times \text{Close}_i) + \beta_7\,(\text{Postdump}_{it} \times \text{High bomb}_i \times \text{Close}_i) \\ & + \beta_8\,X_{it} + \mu_m + \nu_t + \epsilon_{it}\end{aligned} \tag{2.2}$$

where β_7 is the triple-difference effect (the additional post-opening change in infant mortality for children in heavily bombed municipalities who live close to landfills, net of main effects and pairwise interactions); X_{it} denotes controls; μ_m and ν_t are municipality and year fixed effects; ϵ: Error term, representing unobserved factors affecting infant mortality not accounted for by the model, and standard errors are clustered at the municipality level.

Fixed effects and omitted variable bias A central concern in estimating the effects of landfill proximity in post-conflict settings is omitted variable bias arising from unobserved heterogeneity across municipalities and cohorts. The empirical strategy addresses these concerns through a combination of differencing and fixed effects. First, the DiD and DDD designs difference out all time-invariant characteristics of municipalities and landfill locations—such as geography, baseline health environments, and pre-war infrastructure—that may jointly influence landfill siting and infant mortality. Second, child birth-year fixed effects absorb common time-varying shocks to infant mortality, including nationwide improvements in health care, vaccination campaigns, and post-war reconstruction dynamics.

The role of fixed effects is assessed directly in the data by estimating specifications that vary the inclusion of municipality fixed effects, region fixed effects, and alternative time fixed effects. As shown in Tables 2.3 and 2.4, the coefficient on the triple interaction term (*postdump* $\times$ *high bomb* $\times$ *close*) remains stable in sign, magnitude, and statistical significance across all fixed-effect structures, while several main effects and pairwise interactions are sensitive to the choice of fixed effects. This pattern indicates that fixed effects absorb important sources of confounding variation, but that the core triple-difference estimate is not driven by cross-sectional differences across municipalities or by aggregate time shocks.

Nevertheless, fixed effects cannot eliminate bias from unobserved factors that vary over time at the municipality level and are correlated with both landfill exposure and infant mortality. The results should therefore be interpreted as capturing the interaction between landfill proximity and post-conflict institutional conditions shaped by wartime bombing—operating through persistent institutional and governance weaknesses rather than through direct physical or environmental contamination from the conflict itself.

Triple Difference Design and Identification Assumptions The specification in Equation 2.2 corresponds to a triple-difference (DDD) framework, as outlined by Baker et al. (2025). This design compares changes in infant mortality over time between municipalities differing in bombing intensity (above versus below the median) and landfill proximity (within 6 km versus beyond 6 km). The post-treatment period is defined relative to the timing of landfill construction, which varies across municipalities.

Recent methodological developments highlight that in settings with staggered adoption, using a simple DDD as the difference of two DiD models (or relying on TWFE) is problematic—especially when covariates are included. The approach neglects the nuanced timing and interactions caused by staggered treatment, leading to potential bias. Instead, the relevant estimand is the conditional Average Treatment Effect on the Treated (ATT) for the subgroup of interest—in this case, children in heavily bombed municipalities located within 6 km of a landfill after its opening. The Callaway and Sant'Anna estimator directly targets this ATT, ensuring consistent interpretation despite variation in treatment timing and the inclusion of covariates. This approach also mitigates the biases associated with conventional two-way fixed effects estimators in triple-difference settings (Strezhnev, 2023).

Identification of this DDD estimand relies on a modified parallel trends assumption: in the absence of treatment, infant mortality in bombed and non-bombed municipalities would have followed comparable trajectories both near and far from landfills, conditional on observed covariates. This allows for partition-specific and group-specific level differences but requires that underlying trends remain parallel absent treatment. Event-study estimates presented in Figure 2.6 provide supporting evidence for this assumption in the pre-treatment period.

To test the robustness of the findings, additional analyses were conducted with alternative specifications. These included variations in the distance threshold, using radii both smaller (4 km) and larger (8 km) than 6 km, to ensure that results were not overly sensitive to distance choices, please see Section 2.8.

2.5.4 Dealing with the Staggered treatment timing

The staggered Callaway and Sant'Anna (2021) estimator is used to account for differences in the timing of landfill openings across municipalities during the post-conflict period. This approach exploits the staggered nature of treatment implementation, where treatment effects are estimated by comparing units treated at different times to those not yet treated. The not-yet-treated group functions as a dynamic control, including municipalities that will receive treatment in the future but are untreated in the current period.

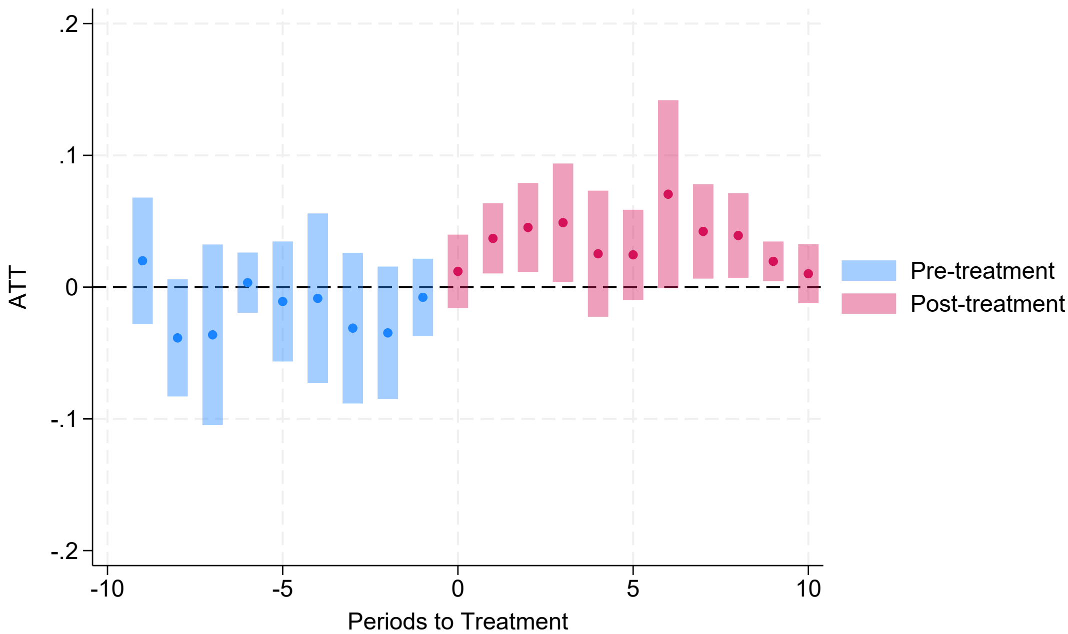
To do so, the following steps have been followed. First, variables are created to define the timing of treatment and differentiate between treated, not-yet-treated, and never-treated units. The variable that defines the first treatment captures the year of landfill construction for municipalities classified as treated, and then further conditions, such as high bombing intensity, are incorporated to refine the treatment definition. These variables allow the model to dynamically assign control and treatment statuses based on the timing and characteristics of treatment.

Estimation follows the staggered DiD framework of Callaway and Sant'Anna (2021), extended to allow differential effects by bombing intensity and distance. Treatment is birth after the opening of the nearest landfill. For each first-treated cohort and year, treated observations are compared with not-yet-treated observations to identify ATT effects. Results are reported as event-time profiles and as an average post-opening effect using the doubly robust inverse-probability-weighted implementation; standard errors are clustered at the municipality level. Implementation details appear in Appendix B.5.

Figure 2.6 reports cohort-aggregated event-time ATT estimates for the main group: children residing within 6 km of a landfill in municipalities with above-median bombing intensity. Event time is defined at the child level relative to the opening year of the nearest landfill: 0 denotes births in the opening year; $+k$ denotes births k years after opening; and $-k$ denotes births k years before opening. Table 2.2 presents the corresponding average post-opening ATT for the same group. Placebo estimates for children farther than 6 km and/or in low-bombing municipalities appear in Appendix Table B.8 and associated figures; these serve as falsification checks and are not used to form the counterfactual for the main estimate.

For the main group (within 6 km and above-median bombing), the average post-opening ATT equals 0.034 (s.e. 0.010), corresponding to roughly 3.4 additional infant deaths per 1,000 live births after opening. Pre-treatment coefficients are statistically indistinguishable from zero, while post-treatment coefficients are positive. See Figure 2.6 (event-time profile) and Table 2.2 (average post-opening effect); placebo strata appear in Appendix Table B.8.

Figure 2.6: Impact of landfill openings on infant mortality: event-study estimates (children within 6 km in heavily bombed municipalities)



Notes: Points plot cohort-aggregated ATT estimates; bars denote 95% confidence intervals. The horizontal axis measures event time relative to the opening year of the nearest landfill (0 = birth in the opening year; $+k = k$ years after; $-k = k$ years before). Estimates use the staggered DiD estimator of Callaway and Sant'Anna (2021) with doubly robust inverse-probability weighting; standard errors are clustered at the municipality level.

2.5.5 Falsification Checks

To assess whether the estimated effects are specific to the relevant treatment group, alternative exposure groups are examined. Figure 2.6 presents the main specification for children born within 6 km of a landfill in heavily bombed municipalities, where pre-treatment coefficients are flat and post-treatment effects are positive and significant.

Table 2.2: Average Treatment Effect on Infant Mortality: Children within 6 km and High bombing (Main Specification)

ATT	**Pre-treatment avg.**	**Post-treatment avg.**
0.034***	-0.016	0.034***
(0.010)	(0.014)	(0.010)

Notes: Estimates are obtained using the Callaway and Sant'Anna (2021) staggered DiD estimator with doubly robust inverse probability weighting. Treatment is defined as being born within 6 km of a landfill in a municipality with above-median wartime bombing intensity. Standard errors clustered at the municipality level are reported in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 2.7 reports results for children born in heavily bombed municipalities but residing farther than 6 km from a landfill. In this group, event-time coefficients fluctuate around zero; a few post-opening periods deviate from zero, but there is no persistent post-opening shift, and magnitudes are smaller than in the main treated group. Figure 2.8 shows sloped pre-treatment coefficients for the non-bombed, >6 km placebo group, indicating distinct secular trends; accordingly, this group is used only as a falsification check and not to form the counterfactual for the main estimate. Together, these results are consistent with effects concentrating where proximity and high bombing overlap, though smaller effects elsewhere cannot be ruled out.

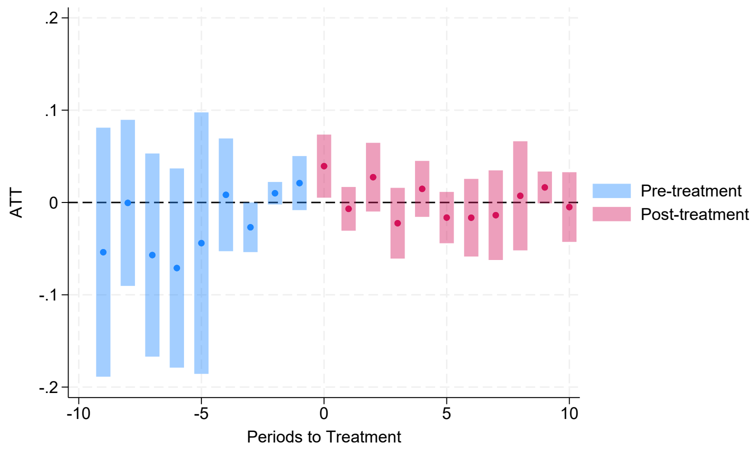
For completeness, Appendix Figure B.6 shows results for children living within 6 km of landfills in municipalities with low bombing intensity. While post-treatment coefficients are positive, significant differences are also observed in the pre-treatment period, indicating violations of the parallel trends assumption. This reinforces that the low-bomb group is not a valid counterfactual for the main treated units and underscores the importance of conditioning treatment on both landfill proximity and bombing intensity.

2.6 Results

This section reports the main estimates quantifying how bombing intensity, landfill proximity, and post–opening exposure jointly relate to infant mortality.

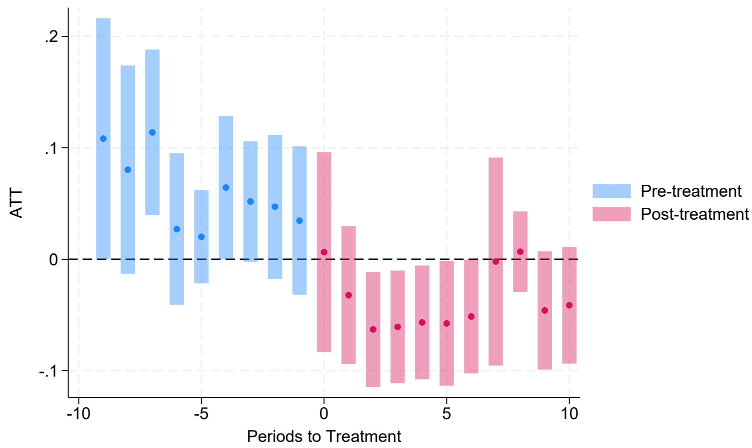
Panel A of Table 2.3 shows that the triple interaction— *Postdump* × *High bomb* × *Close*—is positive and statistically significant across specifications, with point estimates between 0.033 and 0.044 (p-values 0.01 to < 0.001). In levels, this corresponds to 3.3–4.4 percentage points, or roughly 33–44 additional deaths per 1,000 live births. Estimates are similar in covariate-rich specifications (adding maternal education and wealth, municipality, birth-year, and landfill-opening fixed effects). These patterns indicate that, in

Figure 2.7: Event-study estimates: Children in heavily bombed municipalities, >6 km from a landfill



Note: Estimates obtained using the Callaway and Sant'Anna (2021) staggered DiD estimator with doubly robust inverse probability weighting. Treatment is defined as landfill openings in municipalities with above-median wartime bombing intensity. The sample is restricted to children residing more than 6 km from a landfill. Standard errors clustered at the municipality level.

Figure 2.8: Event-study estimates: Children in non-bombed municipalities, >6 km from a landfill



Note: Estimates obtained using the Callaway and Sant'Anna (2021) staggered DiD estimator with doubly robust inverse probability weighting. Treatment is defined as landfill openings. The sample is restricted to children residing more than 6 km from a landfill in municipalities with below-median wartime bombing intensity. Standard errors clustered at the municipality level.

heavily bombed municipalities, proximity to post-war landfills is associated with higher infant mortality.

Panel B (DiD within the 0–6 km sample) compares high- and low-bombing municipalities without an explicit proximity interaction. The *Postdump* $\times$ *High bomb* coefficient ranges from 0.017 (p = 0.12) without covariates to 0.032 (p = 0.004) with municipality and birth-year fixed effects. Main effects (*Postdump*, *High bomb*) are not causally interpretable in this design; the parameter of interest is the triple interaction in Panel A. The smaller and less precise DiD estimates underscore that excess mortality risk concentrates among children living near landfills in heavily bombed municipalities.

Table 2.4 isolates landfills externally financed by the European Agency for Reconstruction (EAR, 2003–2007). Triple-interaction estimates range from 0.041 to 0.046 (4.1–4.6 percentage points; 41–46 per 1,000) and are stable across specifications with opening-year, birth-year, and landfill fixed effects. Panel B of Table 2.4 reports consistent DiD interactions (0.031–0.039, $p < 0.01$), reinforcing that infant mortality rose when high bombing coincided with landfill exposure, even absent an explicit proximity term. Specifications that include birth-year fixed effects yield similar triple-interaction estimates in the full sample ($approx 0.039$, $p = 0.002$) and the EAR-financed subsample (≈0.044, $p = 0.001$), with near-zero estimates for post-EAR landfills (Appendix Table B.4).

Figure 2.6 presents event-study estimates for children within 6 km in heavily bombed municipalities. Pre-treatment coefficients are small and statistically indistinguishable from zero, supporting parallel trends. After landfill openings, the coefficients turn positive and remain so, mirroring the DDD results.

Overall, the DDD estimates imply increases of 3.3–4.6 percentage points (roughly 33–46 per 1,000) in infant mortality for children born near landfills in heavily bombed municipalities after landfill construction. These magnitudes are large relative to a baseline infant mortality rate of about 3%, indicating substantial health risks at the intersection of conflict intensity and post-conflict environmental infrastructure.

For comparison, Appendix Table B.9 replicates the same specifications for landfills built without EAR financing. Across columns, the triple–interaction estimates are small (≈0.02) and statistically indistinguishable from zero, and the corresponding DiD interactions are likewise near zero. This contrast isolates the association to EAR–financed sites and is consistent with heterogeneity in siting and implementation across landfill regimes.

Table 2.3: Difference in Difference results

Panel A: D×D×D						
Infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.018**	0.000	-0.016	-0.012	-0.016	-0.012
	(0.008)	(0.014)	(0.011)	(0.019)	(0.011)	(0.016)
High bomb	-0.032***	-0.018	-0.026***	-0.017**	-0.020*	-0.008
	(0.008)	(0.021)	(0.008)	(0.007)	(0.011)	(0.008)
Close	-0.012	-0.013*	-0.009	-0.008	-0.002	0.003
	(0.008)	(0.007)	(0.010)	(0.009)	(0.010)	(0.009)
Postdump × High bomb	-0.022	-0.018*	-0.019	-0.018	-0.019	-0.017
	(0.013)	(0.010)	(0.014)	(0.012)	(0.014)	(0.012)
Postdump × Close	-0.005	-0.007	-0.007	-0.009	-0.006	-0.008
	(0.008)	(0.009)	(0.011)	(0.012)	(0.011)	(0.012)
High bomb × Close	0.003	0.002	-0.003	-0.002	-0.013	-0.016**
	(0.011)	(0.011)	(0.008)	(0.008)	(0.008)	(0.008)
Postdump × High bomb × Close	0.037***	0.033***	0.044***	0.043***	0.044***	0.043***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)
Dump Opening year FE	X		X		X	
Child Birth Year FE		X		X		X
Municipality FE	X	X	X	X		
Region FE					X	X
Covariates		X		X		X
N	5373	5373	4466	4466	4466	4466
N cluster	26	26	20	20	20	20
Panel B: D×D						
Infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.022***	-0.011	-0.024***	0.013	-0.024***	0.003
	(0.004)	(0.013)	(0.005)	(0.017)	(0.005)	(0.017)
High bomb	-0.025***	-0.027***	-0.026**	0.024	-0.021**	-0.015
	(0.005)	(0.005)	(0.009)	(0.016)	(0.009)	(0.017)
Postdump × High bomb	0.017	0.019*	0.026**	0.032***	0.026**	0.033***
	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)
Dump Opening year FE	X		X		X	
Child Birth Year FE		X		X		X
Municipality FE	X	X	X	X		
Region FE					X	X
Covariates		X		X		X
N	2945	2945	2366	2365	2366	2365
N cluster	15	10	10	10	10	10

Notes: The primary estimand is the triple interaction (*postdump* × *high bomb* × *close*); main effects are not directly interpretable as causal impacts in this specification. The table displays coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, robust standard errors clustered at the municipality level are reported in parentheses. *Panel A* represents the triple difference in difference, while the simple difference in difference is reported in *Panel B*. *Panel B* sample is obtained by subsampling the children living within 6 km of a waste landfill. Furthermore, it also includes weights to account for children born in places with more than one dump within 6 km. The table shows the results for infant mortality where (1) represents the difference in difference without covariates but with the municipality and opening year fixed effects: (2) is the same as the previous one with birth year fixed effects instead of opening year: (3) same as (1) but with covariates; (4)same as (2) but with covariates; (5) instead of the municipality the region is used as FE and also opening year is included; (6) is the same as (5) but with birth year fixed effect.

Table 2.4: Difference in Difference results for waste landfills externally financed (EAR between 2003 and 2007)

Panel A: D×D×D						
Infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.021***	-0.005	-0.021**	-0.018	-0.021**	-0.017
	(0.007)	(0.013)	(0.009)	(0.016)	(0.008)	(0.014)
High bomb	-0.026**	-0.014	-0.023**	-0.017**	-0.018	-0.007
	(0.010)	(0.021)	(0.010)	(0.008)	(0.013)	(0.008)
Close EAR	-0.012	-0.015*	-0.011	-0.012	-0.004	0.001
	(0.009)	(0.007)	(0.011)	(0.010)	(0.010)	(0.007)
Postdump × High bomb	-0.016	-0.012	-0.013	-0.011	-0.012	-0.010
	(0.009)	(0.008)	(0.010)	(0.010)	(0.010)	(0.010)
Postdump × Close EAR	0.001	0.000	0.000	0.000	0.001	0.001
	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)	(0.010)
High bomb × Close EAR	-0.016	-0.016	-0.013	-0.011	-0.022	-0.024***
	(0.012)	(0.014)	(0.014)	(0.012)	(0.013)	(0.008)
Postdump × High bomb × Close EAR	0.046***	0.044***	0.044***	0.042***	0.044***	0.041***
	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Dump Opening year FE	X		X		X	
Child Birth Year FE		X		X		X
Municipality FE	X	X	X	X		
Region FE					X	X
Covariates		X		X		X
N	3249	3118	3118	3118	3118	3118
N cluster	17	15	15	15	15	15

Panel B: D×D						
infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.020***	-0.005	-0.021**	-0.009	-0.021**	0.009
	(0.005)	(0.016)	(0.006)	(0.012)	(0.006)	(0.012)
High bomb	-0.032***	-0.035***	-0.033***	0.012	-0.027***	0.10
	(0.002)	(0.004)	(0.008)	(0.009)	(0.006)	(0.008)
Postdump× High bomb	0.031***	0.037***	0.033***	0.039***	0.033***	0.039***
	(0.006)	(0.009)	(0.006)	(0.009)	(0.006)	(0.009)
Dump Opening year FE	X		X		X	
Child Birth Year FE		X		X		X
Municipality FE	X	X	X	X		
Region FE					X	X
Covariates		X		X		X
N	1800	1800	1780	1799	1780	1779
N cluster	8	8	6	6	6	6

Notes: The table displays coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, robust standard errors clustered at the municipality level are reported in parentheses. *Panel A* represents the triple difference in difference, while the simple difference in difference is reported in *Panel B*. *Panel B* sample is obtained by subsampling the children living within 6 km of a waste landfill. Furthermore, it also includes weights to account for children born in places with more than one dump within 6 km. The table shows the results for infant mortality where (1) represents the difference in difference with covariates but with the municipality and opening year fixed effects: (2) is the same as the previous one with birth year fixed effects instead of opening year: (3) same as (1) but with covariates; (4)same as (2) but with covariates; (5) instead of the municipality the region is used as FE and also opening year is included; (6) is the same as (5) but with birth year fixed effect.

As a robustness check, replacing bombing intensity with a high-casualties proxy yields similar positive DDD effects, see Table 2.7.

2.7 Mechanisms

2.7.1 Water contamination

According to the World Bank (2018) report 'Water Security Outlook for Kosovo', most of the population of Kosovo has access to safe water and sanitation services, with 92% of the population having access to piped water service within their dwellings, and only 8.42% of the population has no access to a water service network. Municipalities in Kosovo are responsible for managing local water resources, including natural springs, public springs, wells, and ditches, to ensure an adequate water supply at the local level. Additionally, under the Law on Local Self-Government, municipalities are tasked with providing public water supply services. Kosovo's major river basins are widely recognized as being moderately to heavily polluted. As economic development progresses, water use and pollution are expected to increase. Some water sources are already reported as polluted or at risk of organic contamination. The quality of raw surface water in Kosovo is generally considered moderate, as much of the water is drawn from artificial reservoirs.

Within this context, water contamination cannot be excluded as a potential risk factor. One possible mechanism by which water contamination may impact child and maternal health is the presence of Escherichia Coli (E.Coli) bacteria, either at the water source or within households.

The landfill environment itself appears to promote the proliferation and spread of antibiotic-resistance genes in E.Coli and other microbes. Hui et al. (2023) and Threedeach et al. (2012) confirm that the aging of landfilled waste is associated with increased levels of antibiotic-resistance genes. Considering this hazard, seems reasonable to analyze the count of E.Coli present in the water. In fact, during their interviews, UNICEF Kosovo Programme (2020) have asked for a glass of water in order to collect water samples in most of the households they have visited. Furthermore, they have also requested to see the source of the water and have collected a sample from there as well. UNICEF (2017) provides levels of riskiness of E.Coli in water. It is considered low when the E.Coli count is less than 1 E.Coli per 100 mL; moderate when it ranges from 1 to 10 E.Coli per 100 mL; high when it ranges between 11 to 100 E.Coli per 100 mL, and very high risk if it is higher than 101 E.Coli per 100 mL. Contamination may occur between the source and the household during transport, handling, and storage.

$$
\begin{aligned}
Y_i = \beta_0 &+ \beta_1 \, \text{High bomb}_i + \beta_2 \, \text{Close}_i + \beta_3 \, (\text{High bomb}_i \times \text{Close}_i) \\
&+ \beta_4 \, \text{Piped water}_i + \beta_5 \, \text{Public tap water}_i + \beta_6 \, \text{Well water}_i + \beta_7 \, \text{Spring water}_i \\
&+ X_i'\theta + \gamma_{\ell(i)} + \varepsilon_i.
\end{aligned}
\tag{2.3}
$$

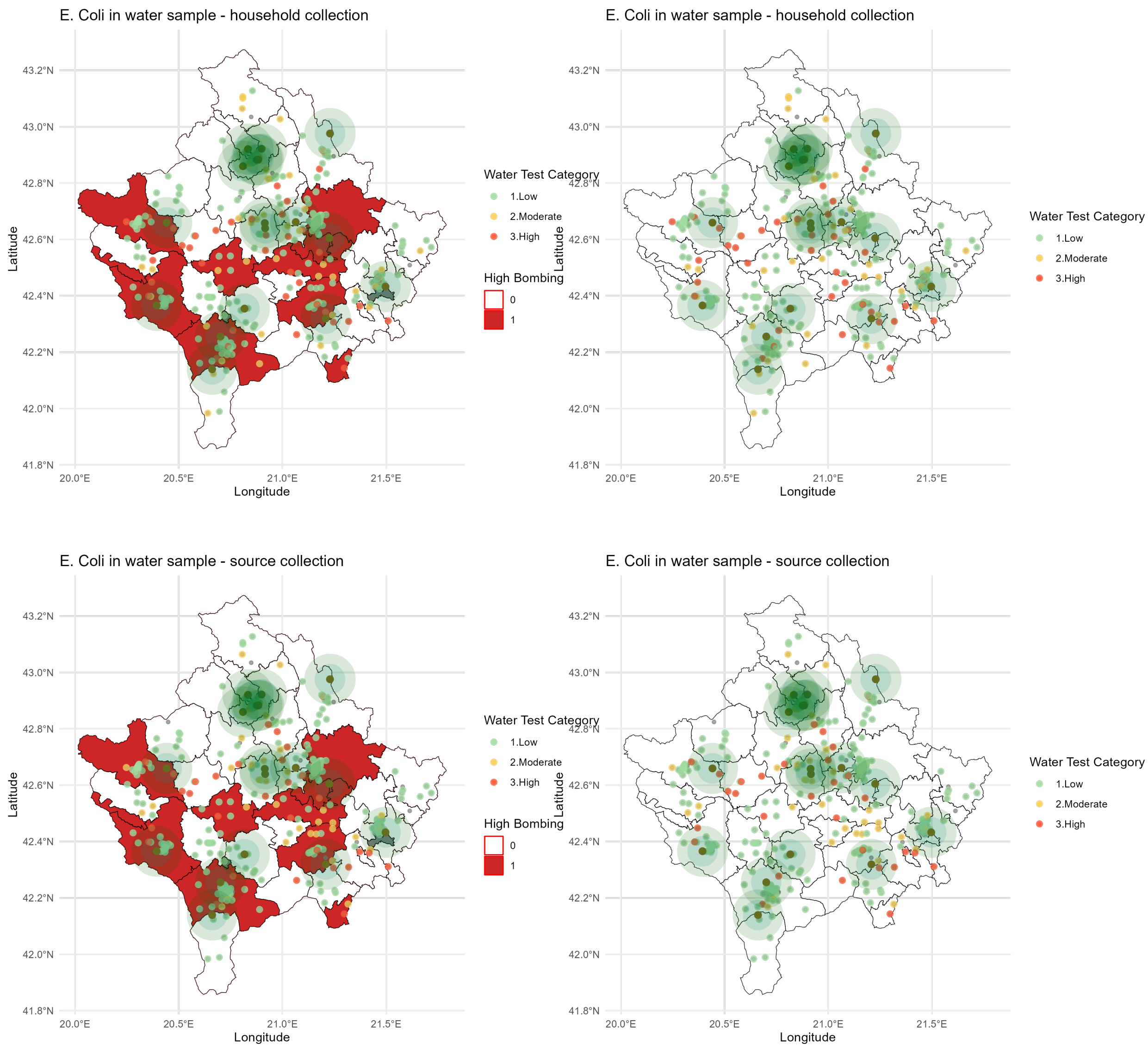
Two outcome definitions are used: (i) $Y_i = \textit{water test}_i$ (the E.Coli count, estimated by OLS) and (ii) $Y_i = \log(\textit{water test}_i)$. In addition, a Poisson GLM with a log link is estimated; coefficients are reported on the log scale. For interpretability, incidence rate ratios (IRR) can be obtained as $\exp(\hat{\beta})$. Water-source indicators (piped, public tap, well, spring) enter with bottled water as the omitted category. $\gamma_{\ell(i)}$ denotes nearest–landfill fixed effects, and X_i includes maternal education, household wealth, ethnicity, and a rural indicator. Standard errors are clustered at the landfill level. All specifications are estimated separately for samples collected at the household and at the source.

Table 2.5 reports OLS and Poisson regressions for E.Coli counts in water collected at the source. Results are shown for the full sample ('All') and, separately, for strata defined by funding of the nearest landfill: External funding denotes landfills financed by the European Agency for Reconstruction (EAR, 2003–2007); Internal funding denotes landfills constructed later without external aid. The dependent variable is the E.Coli count; covariates include above-median bombing intensity, proximity to the nearest landfill ($\leq$6km), water-source type indicators, and a rural indicator. Poisson estimates (on the log scale) are interpreted as incidence rate ratios and are well suited to count outcomes; OLS coefficients are reported for comparison. Directions are consistent across approaches. Importantly, both OLS and Poisson show consistent trends in the direction of effects, particularly for 'high bomb' and 'close', reinforcing the validity of the findings. An effect is expected if conflict intensity strains local governance—diverting resources from monitoring and maintenance and weakening enforcement—so that proximity to landfills is more likely to coincide with compromised water quality

Focusing on the comparison between the overall sample and the externally funded sources, the interaction term between bombing intensity and proximity to the waste landfill exhibits notable differences. In the OLS model for the overall sample, the interaction term is positive but insignificant, whereas in the Poisson model, the coefficient is 0.880 ($p<0.001$), implying an incidence rate ratio (IRR) of $\exp(0.880) \approx 2.4$. This means that, conditional on controls, expected E.Coli counts are about 2.4 times higher in municipalities that are both heavily bombed and close to a landfill.

For externally funded sources, the interaction term is large and precisely estimated in both the OLS (coef. 5.043, $p<0.01$) and Poisson (coef. 2.171, $p<0.001$) models. The Poisson estimate corresponds to an IRR of $\exp(2.171) \approx 8.8$, indicating that contamination levels are nearly nine times higher when bombing intensity and landfill proximity coincide. This finding may reflect the persistent challenges of mitigating contamination in severely bombed areas, even with improved infrastructure provided through external interventions.

Figure 2.9: E. coli contamination in water samples by bombing intensity and proximity to landfills



Notes: Each panel maps *E. coli* contamination levels in water samples collected during the MICS survey. Top row: household water samples. Bottom row: water collected at the source. Left column: municipalities shaded by NATO bombing intensity (red = above median). Right column: same data without bombing overlay. Circles represent survey clusters; size reflects the number of samples and color indicates contamination risk (green = low, orange = moderate, red = high, very high >100). Categories follow UNICEF guidelines: low (0–1 *E. coli* per 100 mL), moderate (1–10), high (11–100), very high (>100).

Table 2.5: E.Coli count of water collected at the source

	All		External funding		Internal funding	
E.Coli count	OLS	Poisson	OLS	Poisson	OLS	Poisson
High bomb	-0.007	0.074	-2.707*	-0.671***	5.353	2.482***
	(1.500)	(0.087)	(1.484)	(0.103)	(3.549)	(0.140)
Close	-0.302	-0.045	-1.730	0.275**	1.438	1.984***
	(1.426)	(0.088)	(1.564)	(0.115)	(3.434)	(0.183)
High bomb × Close	1.952	0.880***	5.043***	2.171***	5.017	-2.074***
	(1.885)	(0.118)	(1.949)	(0.142)	(6.416)	(0.243)
Piped water	-0.143	3.550***	0.963	3.699***	0.542	14.344
	(1.366)	(1.001)	(1.579)	(1.002)	(3.372)	(673.462)
Public tap water	-2.738	2.192**	-7.346*	-0.340	-6.825	11.747
	(3.305)	(1.048)	(3.996)	(1.429)	(5.710)	(673.462)
Well water	5.459***	5.377***	3.707*	4.878***	10.334**	16.531
	(1.819)	(1.002)	(2.209)	(1.005)	(4.443)	(673.462)
Spring water	6.778***	4.861***	26.622***	6.023***	4.214	15.866
	(2.383)	(1.003)	(3.611)	(1.007)	(4.822)	(673.462)
N	881	881	580	580	330	330
Mean	3.220		2.465		3.914	

Notes: The table reports coefficients with standard errors in parentheses. Standard errors in parentheses are heteroskedasticity-robust and clustered at the landfill level. All water-source indicators are measured relative to the omitted category (bottled water). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. For OLS, coefficients reflect average changes in the E.Coli count associated with each variable. For Poisson models, coefficients are expressed on the log scale, and incidence rate ratios can be obtained as $\exp(\hat{\beta})$. For example, a coefficient of 0.880 corresponds to an IRR of approximately 2.4, meaning that expected contamination counts are 2.4 times higher in the treated group. All water-source indicators are measured relative to the omitted category, bottled water. Reported p-values are based on robust standard errors clustered at the landfill level.

2.7.2 Corruption

Corruption is examined as a potential mechanism linking bombing intensity, landfill construction, and adverse health outcomes. Landfill projects were often financed through aid flows, notably by the European Agency for Reconstruction (EAR), but the effectiveness of these investments may have been undermined if resources were diverted through corrupt practices. In such cases, infrastructure was formally delivered, but compromised implementation could reduce the expected health benefits.

$$\text{Corruption}_{jmt} = \beta_0 + \beta_1 \text{High bomb}_m + \beta_2 \text{Landfill}_m + \beta_3 (\text{High bomb}_m \times \text{Landfill}_m) \\ + X_{it} + \gamma_t + \epsilon_{jmt} \tag{2.4}$$

Equation 2.4 models perceptions of corruption across institution j in municipality m and year t. The dependent variable equals one if the respondent perceives moderate or severe corruption and zero otherwise. Key explanatory variables are a dummy for municipalities exposed to high bombing intensity and an indicator for the presence of a landfill. Their interaction captures whether corruption perceptions differ systematically in municipalities with both heavy wartime bombing and post-war landfill construction. The vector X_{it} includes individual-level controls (ethnicity, employment, age, education, marital status, and urban residence).

Panel A of Table 2.6 indicates that in local government the main effect of high bombing is negative (–0.030 to –0.026, $p<0.05$), whereas for central government it is positive (0.022 to 0.031, $p<0.05$). This pattern is consistent with Chapter 3 when institutional tiers and the type of conflict exposure are distinguished: war-era intensity is associated with higher central corruption perceptions, whereas effects on local institutions are weak or negative (as in the armed-forces casualty specifications). Notably, the interaction between bombing intensity and landfill presence (columns 1 and 2) is positive for local government (≈0.072–0.074, $p<0.01$) and negative for central government, IOs, EULEX, and police (all $p<0.05$). Thus, in municipalities that were heavily bombed and host a landfill, attribution shifts toward local actors and away from central/IOs. The negative interaction between bombing intensity and landfill presence for central government corruption (columns 3 and 4) suggests that municipalities with both high bombing intensity and landfills funded by international organizations may perceive lower levels of corruption. Given the negative interaction at the central/IO level and positive interaction locally, a more straightforward interpretation is that post-war, project-proximate frictions make corruption more salient locally, while dampening blame toward central/IO institutions in those same municipalities. The negative interaction for central/IO outcomes is consistent with reduced attribution to those institutions in municipalities combining high bombing and landfills. While this could arise from oversight, contracting structure, or expectations, the data here identify perceptions, not mechanisms. In areas with foreign-funded landfills, these international bodies may have implemented measures aimed at curbing corruption, which could explain why municipalities with both high bombing intensity and foreign-backed infrastructure projects might perceive corruption as lower, despite the complex dynamics at play. Thus, the results could be interpreted through two lenses: (1) the normalization of corruption in areas with high bombing intensity, where the local population might become more tolerant of corrupt practices, and (2) the mitigating effect of international organizations and their influence on reducing corruption in areas with foreign aid. This dual interpretation highlights the complexity of the relationship between conflict, corruption, and foreign aid, suggesting that the perception of corruption is influenced by both historical exposure to violence and the presence of international institutions aimed at promoting transparency and reducing corruption.

Panel B reveals interesting contrasts between externally funded and internally funded landfills in municipalities affected by high bombing intensity, shedding light on how the source of funding influences the perception of corruption. Focusing on the perception of corruption in local government (Column 2), the interaction is positive for both funding types and larger for internally funded landfills than for externally funded. Thus, in high-bomb municipalities, either funding source raises local corruption perceptions, with a stronger amplification for internally funded projects. This pattern is consistent with Panel A, but Panel B makes clear that the stronger local effect is driven primarily by the internally funded landfills, which amplify corruption perceptions more than externally funded projects. The increase in perceived corruption in these areas could stem from a variety of factors, such as skepticism about the effectiveness or motives behind international aid, or a mismatch between local expectations and the outcomes of externally funded projects. Although no significant difference is displayed in the perception of corruption for the central government (Column 4), some difference is displayed in Column 6 where the corruption in International Organizations is perceived as negative by both people who live in municipality internally and externally funded waste landfills, but the only significant one is the related to the internal funded with -0.081 (vs a -0.041 of the externally funded).

Note that main effects and interactions should be read jointly. In high-bomb municipalities with a landfill, the marginal effect on local corruption combines the negative main effect of In high-bomb municipalities with a landfill, the marginal effect on local corruption combines the negative main effect of Landfill with the positive high bomb $\times$ Landfill interaction; the resulting net effect is an empirical matter, summarized in the table by the interaction's sign and significance.

Table 2.6: Corruption

	Local government		Central government		International Org (IOs)		EULEX		Police	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: heavily bombed										
High bomb	-0.030***	-0.026**	0.022**	0.031***	0.107***	0.110***	0.076***	0.072***	0.075***	0.072***
	(0.010)	(0.010)	(0.009)	(0.009)	(0.015)	(0.015)	(0.013)	(0.013)	(0.011)	(0.011)
Landfill	-0.053***	-0.045***	-0.016**	-0.003	-0.074***	-0.066***	-0.007	-0.008	0.018*	0.015
	(0.009)	(0.009)	(0.008)	(0.008)	(0.013)	(0.013)	(0.012)	(0.012)	(0.009)	(0.010)
High bomb×Landfill	0.072***	0.074***	-0.031**	-0.032**	-0.067***	-0.068***	-0.075***	-0.068***	-0.046***	-0.045***
	(0.015)	(0.015)	(0.013)	(0.013)	(0.022)	(0.022)	(0.020)	(0.020)	(0.016)	(0.016)
N	18967	18967	18108	18108	9996	9996	10937	10937	18498	18498
Controls	X	X	X	X	X	X	X	X	X	X
Year FE		X		X		X		X		X
Panel B: extern vs intern fund										
High bomb	-0.030***	-0.025**	0.022**	0.031***	0.110***	0.113***	0.078***	0.074***	0.075***	0.072***
	(0.010)	(0.010)	(0.009)	(0.009)	(0.015)	(0.015)	(0.013)	(0.013)	(0.011)	(0.011)
Landfill (external fund)	-0.028***	-0.021**	-0.013	-0.001	-0.061***	-0.052***	-0.008	-0.008	0.004	0.001
	(0.010)	(0.010)	(0.009)	(0.009)	(0.016)	(0.016)	(0.014)	(0.014)	(0.011)	(0.011)
Landfill (internal fund)	-0.095***	-0.085***	-0.022*	-0.005	-0.099***	-0.091***	-0.009	-0.009	0.042***	0.037***
	(0.013)	(0.013)	(0.012)	(0.012)	(0.020)	(0.020)	(0.018)	(0.018)	(0.014)	(0.014)
High bomb×Landfill (external fund)	0.045***	0.052***	-0.042***	-0.032**	-0.045*	-0.041	-0.022	-0.013	-0.034*	-0.034*
	(0.017)	(0.017)	(0.016)	(0.015)	(0.027)	(0.027)	(0.024)	(0.024)	(0.018)	(0.018)
High bomb×Landfill (internal fund)	0.115***	0.108***	-0.016	-0.031*	-0.076***	-0.081***	-0.122***	-0.117***	-0.067***	-0.063***
	(0.019)	(0.019)	(0.018)	(0.018)	(0.029)	(0.028)	(0.026)	(0.026)	(0.021)	(0.021)
N	18967	18967	18108	18108	9996	9996	10937	10937	18498	18498
Controls	X	X	X	X	X	X	X	X	X	X
Year FE		X		X		X		X		X
outcome mean	0.698	0.698	0.793	0.793	0.603	0.603	0.696	0.696	0.584	0.584

Standard errors in parentheses. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$. All regressions include individual-level controls (ethnicity, employment, age, schooling years, marital status, urban/rural).

2.8 Robutsness checks

2.8.1 Memory book approach - high casualties

As a robustness check, bombing intensity is replaced with an alternative exposure measure: the municipality-level total number of conflict casualties (civilian and armed forces). The outcome, sample (0–10 km with 0–6 km 'close'), event-time definition, fixed effects, and estimator are unchanged from Equations 2.1 and 2.2; only the treatment dimension is altered by substituting a high-casualties indicator for high bombing (median split). High casualties equal one if a municipality's casualty count exceeds the sample median. An effect is expected if elevated casualties proxy conflict intensity and governance strain—diverting resources from monitoring and maintenance, weakening enforcement, and slowing remediation—so that proximity to landfills interacts with lower operational quality.

Panel A of Table 2.7 reports the DDD estimates. In the EAR-funded subsample (Columns (3)–(4)), the triple interaction term (Postdump $\times$ High casualties $\times$ Close) is positive (≈ 0.039, $p = 0.002$), implying about 39 additional deaths per 1,000 live births for children born near landfills in high-casualty municipalities. In the non-EAR subsample (Columns (5)–(6)), the estimate is negative ($\approx$–0.081, $p = 0.001$); given differences in siting and timing, this is interpreted as heterogeneity across landfill regimes rather than a protective effect. Relative to the bombing-intensity benchmark (Table 2.3), the casualty-based DDD estimates at EAR-funded sites are similar in sign and magnitude, indicating that casualty counts capture a related dimension of conflict intensity. Differences for non-EAR sites are consistent with heterogeneity in siting and implementation.

Panel B of Table 2.7 reports the DiD estimates within the 0–6 km sample. In EAR-funded municipalities (Columns (3)–(4)), the interaction Postdump $\times$ High casualties is positive (0.036–0.045, $p < 0.01$), implying 36–45 additional deaths per 1,000 live births. In the non-EAR subsample (Columns (5)–(6)), the interaction is negative ($\approx$ –0.022, $p = 0.04$), again indicating heterogeneity by financing and timing.

Taken together, the results are consistent with conflict exposure (proxied by casualties) interacting with landfill proximity, with effects concentrated at EAR-funded sites. No comparable association is observed for later, non-ERA landfills. The pattern aligns with a governance channel in which conflict intensity and infrastructure management jointly shape health risks.

Table 2.7: Difference in Difference - Memory book casualties

Panel A: D×D×D						
infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.031***	-0.024	-0.019*	-0.015	-0.038***	-0.029***
	(0.008)	(0.014)	(0.010)	(0.022)	(0.007)	(0.009)
Postdump × High cas	0.010	0.006	-0.005	-0.009	0.057***	0.050**
	(0.010)	(0.010)	(0.012)	(0.013)	(0.017)	(0.018)
Close	0.000	0.002	-0.002	-0.002	-0.001	-0.003
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Postdump × Close	0.009	0.007	-0.005	-0.005	0.021**	0.021**
	(0.009)	(0.009)	(0.010)	(0.011)	(0.008)	(0.007)
High cas × Close	-0.014	-0.015	-0.023***	-0.023***	0.078***	0.082***
	(0.010)	(0.010)	(0.006)	(0.005)	(0.012)	(0.012)
Postdump × High cas × Close	0.011	0.013	0.039***	0.039***	-0.081***	-0.082***
	(0.016)	(0.015)	(0.010)	(0.011)	(0.017)	(0.017)
Dump Opening year FE	X		X		X	
Child Birth Year FE		X		X		X
Region FE	X	X	X	X	X	X
Covariates	X	X	X	X	X	X
N	4466	4466	3118	3118	1657	1657
N cluster	20	20	15	15	11	11

Panel B: D×D						
infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.024***	0.005	-0.026**	-0.001	-0.020**	0.002
	(0.004)	(0.018)	(0.005)	(0.015)	(0.004)	(0.033)
High cas	-0.18***	-0.19**	-0.10	-0.014**	-0.000	-0.002
	(0.004)	(0.006)	(0.006)	(0.005)	(0.017)	0.023
Postdump × High cas	0.023	0.027	0.036***	0.045***	-0.022***	-0.022
	(0.014)	(0.015)	(0.008)	(0.008)	(0.004)	(0.013)
Dump Opening year FE	X		X		X	
Child Birth Year FE		X		X		X
Region FE	X	X	X	X	X	X
Covariates	X	X	X	X	X	X
N	2366	2365	1780	1799	804	804
N cluster	10	10	6	6	6	6

Notes: The table displays coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, robust standard errors clustered at the municipality level are reported in parentheses. *Panel A* represents the triple difference in difference, while the simple difference in difference is reported in *Panel B*. *Panel B* sample is obtained by subsampling the children living within 6 km of a waste landfill. Furthermore, it also includes weights to account for children born in places with more than one dump within 6 km. The table shows the results for infant mortality where (1) represents the difference in difference outcome considering all the waste landfill with Region and opening year fixed effect; (2) is the same as the previous one with birth year fixed effects instead of opening year; (3) same as (1) but only considering waste landfill constructed with the EAR funding ; (4)same as (2) waste landfill constructed with the EAR funding; (5) same as (1) but only considering waste landfill *not* constructed with the EAR funding (so the internally funded); (6) is the same as (5) but with birth year fixed effect.

2.8.2 Geographic Anonymization and Measurement Precision

An important feature of the analysis is that MICS cluster coordinates are deliberately displaced to protect confidentiality (UNICEF, 2024). Urban clusters are randomly displaced up to 2 km; rural clusters up to 5 km, with 1% of rural clusters displaced up to 10 km. Crucially, clusters are never displaced across municipal boundaries; therefore, municipal-level bombing exposure is unaffected by anonymization.

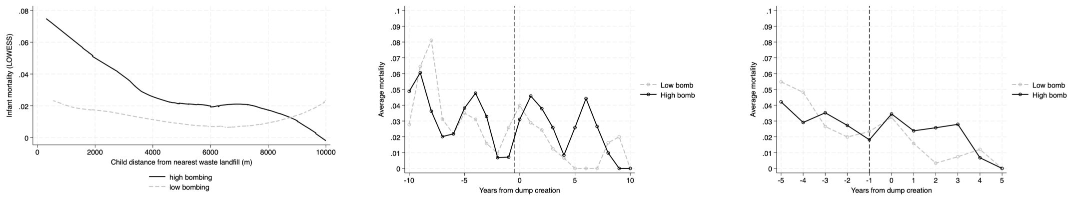
Displacement introduces non-systematic noise in proximity measures (e.g., distance to the nearest landfill), which can attenuate estimates. To assess robustness, the official UNICEF MICS GIS displacement routine is re-applied to the provided coordinates, and all proximity-based results are re-estimated under repeated randomized displacements within the allowed ranges. While this imprecision is non-systematic and unlikely to bias the estimates, it could attenuate the measured effects of landfill proximity on infant mortality. To address this, the randomization tool provided on the UNICEF website has been implemented and reapplied to the geographical data received from the MICS team to check if the results provided in the paper are robust to the displacement. To implement the simulation, the MICS coordinates supplied by UNICEF are treated as the true locations, and the original anonymization procedure is replicated by displacing each cluster centroid using a randomly drawn angle and distance.

All clusters are displaced at an angle ranging from 0 to 360 degrees. Urban cluster centroids are displaced up to 2 km, while rural cluster centroids are displaced up to 5 km, with 1% of clusters randomly selected for displacement up to 10 km. Importantly, all clusters remain within their original subnational area, typically corresponding to one administrative level below the survey's sample stratification (in this case, the municipality). If a cluster falls outside the correct administrative unit after displacement, the process is repeated until it falls within the correct boundary.

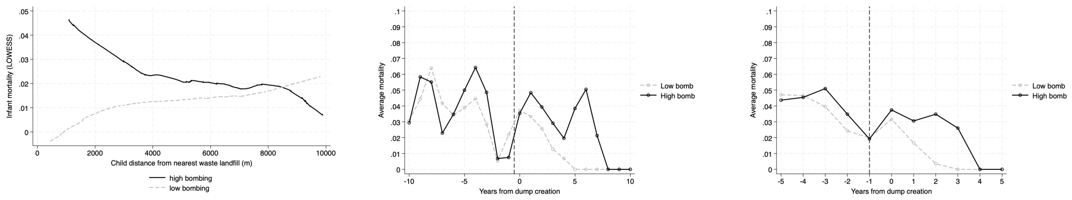
The MICS Geocode Plugin generates multiple outputs, some of which are added to QGIS as temporary vector layers, while others are saved as permanent layers. The number of outputs remains the same regardless of the input type, whether it is a CSV file or a Shapefile.

Across five displacement simulations, the DDD and DD coefficients remain stable in sign and significance, and qualitative patterns in event-study plots are unchanged. Figure 2.10 visualizes one set of simulations: (left) LOWESS of mortality versus distance supports the 6 km threshold; (middle) two-year averages around opening; (right) rolling averages around opening. These results indicate that the findings are not artifacts of geographic anonymization.

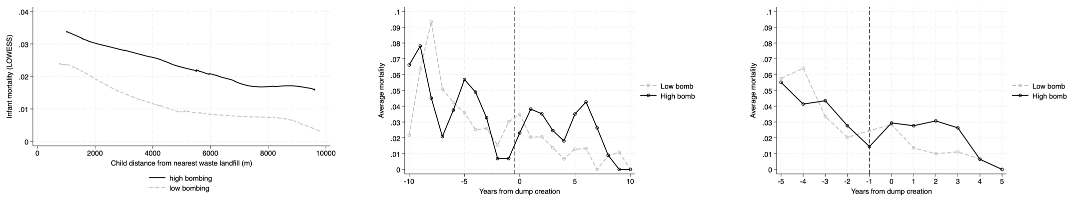
Figure 2.10: Simulations of anonymized displacement



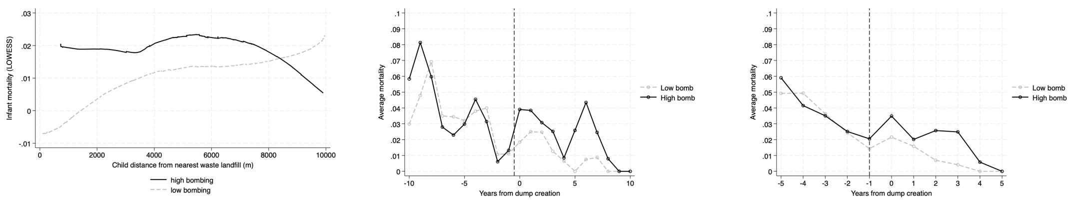
Simulation 1



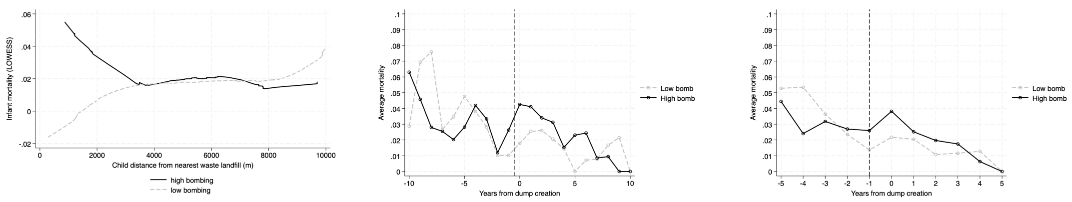
Simulation 2



Simulation 3



Simulation 4



Simulation 5

Notes: Left: LOWESS of infant mortality vs. minimum distance to the nearest landfill (supports the 6km cutoff). Middle: two-year averages around opening (vertical line at 0). Right: rolling averages around opening.

2.8.3 Alternative bombing cutoffs: 75th and 25th percentiles

Sensitivity to the definition of 'high bombing' is assessed by re-estimating the main models using the 75th and 25th percentiles as alternative cutoffs (baseline: median split). Panel A of Table 2.8 shows that the DDD triple interaction remains positive and is statistically meaningful under both tighter (75th) and looser (25th) definitions, though precision varies with sample size. Panel B (DD within the 6 km sample) shows the same concentration of excess mortality where bombing intensity is high.

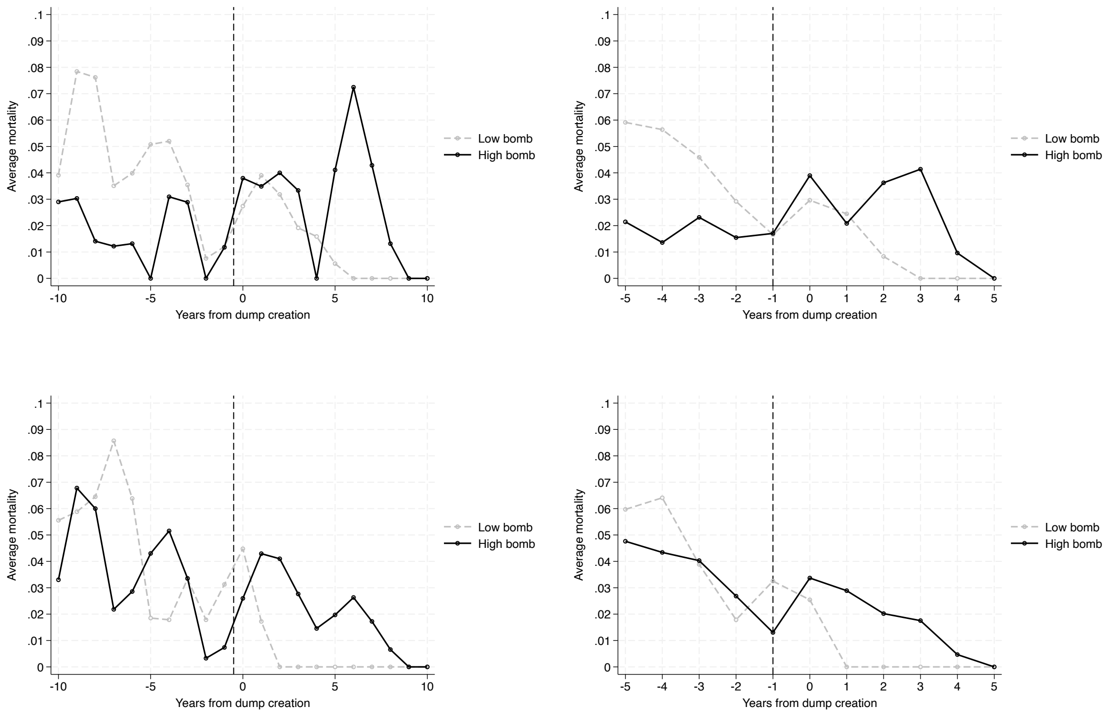
Table 2.8: Robustness to alternative bombing cutoffs (75th and 25th percentiles)

Panel A: D×D×D						
Infant mortality	75 percentile			25 percentile		
	(1)	(2)	(3)	(1)	(2)	(3)
Postdump	-0.013	-0.021	-0.005	-0.015	-0.029*	-0.009
	(0.010)	(0.013)	(0.011)	(0.013)	(0.015)	(0.018)
High bomb	-0.001	-0.013	0.017	0.006	0.000	0.012
	(0.017)	(0.009)	(0.040)	(0.011)	(0.015)	(0.018)
Postdump× High bomb	-0.003	0.009	-0.026	-0.002	0.004	-0.003
	(0.019)	(0.010)	(0.040)	(0.011)	(0.015)	(0.019)
Close	0.004	-0.005	0.010	0.006	0.004	0.012
	(0.006)	(0.007)	(0.008)	(0.010)	(0.016)	(0.016)
Postdump × Close	0.004	0.007	0.002	-0.006	-0.001	-0.008
	(0.008)	(0.010)	(0.011)	(0.014)	(0.024)	(0.018)
High bomb × Close	-0.017	-0.013	-0.016	-0.012	-0.018	-0.007
	(0.019)	(0.009)	(0.040)	(0.012)	(0.017)	(0.020)
Postdump × High bomb × Close	0.027	0.027**	0.021	0.026*	0.021	0.021
	(0.021)	(0.011)	(0.040)	(0.014)	(0.022)	(0.024)
Covariates	X	X	X	X	X	X
Birth Year FE	X	X	X	X	X	X
N	5373	3249	2564	5373	3249	2564
N cluster	26	17	18	26	17	18
Infant mortality mean	0.029	0.026	0.031	0.029	0.026	0.031
High bomb mean	0.302	0.332	0.218	0.713	0.776	0.677

Panel B: D×D						
infant mortality	75 percentile			25 percentile		
	(1)	(2)	(3)	(1)	(2)	(3)
Postdump	-0.001	-0.006	-0.003	-0.011	-0.019	-0.010
	(0.011)	(0.014)	(0.011)	(0.014)	(0.022)	(0.013)
High bomb	-0.021*	-0.033***	-0.008	-0.007	-0.022**	0.006
	(0.010)	(0.007)	(0.006)	(0.007)	(0.008)	(0.007)
Postdump × High bomb	0.028**	0.047***	0.006	0.022**	0.023	0.015
	(0.012)	(0.007)	(0.009)	(0.008)	(0.016)	(0.012)
Covariates	X	X	X	X	X	X
Child Birth Year FE	X	X	X	X	X	X
N	2944	1799	1383	2944	1799	1383
N cluster	15	8	11	15	8	11
Infant mortality mean	0.028	0.024	0.031	0.028	0.0244	0.031
High bomb mean	0.364	0.378	0.283	0.792	0.869	0.732

Notes: The table displays coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, robust standard errors clustered at the municipality level are reported in parentheses. *Panel A* represents the triple difference in difference while the simple difference in difference is reported in *Panel B*. *Panel B* sample is obtained by subsampling the children living within 6 km to a waste landfill. Furthermore, it also includes weights to account for children born in places with more than one dump within 6 km. The table shows the results for infant mortality considering (1) all the waste landfills; (2) the externally funded landfills ; (3) the waste landfill non-constructed with the aid money.

Figure 2.11: Average mortality and rolling window for different levels of high bombing



Notes: Each panel plots two-year rolling averages of infant mortality by event time (years relative to landfill opening). Curves are shown separately for municipalities above the bombing cutoff ("high bombing," black, solid) and below the cutoff (light-grey, dashed). Panels use alternative definitions of "high bombing": the 75th percentile (top row) and the 25th percentile (bottom row) relative to the baseline median split. The analysis frame is limited to births within 10 km of the nearest landfill. The figure is descriptive; inference relies on the DD/DDD estimates reported in Table 2.8.

2.8.4 Alternative proximity thresholds: 4 km and 8 km

The proximity threshold used to define exposure to landfill sites is varied from the baseline radius of 6 km to both tighter (4 km) and looser (8 km) definitions. Varying the radius affects the classification of children as exposed to landfill proximity, and therefore changes the size and composition of the treated sample, while holding constant the underlying time period, outcome definition, and empirical specification.

Table 2.9 reports the corresponding Difference-in-Difference-in-Differences (Panel A) and Difference-in-Differences (Panel B) estimates. Across specifications, the sign and qualitative pattern of the coefficients remain consistent with the main results. Estimated effects are attenuated when exposure is defined within a 4 km radius, reflecting the substantially smaller exposed sample and reduced statistical power at tighter cutoffs. By contrast, when the radius is expanded to 8 km, the triple interaction term remains positive and is statistically significant in several specifications, indicating that the adverse effects associated with landfill proximity in highly bombed municipalities persist when a broader exposure window is considered.

This pattern is consistent with the descriptive LOWESS evidence in Figure 2.3, which shows that infant mortality risk is highly localized around landfill sites—particularly in high-bombing municipalities—and stabilizes or gradually declines beyond approximately 6–8 km. Figure 2.12 further illustrates these dynamics using rolling averages around landfill opening for alternative proximity thresholds. Taken together, the regression and graphical evidence indicate that the main findings are not driven by an arbitrary choice of proximity threshold.

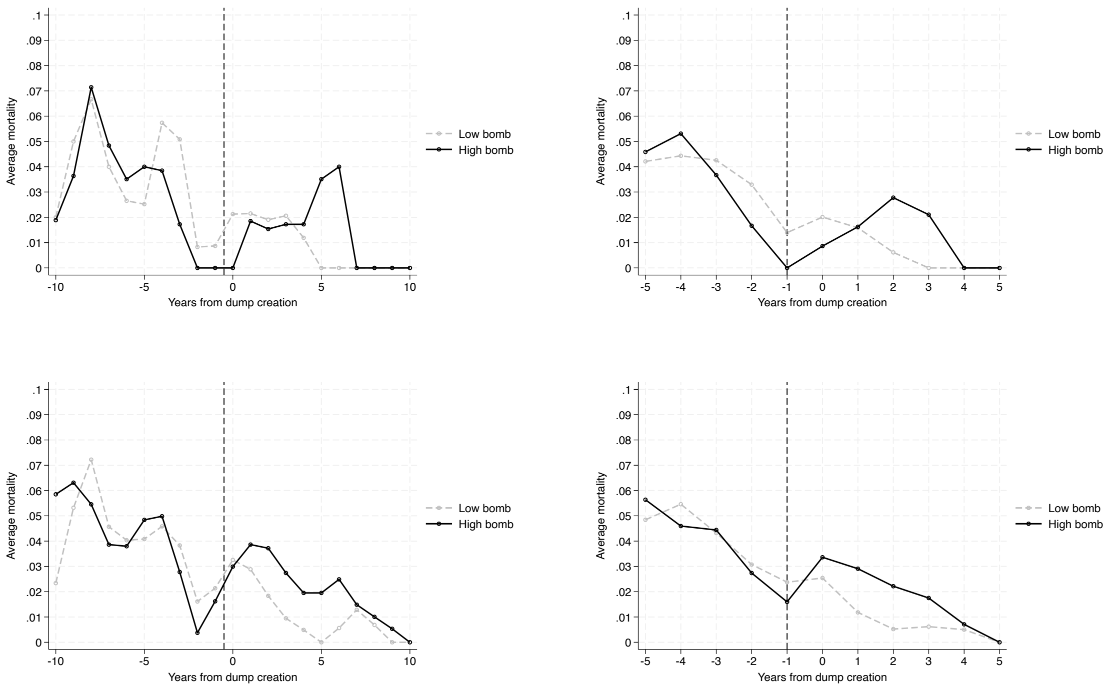
Table 2.9: Robustness to alternative proximity thresholds (4 km and 8 km)

Panel A: D×D×D						
infant mortality	4km radius			8km radius		
	(1)	(2)	(3)	(1)	(2)	(3)
postdump	-0.009	-0.014	-0.004	0.010	-0.007	0.033*
	(0.009)	(0.010)	(0.015)	(0.014)	(0.018)	(0.017)
high bomb	-0.003	-0.013	0.008	0.032*	0.009	0.053**
	(0.011)	(0.014)	(0.015)	(0.017)	(0.022)	(0.021)
postdump× highbomb	0.002	0.011	-0.008	-0.040**	-0.016	-0.067**
	(0.011)	(0.013)	(0.015)	(0.016)	(0.019)	(0.025)
close	-0.002	-0.006	0.010	0.026**	0.011	0.049***
	(0.010)	(0.012)	(0.016)	(0.010)	(0.011)	(0.009)
postdump×close	0.002	0.000	-0.006	-0.021	-0.008	-0.049**
	(0.010)	(0.018)	(0.015)	(0.013)	(0.017)	(0.018)
high bomb×close	-0.001	-0.002	-0.012	-0.043**	-0.025	-0.067**
	(0.014)	(0.022)	(0.020)	(0.018)	(0.020)	(0.023)
postdump×high bomb×close	0.002	0.006	0.008	0.051**	0.032	0.081**
	(0.016)	(0.023)	(0.025)	(0.019)	(0.024)	(0.029)
Covariates	X	X	X	X	X	X
Birth Year FE	X	X	X	X	X	X
N	5373	3249	2564	5373	3249	2564
N cluster	26	17	18	26	17	18
Infant mortality mean	0.029	0.026	0.031	0.029	0.026	0.031
High bomb mean	0.371	0.419	0.567	0.521	0.419	0.567

Panel B: D×D						
infant mortality	4km radius			8km radius		
	(1)	(2)	(3)	(1)	(2)	(3)
postdump	-0.005	-0.038*	-0.006	-0.008	-0.010	-0.015**
	(0.010)	(0.014)	(0.009)	(0.010)	(0.013)	(0.007)
high bomb	-0.005	-0.017	-0.003	-0.012*	-0.020**	-0.013
	(0.009)	(0.014)	(0.013)	(0.006)	(0.008)	(0.008)
postdump×high bomb	0.004	0.025	0.003	0.011	0.022*	0.013
	(0.014)	(0.012)	(0.019)	(0.009)	(0.011)	(0.010)
Covariates	X	X	X	X	X	X
Child Birth Year FE	X	X	X	X	X	X
N	1556	693	963	4435	2728	2054
N cluster	13	5	9	22	13	14
Infant mortality mean	0.024	0.0213	0.026	0.029	0.026	0.032
High bomb mean	0.521	0.244	0.362	0.54	0.44	0.582

Notes: The table displays coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, robust standard errors clustered at the municipality level are reported in parentheses. *Panel A* represents the triple difference in difference, while the simple difference in difference is reported in *Panel B*. *Panel B* sample is obtained by subsampling the children living within 6 km to a waste landfill. Furthermore, it also includes weights to account for children born in places with more than one dump within 6 km. The table shows the results for infant mortality considering (1) all the waste landfills; (2) the externally funded landfills ; (3) the waste landfills constructed without the aid money. Differences in coefficient magnitude across distance thresholds partly reflect changes in the size and composition of the exposed sample induced by alternative proximity definitions.

Figure 2.12: Average mortality and rolling window for different levels of distance from the landfill



Notes: Each panel plots two-year rolling averages of infant mortality by event time (years relative to landfill opening). The treated radius defining *close* is varied across panels: 4 km (tighter) and 8 km (looser); the analysis frame remains within 10 km of the nearest landfill. Curves are shown separately for high-bombing municipalities (black, solid) and lower-bombing municipalities (light grey, dashed). These patterns are descriptive and complement the DD/DDD estimates in Table 2.9.

2.9 Conclusions

This study examines the long-term effects of conflict on environmental management and public health, focusing on the unique post-conflict setting of Kosovo. It investigates how the intense bombing during the Kosovo War and subsequent mismanagement of waste facilities have contributed to lasting environmental and health challenges, particularly concerning infant mortality. The analysis bridges conflict studies and public health by exploring an overlooked dimension of war's legacy—the environmental hazards posed by poorly managed post-conflict waste landfills. By examining municipalities with varying intensity of bombing exposure and their proximity to waste landfills, this research uncovers a concerning link between conflict-driven environmental degradation and adverse health outcomes.

Using Difference-in-Difference (DD) and Triple Difference-in-Difference (DDD) methodologies, the paper establishes a robust relationship between bombing intensity, focusing on whether high-intensity bombing and proximity to post-conflict landfills jointly exacerbate health risks for infants. The findings reveal a significant increase in infant mortality for children born in proximity to waste landfills in municipalities heavily bombed during the war. Specifically, the triple interaction term indicates a 0.33–0.044 percentage point increase in infant mortality relative to the mean rate. Given an average infant mortality rate of 0.028, this represents a striking increase of about 120–160% before their first birthday compared to children in less affected regions. These effects remain statistically significant and robust across model specifications, including the incorporation of fixed effects and control variables such as maternal education, household wealth, and rural residence.

The findings suggest that localized exposure to poorly managed post-conflict waste infrastructure exacerbates adverse health outcomes, particularly when combined with the legacies of intense wartime violence. In municipalities that were heavily bombed and subsequently received international aid for waste management projects, corruption and insufficient regulatory oversight appear to have intensified the health risks. The study highlights how large inflows of aid funds, similar to findings in other post-conflict environments, can create opportunities for local corruption, further undermining environmental and health outcomes. In this context, waste facilities financed by international aid, such as those managed by the European Agency for Reconstruction (EAR), may have inadvertently contributed to the health challenges faced in these areas, as ineffective management compounded the environmental risks of landfills.

The paper also addresses the role of community trust and local governance, arguing that heightened distrust in institutions following the war has hindered effective waste management. The legacy of corruption and distrust in post-conflict Kosovo continues to affect both environmental quality and public health, creating persistent vulnerabilities for populations in heavily bombed areas. From a policy perspective, the findings underline the importance of integrating public health considerations into post-conflict reconstruction frameworks. Improved accountability and targeted investment in waste management systems are essential to mitigate the long-term health costs of war. Future research should further investigate the heterogeneous effects of aid-financed infrastructure and the interplay between conflict intensity, institutional capacity, and health outcomes.

In conclusion, this research sheds light on the lasting public health impacts of war-driven environmental degradation, calling for more sustainable and accountable management practices in post-conflict regions. For policymakers and international organizations, the findings underscore the need for targeted health and environmental interventions that address the compounded risks faced by communities near hazardous waste sites in post-conflict zones. Additionally, robust governance and anti-corruption measures should be integral to post-conflict recovery efforts to prevent such outcomes in the future.

Chapter 3

Long-term impact of conflict violence on trust

3.1 Introduction

Understanding how exposure to war and political violence shapes public attitudes toward institutions decades later remains a significant and underexplored challenge. While considerable attention has been devoted to the reconstruction of physical infrastructure in post-conflict societies, far less is known about how to restore citizens' trust in state institutions, a key ingredient for effective governance in fragile and transitional settings. Even among countries with similar conflict legacies and levels of international support, levels of institutional trust vary widely, often reflecting deep social and ethnic divisions. A key dimension of institutional trust is how citizens perceive the integrity and fairness of public institutions, often measured through perceived levels of corruption.

This study investigates how exposure to wartime violence and human rights violations relates to perceived institutional corruption. The analysis distinguishes between two forms of conflict exposure: municipality-level casualty rates and individual-level reports of human rights violations (HRV). It further explores whether perceptions vary by the level of government (local versus central) and by the type of violence experienced (e.g., targeting civilians versus armed forces). This context provides a unique opportunity to study how individual exposure to wartime violence and group narratives of victimization shape long-term attitudes toward institutional integrity.

To evaluate these dynamics, the analysis combines two sources. The United Nations Development Programme's Public Pulse surveys (2010–2023) provide individual-level measures of perceived corruption and, from 2018 onward, questions on personal or household exposure to HRV. The Kosovo Memory Book supplies municipality-level wartime casualty counts, disaggregated into civilian and armed forces deaths.

By combining localized measures of conflict exposure with individual perceptions of institutions, the study provides new micro-level evidence on how violence shapes long-term political attitudes.

The chapter makes two contributions. First, it implements a disaggregated measurement strategy that explicitly distinguishes institutional tier (local vs. central) and violence type (civilians vs. armed forces), revealing patterns masked in aggregate exposure. Second, it combines standard regressions with conditional randomization inference in a joint-exposure design, clarifying which patterns are robust to high–low contrasts and which rely on assumptions about linear functional form.

By connecting insights from the conflict and corruption literatures, this study highlights how wartime exposure shapes institutional trust. Given Kosovo's salient ethnic cleavages, the analysis centers on violence exposure and tests whether estimated effects differ by ethnicity. This research highlights how individual-level exposure to conflict violence shapes perceptions of corruption and institutional trust across municipalities. These contributions extend a growing literature in political economy and conflict studies on the long-term effects of violence by bringing in new micro-level evidence from a post-conflict context.

Prior studies have shown that civil war exposure can increase political participation and collective action (Bellows and Miguel, 2009). More recent research highlights that the relationship between conflict exposure and post-conflict attitudes is highly heterogeneous, varying based on the perpetrator of violence, victim identity, and local social structures (Voytas and Crisman, 2024; Hadzic and Tavits, 2019; Dorff, 2017; Gallego, 2018). Work on transitional justice further emphasizes that the perceived fairness and inclusivity of post-conflict institutions critically shape public trust (Loyle and Appel, 2017; Fernandes, 2011). While extensive research has focused on the political and institutional strategies of elites, such as informal power-sharing arrangements, the persistence of patronage networks, or the co-optation of state institutions for personal or factional gain (Spector, 2005; Ledeneva, 2017)—far less attention has been paid to the lived experiences and beliefs of ordinary citizens. Three gaps in the existing literature motivate the approach. First, macro-level approaches often miss how individual exposure maps into legitimacy beliefs. Existing work on post-conflict institutions largely relies on macro or cross-country indicators (e.g., Collier et al. (2003); Fearon and Laitin (2003)) and seldom connects personal exposure to violence—especially state-perpetrated abuses—to individual beliefs about legitimacy. Second, work on corruption perceptions seldom links measured exposure to institutional abuse at the micro level. Likewise, the corruption-perceptions literature is extensive (Treisman, 2000; Rose-Ackerman, 2008) but rarely links perceptions to micro-level exposure to institutional abuse. Third, identity and cleavage structures condition how violence translates into post-war trust, with cultural distance and ethnic conflict producing persistent divisions (Guarnieri, 2025), and exposure to state violence shaping views of justice and institutions (Voytas and Crisman, 2024). In Kosovo's contested state formation, evaluations plausibly depend on who used violence and which tier of government is judged.

Theoretically, the impact of violence exposure on institutional trust is ambiguous. On one hand, witnessing or suffering HRV may heighten political awareness and lead to increased scrutiny of institutions. Such exposure can catalyze civic engagement and vigilance, as shown in Sierra Leone (Bellows and Miguel, 2009). On the other hand, repeated trauma may erode trust and political efficacy, fostering apathy, withdrawal, or fatalism, particularly where reconstruction is seen as unequal or exclusionary. This process has been linked to learned helplessness and diminished social capital (De Luca and Verpoorten, 2015; Kijewski and Freitag, 2018). These opposing pathways suggest that high corruption perceptions may signal either active political vigilance or deep institutional alienation. Disentangling these mechanisms empirically is essential for understanding the diversity of post-conflict attitudes and the complexities of trust formation.

This study uses perceptions of corruption as a proxy for institutional trust, capturing skepticism toward institutional fairness, accountability, and legitimacy in a context where direct trust measures may be abstract or idealized.[1] Kosovo's post-conflict trajectory—marked by mass displacement, documented abuses, and prolonged international administration—frames how citizens judge fairness and accountability today. The results speak to the design of post-conflict governance and anti-corruption policy in Kosovo and similar contexts. Associations based on individual-level HRV are robust, whereas those based on spatially aggregated casualties warrant caution given potential residual spatial confounding.

The empirical strategy proceeds in two steps. First, cross-sectional joint specifications

[1] While not identical, perceptions of corruption are closely related to institutional trust. Individuals who see institutions as corrupt effectively express skepticism about their fairness and legitimacy. This proxy is especially valuable in post-conflict settings where direct trust questions may be abstract or idealized. The literature recognizes corruption perceptions as indicators of institutional quality and state–society relations (Treisman, 2000; Rose-Ackerman, 2008). In fragile, polarized states like Kosovo, these perceptions reflect lived experiences and broader normative beliefs about governance, varying by institutional level and influenced by personal and group-based histories (e.g., De Luca and Verpoorten, 2015; Kijewski and Freitag, 2018).

relate perceived corruption to civilian and armed forces casualties (per 1990 municipal population) entered jointly. Because these exposures are time-invariant at the municipality level, municipality and year fixed effects are not included; standard errors are clustered at the municipality level. Second, models using individual-level HRV exposure include municipality and year fixed effects and identify from within-municipality individual variation; specifications are shown with and without individual covariates. Randomization-inference placebos complement each design: for casualties, conditional permutation tests shuffle one municipal exposure while holding the other fixed; for HRV, the exposure is permuted across individuals within the fixed-effects framework.

The empirical analysis yields three main findings[2] that motivate the empirical design and its interpretation. First, in joint municipality-level casualty models, civilian casualties are positively associated with perceived corruption—especially for central institutions (4.34–4.57 index points per 1 pp increase in civilian-death share; $p \leq 0.01$)—and smaller but positive locally (1.77–2.42; $p \leq 0.05$). Armed forces casualties are negatively associated locally (–8.84 to –5.04; all significant) and negative but imprecise for central institutions. Second, in fixed-effects models, wartime HRV is negatively associated with perceived corruption at both tiers (local: –0.096 to –0.090; central: –0.063 to –0.055), whereas post-war HRV is positive locally (+0.039 to +0.042) and near zero for central institutions. Third, cumulative exposure matters: wartime $\times$ post-war HRV interactions are positive (local: 0.086–0.100; central: 0.038–0.063). Inference supports these patterns. In the joint casualty specifications, conditional placebos using binary (above-median) municipal exposures reject the sharp null when permuting one exposure and holding the other fixed (civilian: $p = 0.0039$ local, 0.0002 central, 0.0001 all; armed-force: $p = 0.0017$ local, 0.0144 central, 0.0014 all). By contrast, analogous placebos based on continuous casualty shares yield large permutation p-values (0.22–0.58), consistent with the signal operating through high-versus-low contrasts rather than a linear dose–response. In the HRV fixed-effects framework, permuting individual HRV exposure places the observed coefficients in the extreme tails of the placebo distributions (empirical $p \approx 0$ across outcomes).

The rest of the chapter proceeds as follows. Section 2 provides historical background on the conflict and its aftermath. Section 3 reviews the related literature. Section 4 presents the data sources and construction of key variables. Section 5 outlines the empirical strategy. Section 6 discusses the main results and robustness checks. Section 7 concludes.

[2]PCA/FA rescaled to the Simple Average scale and sign-aligned; see Footnote 8

3.2 Historical background

Kosovo's path to statehood has been conditioned by a legacy of ethnic conflict, contested sovereignty, and international intervention. In the late 1980s, the revocation of Kosovo's autonomous status by Serbian authorities marked the beginning of a decade of repression, during which ethnic Albanians were systematically excluded from public employment and education. This led to the formation of a parallel society and nonviolent resistance. However, by the mid-1990s, the failure of peaceful efforts to secure autonomy prompted the emergence of the Kosovo Liberation Army (KLA), an armed insurgency demanding independence.

By 1998, open conflict erupted between the KLA and Yugoslav/Serbian forces. The war escalated quickly, characterized by mass killings, forced displacement, and widespread destruction. The KLA, while framed by many Kosovo Albanians as a liberation movement, was also involved in contentious practices, and its growing influence blurred the lines between wartime heroism and postwar political dominance. Serbian forces responded with a campaign of violence that Human Rights Watch and the OSCE later documented as constituting systematic ethnic cleansing. Approximately 1.5 million ethnic Albanians, about three-quarters of the population, were expelled or fled their homes. Accounts from survivors describe terror campaigns involving looting, arson, executions, and forced expulsion at gunpoint.

The conflict in Kosovo presents a unique case. While it began as a civil war between ethnic Albanians and Serbian state forces, it ultimately culminated in Kosovo's declaration of independence in 2008, recognized by the majority of sovereign states. This trajectory distinguishes Kosovo from other post-conflict settings. During the war, the Kosovo Liberation Army operated as a guerrilla movement fighting against Serbian authorities. Unlike many civil conflicts, the aftermath in Kosovo led to the creation of a new sovereign state, rather than a reformed political arrangement within an existing one.

Such distinctions are important in Kosovo, where civilian casualties are often interpreted as evidence of institutional victimization, while armed forces deaths are more likely to be viewed as legitimate wartime losses. These contrasting experiences shape how violence is embedded in post-war narratives of justice and legitimacy, particularly in a society marked by ethnic divisions and contested institutional authority. Ethnic fractionalization and polarization—factors shown to undermine trust and increase conflict risk (Alesina and La Ferrara, 2005; Montalvo and Reynal-Querol, 2005)—remain deeply entrenched.

The scale of civilian suffering in Kosovo prompted NATO's intervention in March 1999, through a 78-day aerial bombing campaign (Operation Allied Force). Although the aim was to halt ethnic cleansing and enforce Serbian withdrawal, the intervention occurred without UN Security Council authorization, making it a precedent-setting case of humanitarian intervention. As detailed in international reports, NATO's targeting strategy often failed to distinguish between military and civilian assets. While war crimes tribunals later found no grounds for prosecuting NATO, collateral damage was considerable: approximately 500 civilians were killed in Yugoslavia, with over half of those deaths occurring in Kosovo. This loss of life should not be downplayed. However, it is important to contextualize it within the broader scale and asymmetry of the conflict: according to the Kosovo Memory Book, out of the 13,154 documented deaths, 10,094 were civilians, and an estimated 86% of all victims were ethnic Albanians. This overwhelming concentration of civilian suffering among Kosovar Albanians helps explain the urgency and moral imperative behind NATO's intervention, despite its controversial legal standing.

Following Serbia's withdrawal, Kosovo came under the authority of the United Nations Interim Administration Mission in Kosovo (UNMIK). UNMIK assumed full executive, legislative, and judicial powers, overseeing the reconstruction of institutions and public services. The European Union Rule of Law Mission (EULEX) later supported these efforts. While these missions provided a vital stabilizing presence, they also created a governance gap: the state was administered largely by international actors, making accountability diffuse and weakening the development of local legitimacy.

Kosovo's post-conflict governance is characterized by fragmented authority, weak accountability, and pervasive public perceptions of corruption, especially in areas most affected by violence. These persistent challenges underscore the need to understand how direct exposure to state-perpetrated violence shapes citizens' beliefs about government legitimacy.

This institutional rupture complicates notions of state legitimacy. The Serbian state, widely perceived by Kosovo Albanians as the main perpetrator of violence, contrasts with the emergent Kosovar institutions, which, despite being successors in governance, had to establish legitimacy from the ground up. Moreover, while Serbian police and military forces were responsible for much of the violence, credible reports have also documented HRV by the KLA against ethnic Serbs.[3]. As such, the perpetrating actors, whether Serbian or Kosovar, were organized institutions, blurring the boundaries of state and non-state violence in the eyes of victims and survivors. Post-war reconstruction in Kosovo

[3]For example: Reuters, 2024; Human Rights Watch, 2001; BBC News, 2012

was accompanied by large-scale international aid and formal state-building reforms. Yet, persistent institutional fragility and weak oversight created fertile ground for corruption.

The complexity of Kosovo's post-conflict governance is compounded by its ethnic geography. Serb-majority municipalities, especially in the north, have resisted integration into the Republic of Kosovo's institutions, operating under parallel structures supported by Belgrade. To address some of these issues, Kosovo and Serbia signed the Brussels Agreement in 2013 under EU mediation, aiming to normalize relations and integrate Serb-majority areas into Kosovo's legal framework. However, implementation has been partial, and tensions between the two sides remain high. This incomplete integration continues to affect perceptions of state legitimacy across ethnic communities.

Interpreting institutional trust in Kosovo also requires recognizing its unique and contested political history. The conflict began as an ethnopolitical civil war and fundamentally reshaped the locus of legitimate authority. For many, especially Kosovo Albanians, the pre-independence state was viewed as a foreign or hostile authority rather than a legitimate government. Consequently, the notion of "my government" only emerged after the war, complicating retrospective assessments of state responsibility and adding depth to current perceptions of corruption and trust.

This study explores whether and how these legacies continue to influence perceptions of public institutions. Post-war reconciliation efforts have been fragmented and limited. While transitional justice initiatives were implemented, such as truth commissions and victim compensation schemes, they often lacked political backing or credibility across ethnic lines. Crucially, the absence of comprehensive accountability for wartime abuses, particularly for crimes committed by the KLA, has undermined public confidence in the justice system and fed perceptions of impunity. This dynamic came to the forefront in 2020, when Kosovo's then-President Hashim Thaçi, a former senior KLA commander, was indicted for war crimes and crimes against humanity by the Kosovo Specialist Chambers in The Hague. Thaçi subsequently resigned to face trial, highlighting the enduring challenge of reckoning with wartime legacies at the highest levels of government.

The political upheaval triggered by Thaçi's resignation led to renewed demands for institutional change. In the 2021 elections, Kosovo saw a major political turnover, with the Vetëvendosje movement winning a decisive victory on an anti-corruption and reformist platform. While this political shift signals potential for greater accountability and citizen engagement, historical legacies of conflict, ethnic division, international governance, and incomplete reconciliation continue to shape public attitudes toward institutions.

These historical legacies are central to understanding how citizens perceive corruption and trust in public institutions today. In post-conflict Kosovo, trust is not simply a matter of service delivery or transparency, but of whether the state is seen as ours, legitimate, and protective.

3.3 Literature review

The relationship between conflict exposure and institutional trust has gained significant attention in political science and economics. Empirical studies have explored how experiences of violence influence political behavior and perceptions of governance.

Conflict Exposure and Institutional Trust

Research indicates that exposure to violence can have complex and sometimes contradictory effects on political engagement and trust in institutions. In some cases, such exposure can stimulate prosocial behavior and political participation. For instance, Blattman (2009) finds that individuals abducted by rebels in Uganda exhibited higher voting rates and greater community leadership roles despite the trauma endured. Similar findings from Sierra Leone indicate that exposure to civil war can enhance local collective action and community rebuilding (Bellows and Miguel, 2009). However, other studies emphasize that violence often erodes trust in state institutions, especially when the state is seen as complicit or directly responsible for harm. Citizens exposed to state-perpetrated violence often display lower trust in government, greater skepticism toward state-led initiatives, and reduced engagement with transitional justice mechanisms (Voytas and Crisman, 2024; De Juan and Pierskalla, 2016; Wang, 2021). In Colombia, exposure to state violence significantly decreased participation in state-sponsored transitional justice programs, underlining the role of perpetrator identity in shaping post-conflict attitudes. Moreover, studies increasingly recognize the heterogeneous effects of violence based on individual characteristics and local contexts. For example, Hadzic and Tavits (2019) find that gender moderates the political consequences of violence, while Dorff (2017) and Gallego (2018) highlight the role of kinship networks and armed group strategies, respectively. These insights suggest that the relationship between violence exposure and civic engagement is highly context-dependent and cannot be fully understood without considering perpetrator identity, social structures, and post-conflict dynamics.

The Kosovo case fits within this more complex framework. Exposure to violence—whether perpetrated by Serbian forces or Kosovar actors—may generate heightened political vigilance or deep civic disillusionment, depending on the perceived legitimacy of the perpetrator and broader group narratives. While this study does not disaggregate violence by perpetrator ethnicity, it captures variation in the type of victims (civilian vs. armed group), which may serve as a proxy for how violence is perceived in terms of legitimacy or victimization.

Determinants of Corruption Perceptions

Perceptions of corruption are shaped by a wide array of factors, including historical legacies, institutional quality, cultural norms, and socioeconomic conditions. Cross-country studies highlight that Protestant traditions, British colonial legacies, and economic development are associated with lower perceived corruption (Treisman, 2000). Long-standing democratic institutions are also linked to lower corruption perceptions, emphasizing the importance of institutional maturity and accountability mechanisms.

In post-conflict settings, the establishment of transparent, inclusive, and accountable governance structures becomes particularly vital. While international organizations often support anti-corruption initiatives, their effectiveness can vary based on local conditions, including societal divisions and the inclusivity of political processes (Rose-Ackerman, 2008). Research also indicates that conflict exposure can shape perceptions of institutional corruption. Studies find that trust in the state is substantially undermined when governments are perceived to have failed to prevent violence or, worse, when they are seen as perpetrators (Gates and Justesen, 2020; Ishiyama et al., 2018). In Colombia, for example, policies perceived as biased or externally imposed, such as extradition arrangements, further damaged trust among affected populations (Kreutz and Nussio, 2019). Corruption perceptions are further conditioned by the effectiveness and inclusivity of transitional governance structures. International interventions, such as UNMIK and EULEX in Kosovo, often aim to build transparency and the rule of law, but their success depends heavily on local perceptions of fairness and legitimacy.

Ethnic Cleavages, Polarization, and Institutional Trust

Ethnic divisions significantly impact trust in public institutions. Studies have shown that individuals are more likely to trust members of their own ethnic group, which can translate into varying levels of trust in ethnically aligned institutions (Robinson, 2016). Extensive research has documented that ethnic diversity and polarization can undermine social cohesion and institutional trust (Alesina and La Ferrara, 2005; Montalvo and Reynal-Querol, 2005). In polarized societies, institutions may be perceived as representing the interests of specific groups rather than the collective good, exacerbating perceptions of corruption and illegitimacy.

In Kosovo, the ethnic division between Albanians and Serbs remains a potent source of differentiated trust. Serb-majority areas continue to resist integration into Kosovo's institutions, while many Albanians harbor lingering distrust rooted in wartime experiences and the international administration period. These divisions likely mediate how wartime violence translates into perceptions of corruption today. While the empirical analysis does not disaggregate effects by ethnic identity due to data constraints, the broader context of ethnic polarization informs the interpretation of heterogeneous responses to violence exposure across municipalities.

Post-Conflict Governance, Transitional Justice, and Trust

The pursuit of transitional justice is pivotal in shaping post-conflict institutional trust. Mechanisms such as truth commissions, special war crimes courts, and reparations aim to address past abuses and rebuild legitimacy. However, their success is highly contingent on perceived fairness, inclusivity, and the capacity to deliver meaningful accountability(Fernandes, 2011; Loyle and Appel, 2017).

While some view these actions as steps toward accountability, others perceive them as external impositions. Although this study does not directly assess attitudes toward transitional justice or international institutions, these dynamics provide important context for understanding how individuals in post-conflict Kosovo evaluate institutional integrity. The findings on perceived corruption—especially following exposure to institutionally perpetrated violence—may reflect underlying frustrations with unfulfilled justice or contested authority.

3.4 Data

This study combines repeated cross-sections from the Public Pulse opinion surveys with municipality-level conflict intensity from the Kosovo Memory Book. Together, these sources allow analysis of how variation in wartime violence across municipalities relates to perceptions of institutional corruption observed across survey waves. This section describes data sources, variable construction, and the analytical samples.

Public Pulse Surveys

The main data come from the UNDP Public Pulse surveys, a repeated cross-section fielded since 2010 across all 38 municipalities[4]. The design yields a pooled cross-section: individuals are not tracked across waves, and each wave samples a new, representative set of respondents. This dataset collects information on perceptions of corruption, which serves as the proxy for institutional trust. Respondents are asked to rate corruption of 16 institutions on a four-point scale ('not present at all', 'present at a small scale', 'present at a medium scale', 'present at a large scale') along with personal information such as municpality they live in, gender, age, ethnicity, working status, income, home ownership, etc.

For the sake of the analysis, composite outcomes are constructed at two governance tiers—local and central. Local institutions, include police, education, municipal utilities and services, hospitals, and civil society, which are services predominantly managed at the municipal level and involve frequent citizen contact, providing fertile ground for micro-level corruption[5]. And, central institutions—such as the central government, customs, courts, political parties, central government and agencies (e.g., tax authority, state-owned enterprises) and international organizations—are more distant[6] and often associated with high-level integrity concerns.

However, starting in 2018, the survey discontinued questions on EULEX and international organizations. As a result, these institutions exhibit substantial missingness in later waves and are excluded from the main analysis along with corruption in hospitals[7], hence such items have been removed from the analysis. Summary statistics and empirical results, therefore, focus on the remaining 13 institutions with complete longitudinal coverage. This local-central distinction reflects both theoretical expectations and empirical realities, capturing how proximity and interaction frequency shape citizens' exposure to

[4]While the Public Pulse surveys are described as biannual, fieldwork did not follow a strict twice-per-year schedule. Interviews took place in: June 2010; Jan/Aug 2011; Mar/Oct 2012; May/Dec 2013; Jul 2014; Feb/Sep 2015; Apr/Oct 2016; Oct 2017; May/Nov 2018; May/Nov 2019; Apr 2020; May/Oct 2021; May/Oct 2022; and Apr 2023.

[5]For instance, police officers have been known to solicit small bribes—colloquially referred to as 'for a coffee'—in exchange for overlooking traffic violations; in 2017, more than 50 officers were arrested in a major anti-corruption sweep targeting precisely this behavior (Global Initiative). Similarly, in the health sector, doctors have reportedly demanded informal payments or 'gifts' to provide proper medical treatment or expedite referrals, sometimes even for deceased patients, as revealed in a high-profile bribery case involving major hospitals in 2016 (Reuters). The education system has also faced allegations of corruption, with teachers and school administrators soliciting money or favors from parents in return for better grades or preferential treatment for their children (UNDP).

[6]These forms of corruption are typically less routine but more visible. A notable example occurred in 2021, when senior officials at Kosovo's media regulator were implicated in a bribery scheme involving payments from media companies in exchange for favorable licensing decisions (Transparency International).

[7]Hospitals are unevenly distributed across regions; tertiary services are centralized in Pristina, which generates structural missingness for some municipalities.

and perceptions of institutional corruption.

In addition to the simple-average index, Principal Component Analysis (PCA) and Factor Analysis (FA)[8] are used to construct composite indices. Mean perceived corruption equals 0.76 for central institutions and 0.63 for local institutions (Table 3.1).

The Public Pulse survey employs stratified sampling designed to cover all 38 municipalities. While this ensures broad geographical coverage[9], the survey design supports municipality-level analysis for most locations and waves. The sample size varies by wave, ranging from approximately 1,266 to 2,863 respondents per wave[10].

Observations with missing values on core outcomes or covariates are excluded. Appendix Table C.1 documents missingness by wave and demographics; nonresponse on key variables is limited and shows no systematic patterns.

Summary statistics for key individual- and municipality-level variables are provided in Table 3.1. The full dataset includes over 30,000 observations; however, the estimation sample used in the baseline analysis consists of approximately 25,807 individuals with non-missing values for core outcomes and covariates across the relevant survey waves. All specifications use a consistent sample to avoid estimation differences driven by compositional changes.

Descriptive patterns highlight substantial cross-municipality heterogeneity and gradual declines in average perceived corruption in some locations over time. These patterns motivate disaggregation by institutional tier and support the joint-specification approach at the municipality level[11].

[8] To make coefficients comparable across aggregation methods, PCA/FA scores are linearly rescaled to the Simple Average (SA) scale and sign-aligned so that higher values indicate more perceived corruption:

$$\tilde{y}_{it}^{X} = \frac{\sigma_{\mathrm{SA}}}{\sigma_X}\left(y_{it}^{X} - \mu_X\right) + \mu_{\mathrm{SA}},$$

where μ_{SA} and σ_{SA} are the mean and standard deviation of the SA index and μ_X and σ_X are the mean and standard deviation of the raw $X \in \{\mathrm{PCA}, \mathrm{FA}\}$ index, computed on the corresponding estimation sample. If $\mathrm{corr}(\tilde{y}^{X}, \mathrm{SA}) < 0$, we multiply $\tilde{y}^{X}$ by -1. This affine transformation preserves ranks and t-statistics up to scaling and places all outcomes on SA units.

[9] Public Pulse XVIII Methodological Note: The survey provides sampling weights, and the total numbers in the descriptions are weighted to reflect actual population figures across municipalities, but these are excluded from the main analysis to maintain consistency with the survey and capture any raw heterogeneity in corruption perceptions by exposure.

[10] Larger sample sizes in select years, such as 2,863 respondents in 2011, were strategically implemented to enhance the precision of estimates and enable more detailed subgroup analyses. In contrast, some waves, including 2014 with 1,306 respondents, employed smaller samples due to operational considerations, while still maintaining rigorous sampling protocols. During this period, Kosovo's total population grew modestly from approximately 1.78 million in 2011 to about 1.88 million in 2021. These design features collectively ensure that the survey provides reliable estimates of evolving public opinion across both national and municipal levels.

[11] Some municipalities, such as Gjakovë and Pejë, report consistently lower levels of perceived corruption, while others, including Zubin Potok and Leposaviq, report higher and more persistent concerns. Also, several municipalities show declining corruption perceptions over time, while others exhibit stagnation or increases, potentially reflecting local governance or reform dynamics. This visualization provides a nuanced backdrop for interpreting the empirical results presented in the main analysis. It underscores

Table 3.1: Summary Statistics

	Variable	Obs	Mean	Median	Min	Max
1	Age	30087	40.735	39.000	18.000	99.000
2	Female	29485	0.478	0.000	0.000	1.000
3	Urban	14467	0.479	0.000	0.000	1.000
4	Ethnic Albanian	20202	0.669	1.000	0.000	1.000
5	Ethnic Serb	5239	0.174	0.000	0.000	1.000
6	Other Ethnicity	4726	0.157	0.000	0.000	1.000
7	Unemployed	8115	0.269	0.000	0.000	1.000
8	Pensioner	2742	0.091	0.000	0.000	1.000
9	Years of Schooling	29645	11.395	12.000	0.000	30.000
10	House Owner	24804	0.834	1.000	0.000	1.000
11	Income monthly (EUR)	3952	403.035	350.000	40.000	3500.000
12	Land Owner	3307	0.263	0.000	0.000	1.000
13	Corruption Local (avg)	24229	0.610	0.667	0.000	1.000
14	Corruption Central (avg)	25626	0.762	0.857	0.000	1.000
15	Corruption All (avg)	25994	0.688	0.769	0.000	1.000
16	Local Corruption (PCA)	18264	0.023	0.168	-1.870	1.175
17	Central Corruption (PCA)	17154	-0.026	0.372	-2.546	0.801
18	All Corruption (PCA)	15413	-0.020	0.273	-2.427	1.034
19	Local Corruption (FA)	18264	0.021	0.167	-1.636	1.029
20	Central Corruption (FA)	17154	-0.026	0.366	-2.327	0.724
21	All Corruption (FA)	15413	-0.023	0.257	-2.289	0.951
22	Corruption: Banks	13166	0.582	1.000	0.000	1.000
23	Corruption: Central Gov	18372	0.744	1.000	0.000	1.000
24	Corruption: Civil Society	11675	0.544	1.000	0.000	1.000
25	Corruption: Court	20176	0.798	1.000	0.000	1.000
26	Corruption: Customs	19760	0.793	1.000	0.000	1.000
27	Corruption: Education	15128	0.584	1.000	0.000	1.000
28	Corruption: KEDS	17586	0.732	1.000	0.000	1.000
29	Corruption: Local Gov	17155	0.669	1.000	0.000	1.000
30	Corruption: PAK	17601	0.771	1.000	0.000	1.000
31	Corruption: Political Parties	20441	0.805	1.000	0.000	1.000
32	Corruption: Police	14014	0.557	1.000	0.000	1.000
33	Corruption: PTK	16140	0.694	1.000	0.000	1.000
34	Corruption: TAK	14270	0.697	1.000	0.000	1.000

Note: Most variables are binary indicators taking values between 0 and 1. Exceptions include *age*, *years of schooling*, and *income (EUR)*, which are continuous. PCA and FA variables are standardized indices from principal component and factor analysis, respectively. The variables *Corruption: International Orgs*, *Corruption: EULEX*, and *Corruption: Hospital* are excluded from the analysis due to excessive missing observations, which could bias estimation results.

Public Pulse Survey: Human Rights Violations

From 2018 onward, the Public Pulse survey introduces individual-level data on experiences of HRVs, such as abuse, harassment, threats, murder, deportation, or property destruc-

the importance of local context and motivates the paper's municipality-level approach to studying the long-term effects of conflict and governance on institutional trust.

tion. Specifically, respondents are asked whether they or an immediate family member (e.g., parent, sibling, or child) personally experienced or witnessed a HRV during or after the 1998–1999 conflict. A binary indicator that takes value equal to one is constructed when either the respondent or a family member witnessed any such violation, and zero otherwise. This variable enables the analysis to capture within-municipality variation in conflict-related trauma exposure, complementing the aggregate casualty data. Because HRV questions were fielded only from 2018 onward, regressions using this measure rely on a smaller estimation sample of about 8,500 individuals.

Memory Book and Casualties distinction

Conflict exposure is measured using (i) the individual HRV indicator described above and (ii) the Kosovo Memory Book, a registry of wartime deaths and disappearances compiled by the Humanitarian Law Center. Municipality-level casualties are coded by victim status (civilian versus armed forces) during the 1998–1999 conflict and normalized by the 1990 municipal population. Using the 1990 denominator precedes displacement and reconstruction, facilitates comparison across municipalities of different sizes, and limits endogeneity to wartime migration.

Administrative boundaries changed after the conflict: the number of municipalities today differs from 1999, largely because several larger municipalities were subsequently subdivided. To avoid introducing spurious variation from redistricting, all municipalities created after 1999 inherit the parent municipality's casualty *shares* (civilian and armed forces) computed per 1990 population. This harmonization ensures that exposure is comparable across time and not driven by post-war boundary changes.[12]

Two exposure codings are used. First, continuous casualty shares (per 1990 population) for civilians and armed forces enter jointly in cross-sectional specifications. Second, binary indicators classify municipalities as above-median ("high") exposure by violence type. Conditional randomization inference placebos are implemented for the binary indicators by permuting one exposure while holding the other fixed; for completeness, analogous placebos are reported for the continuous shares.

Selective out-migration is unlikely to drive the results. Post-war return rates were high, with the bulk of displaced persons returning in the months after the conflict, typically to their pre-war municipalities. This reduces the scope for systematic selection in who

[12]This choice is conservative: it preserves the information set available at the time of the conflict and prevents mechanical differences in exposure arising solely from administrative splits.

remains exposed to local institutional environments[13].

3.5 Empirical Strategy

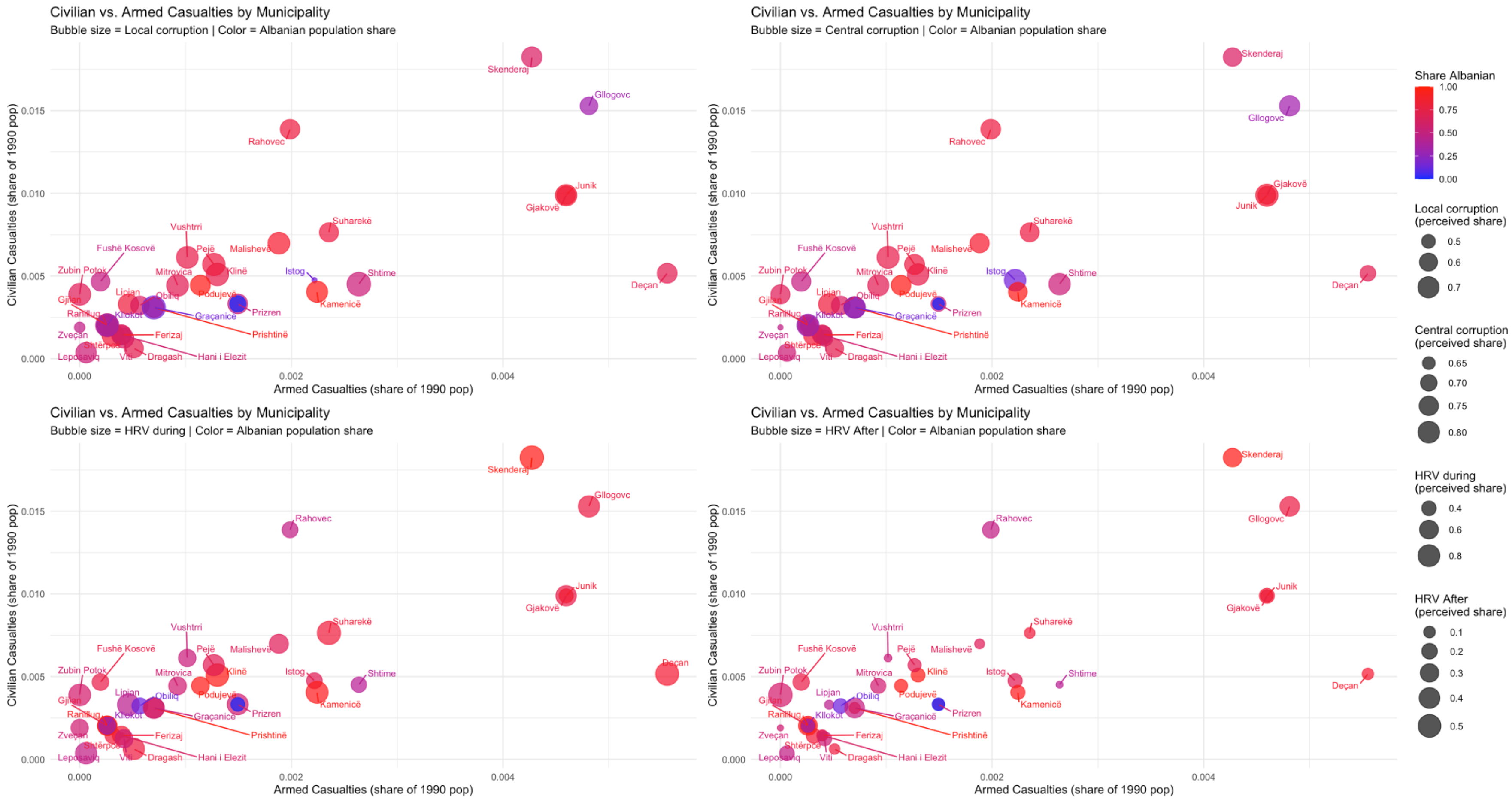
This section examines whether spatial variation in conflict intensity during the 1998–1999 Kosovo War is associated with contemporary perceptions of corruption, interpreted as a reduced-form proxy for institutional trust. The central hypothesis is that exposure to intense wartime violence undermines institutional trust, thereby generating persistently elevated perceptions of corruption.

The empirical strategy exploits cross-sectional variation in conflict exposure that was largely shaped by military strategy, ethnic targeting, and logistical constraints (e.g., troop deployments, proximity to borders, and KLA presence), rather than by pre-existing institutional quality or socio-economic conditions (Judah, 2002). Given the observational setting, estimates are interpreted as robust associations consistent with a causal interpretation under these maintained assumptions.

To examine the relationship between conflict exposure and post-war institutional perceptions, municipality-level bubble plots are presented in Figure 3.1, with normalized armed and civilian casualty shares (relative to the 1990 population) plotted on the horizontal and vertical axes, respectively. Four outcome dimensions are visualized: perceived local corruption, perceived central corruption, HRV reported during the war, and HRV reported after the war. Bubble size corresponds to the average share of respondents reporting each outcome, while bubble color indicates the ethnic composition of each municipality, measured as the share of ethnic Albanians. The plots indicate that municipalities with the highest exposure to both armed and civilian casualties—such as Skenderaj and Gllogovc—also tend to report elevated levels of perceived corruption and HRV, particularly in Albanian-majority areas. While HRV reported during the war closely aligns with conflict exposure, post-war HRV reports appear more spatially dispersed, potentially reflecting institutional fragility or persistent ethnic tensions. These patterns offer descriptive support for the core identification strategy, which relies on spatial heterogeneity in conflict intensity and demographic composition to examine long-term effects on institutional trust. In the regressions, civilian and armed forces casualties enter jointly to separate their conditional associations.

[13] Krieger (2001); UNHCR (2000) report that by November 1999, 76% of the displaced population returned to Kosovo and over 87% resettled in the same municipalities they lived before.

Figure 3.1: Conflict exposure, perceived corruption, and HRV by municipality.



Notes: Each bubble represents a municipality; bubble size reflects the outcome variable, and color indicates the share of ethnic Albanians. The top-left panel shows local corruption, the top-right panel shows central corruption, the bottom-left panel shows human rights violations during the war, and the bottom-right panel shows human rights violations after the war. Casualties are expressed as shares of the 1990 population. The ethnic composition shown in the plots is derived from survey respondents and may differ from official census figures. In municipalities with small sample sizes, these shares should be interpreted with caution.

Corruption Perceptions Variables and Principal Component Analysis

Perceptions of corruption are measured separately for local and central government institutions using multiple survey items from the Public Pulse dataset. The primary dependent variables are the simple average (share) of the relevant binary items within each governance level. Each item is coded as 1 if corruption is reported at a medium or large scale, and 0 otherwise.

However, perceptions of corruption across different institutions may reflect a common underlying dimension of institutional integrity or legitimacy, which may not be perfectly captured by simple counts or averages. To assess whether the items load on a common latent dimension and to verify measurement coherence, Principal Component Analysis (PCA) and Factor Analysis (FA) are used as complementary checks.

PCA is a data reduction technique that transforms a set of possibly correlated variables into a smaller number of uncorrelated components. The first principal component captures the largest possible variance shared among the items, effectively summarizing the core pattern of corruption perceptions across multiple institutions. FA, in contrast, models the shared variance among observed variables as arising from one or more latent factors, representing unobserved constructs like overall perceived corruption. Both methods help identify whether the multiple corruption indicators reflect a single dimension or multiple distinct dimensions of perception. For comparability with the simple average, PCA and FA scores are linearly rescaled to match the simple average's mean and standard deviation and re-oriented so that higher values indicate more perceived corruption.

The first principal component explains roughly half of the total variance in each tier, and FA yields highly similar loadings (Table C.4). The results show that the first principal component accounts for a substantial proportion of the total variance (approximately 46% for local institutions and 52% for central institutions), indicating a strong underlying common factor. The FA yields highly similar loadings, reinforcing the interpretation that these indices reflect a coherent latent construct of perceived corruption. The item loadings are consistent in magnitude across methods and institutional groups, with all variables contributing positively and substantially to the first component/factor (see Table C.4). Given this strong internal consistency, a simple unweighted average of the item responses serves as a transparent and interpretable aggregation method. The average correlates closely with the PCA and FA-derived indices, and is preferred in the main analysis due to its straightforward economic interpretation and ease of communication. Nevertheless, PCA and FA-based scores are included in supplementary analyses to confirm the robustness of the findings.

Model Specification

The main regression model is specified in Equation 3.1, where $Corruption_{imt}$ denotes the perceived corruption score for individual i in municipality m at year t. Each corruption item is coded as zero if corruption is perceived as absent or present on a small scale, and as one if present on a medium or large scale. These binary items are then aggregated into indices using the three methods described above. This coding choice reflects a latent sentiment prevalent in Kosovo that corruption is perceived as widespread and systemic, where even moderate perceptions of corruption signal a meaningful distrust in institutions rather than a benign or neutral view. This interpretation is supported by data from the United Nations Office on Drugs and Crime (UNODC), which reports that approximately 56% of Kosovars believe corruption is increasing, alongside significant distrust in institutions such as the judiciary and local government (UNODC, 2022). Each method is applied separately to three domains of institutional corruption: *local* (e.g., municipal government, local police), *central* (e.g., central government, parliament), and *overall* (all institutions combined). The key independent variables are two municipality-level measures of wartime intensity: civilian casualties per 1990 population (Civilian Casualties$_m$) and Armed forces casualties per 1990 population (Armed forces Casualties$_m$). These are used to measure wartime violence intensity as the number of conflict-related casualties normalized per 1990 municipality m population, constructed separately for civilian and armed group deaths using data from the Kosovo Memory Book.

The model includes a vector of individual-level covariates, X_{it}, capturing age, gender, urban residence, years of schooling, employment status (unemployed, public sector, private sector, student), and homeownership.

$$Corruption_{imt} = \alpha + \beta_1 \text{Civilian Casualties}_m + \beta_2 \text{Armed forces Casualties}_m + \beta_3 X_{it} + \varepsilon_{imt} \tag{3.1}$$

Regressions enter civilian and armed forces casualties jointly in all casualties specifications and the analysis is by corruption level (local vs. central vs. overall), with and without individual-level controls. Standard errors are clustered at the municipality level. Additional heterogeneity analyses include interactions between violence exposure and age-group dummies (30–44, 45–59, 60+; baseline 18–29). For civilian exposure, the interaction estimates indicate an age gradient: the 45–59 and 60+ cohorts exhibit significantly larger associations with perceived corruption—most pronounced for central institutions (PCA/FA, rescaled)—whereas the 30–44 cohort shows no systematic difference. By contrast, for armed forces exposure, age interactions are generally imprecisely estimated and not robust across outcomes; where significant, signs do not align with the civilian pattern. Detailed results are reported in Tables C.10 and C.11.

Although Civilian Casualties_m and Armed forces Casualties_m are correlated across municipalities, variance inflation factors (2.25) and the condition number (4.24) indicate no harmful multicollinearity; see Appendix Section C.3.

To complement the analysis of long-term war casualties, a second model focuses on individual-level variation in post-conflict HRV. In the extended specifications (Equation 3.2), the model replaces war-casualty exposure with an alternative explanatory variable capturing post-war governance quality, measured by the incidence of human rights violations, HRV_{it}, as reported in the Public Pulse survey. More precisely, HRV_{it} is a binary indicator equal to one if individual i reports having experienced or witnessed a human rights violation in period t. The vector X_{it} includes the same set of individual-level controls as in the main specification (age, gender, urban residence, education, employment, and homeownership).

Formally, the regression model for individual-level post-conflict HRV is specified as:

$$Corruption_{imt} = \alpha + \gamma_1 \text{HRV during conflict}_{it} + \gamma_2 \text{HRV after conflict}_{it} + \gamma_3 X_{it} + \mu_m + \lambda_t + \varepsilon_{imt} \tag{3.2}$$

Formally, the extended regression model is specified as:

$$\begin{aligned} Corruption_{imt} = {} & \alpha + \gamma_1 \text{HRV during}_{it} + \gamma_2 \text{HRV after}_{it} + \gamma_3 \left(\text{HRV during}_{it} \times \text{HRV after}_{it}\right) \\ & + \gamma_4 X_{it} + \mu_m + \lambda_t + \varepsilon_{imt} \end{aligned} \tag{3.3}$$

In model 3.3, the interaction term HRV $\text{during}_{it} \times$ HRV after_{it} captures whether combined exposure to HRVs during and after the conflict is associated with different perceptions compared to those with no or single-period exposure.

The vector X_{it} includes the same controls as before. Municipality fixed effects (μ_m) absorb all time-invariant local characteristics such as conflict intensity or governance history. Year fixed effects (λ_t) control for temporal shocks common across municipalities.

Using within-municipality variation, this specification identifies how individual experiences of HRV relate to perceptions of corruption. The coefficient γ_1 reflects the association between wartime HRVs and perceived corruption (for those with no post-conflict HRVs); γ_2 captures the association for individuals who only experienced post-conflict HRVs; and γ_3 shows how combined exposure may have an additional or interactive effect. Unlike the casualties models, the HRV specifications include municipality and year fixed effects, exploiting within-municipality variation over time. They therefore address a different estimand from the casualties models, which capture between-municipality, long-run differences.

Identification Strategy and Causal Inference

The identification strategy assumes that, conditional on observed covariates, cross-municipality variation in wartime violence intensity is plausibly orthogonal to unobserved determinants of contemporary corruption perceptions. This is supported by historical accounts indicating that the geography of violence was shaped primarily by military strategy, ethnic targeting, and logistical constraints (e.g., troop deployments, proximity to borders, and KLA presence), rather than by pre-existing governance quality or socio-economic conditions (Judah, 2002). Post-war displacement and return patterns further suggest limited long-run re-sorting across municipalities, which reduces concerns that the observed cross-section is driven by selective migration.

In the casualties specifications (Equation 3.1), municipality fixed effects and municipality-specific trends are not included as the exposure variables are time-invariant at the municipality level; including such fixed effects would mechanically absorb all variation in the casualty measures. Inference is based on standard errors clustered at the municipality level. Because civilian and armed-force casualties co-vary spatially, both exposures enter jointly to mitigate omitted-variable bias that would arise in separate specifications. Consistent with this, Appendix C.4 reports a sizable cross-municipality correlation alongside variance inflation factors and condition numbers that indicate no harmful multicollinearity.

Separately, Tables 3.2–3.3 report covariate balance across high- and low-exposure groups (median split for casualties; indicators for HRV during/after). These tables are descriptive and intended to document sample comparability; they are not alternative identification strategies. All casualty regressions control for individual covariates, and HRV specifications add municipality and year fixed effects. The casualties models estimate cross-sectional, long-run differences across municipalities with differing wartime intensity (no municipality fixed effects), whereas the HRV models (Equations 3.2–3.3) include municipality and year fixed effects and therefore exploit within-municipality variation over time. The two sets of estimates thus target different estimands and should be interpreted accordingly.

As with any observational design, residual confounding cannot be fully ruled out. Pre-existing institutional weakness, historical ethnic tensions, or unmeasured governance dynamics could correlate with both wartime violence and contemporary perceptions. The estimates are therefore interpreted as robust associations consistent with a causal interpretation under the maintained assumptions but not as definitive causal effects. This cautious interpretation is consistent with best practices in conflict research, where perfect experimental conditions are rarely attainable due to the inherent complexity of war-related processes.

One important concern is that wartime violence may have directly degraded the quality of local institutions—creating a causal pathway from conflict to worse governance and then to negative perceptions. This potential institutional legacy represents a key confounder. While the analysis controls for individual-level characteristics such as age, gender, urban residence, education, employment status, and homeownership, and includes municipality fixed effects in Specifications 3.2 and 3.3 (the HRV models), this risk cannot be fully ruled out. Results disaggregated by institutional level and robustness checks partially address this concern, but residual bias may remain.

Ethnicity as a potential confounder

A further consideration is whether ethnicity, particularly Albanian versus non-Albanian identity, confounds the relationship between wartime exposure and post-conflict corruption perceptions. To act as a confounder, a variable must (i) be correlated with the treatment, (ii) correlated with the outcome, and (iii) not lie on the causal pathway. Ethnicity plausibly satisfies the first two conditions: Albanian-majority municipalities experienced higher levels of civilian targeting and HRV, and descriptive patterns show that Albanian respondents tend to report higher perceived corruption, particularly toward central institutions.

Table C.6 examines this directly by interacting conflict exposure with an indicator for Albanian ethnicity. The results show consistent evidence that ethnicity moderates, but does not confound, the main effects. For civilian and armed casualties, the interaction with Albanian identity is strongly negative, indicating that the positive association between wartime exposure and perceived corruption is concentrated among non-Albanian respondents. For Albanians, the combined main and interaction effects are close to zero, suggesting no substantive relationship.

The same pattern is evident for HRV. During-war HRV is negatively associated with corruption perceptions overall, but the positive interaction with Albanian identity implies that this effect is again driven by non-Albanian respondents. For after-war HRV, the main effect is positive, while the negative Albanian interaction almost fully offsets it, leaving Albanians with no net response. Thus, the observed HRV effects largely reflect the perceptions of Serbs and other non-Albanian minorities.

Taken together, these findings suggest that ethnicity functions as an effect modifier rather than a confounding variable. In other words, ethnic heterogeneity shapes the distribution of treatment effects, but the direction and statistical significance of the main results are not explained away by ethnicity. Accordingly, ethnicity and its interactions are explored in heterogeneity models but are not required to identify unbiased average treatment effects in the baseline specification.

Covariate Balance Across Exposure Groups

Before presenting the estimation strategy, it is important to assess whether individuals in high-exposure municipalities differ systematically from those in less-affected areas along baseline characteristics. Tables 3.2 and 3.3 report mean differences, p-values from two-sample t-tests, and standardized mean differences (SMDs) for a set of demographic and socioeconomic covariates: age, gender, urban residence, ethnicity (Albanian, Serbian, other), years of schooling, home ownership, income, unemployment, and pensioner status. Treatment and control groups are defined by a median split of exposure (civilian casualties, Armed forces casualties, HRV during, HRV after).

Several covariates differ significantly between treatment and control groups at conventional levels. For instance, average years of schooling and age show statistically significant but substantively modest differences, while income and unemployment vary more strongly between groups. Ethnic composition shows the most substantial differences: in the HRV-during specification, SMDs exceed 0.5 for both Albanian and Serbian identity, confirming that ethnicity is a crucial source of imbalance (see Section 3.5). Despite these differences, most SMDs fall below the conventional imbalance threshold of 0.25 (Rosenbaum and Rubin, 1985; Austin, 2009), suggesting that observable characteristics are generally comparable across exposure groups.

To address these differences, all regressions include the full set of individual-level controls (age, gender, urban, ethnicity, schooling, home ownership, income, unemployment, and pensioner status). In addition, ethnicity is explicitly included as a baseline control in all models and further examined in heterogeneity specifications (Section 3.5). The HRV specifications additionally include municipality fixed effects to absorb time-invariant local heterogeneity. Robustness checks, including matching estimators and placebo tests presented in Section 3.5, show that the main results are not driven by residual imbalance in pre-treatment characteristics.

While the HRV-during specification shows a large SMD for ethnicity, reflecting known ethnic targeting during the war, other covariates are relatively well balanced. Explicitly adjusting for ethnicity as a control and examining its interaction with wartime exposure reduces, though does not fully eliminate, concerns over imbalance. This approach follows established best practices in observational causal inference, where treatment assignment is inherently non-random and modest baseline imbalances are expected. Transparent reporting and adjustment for such differences enhance the credibility of the identification strategy.

3.6 Results

This section reports two empirical designs. First, cross-sectional joint specifications relate perceived corruption to *civilian* and *armed forces* casualties per 1990 municipal population (no municipality or year fixed effects; standard errors clustered by municipality). Second, models using human rights-violation exposures include municipality and year fixed effects and identify from within-municipality, individual-level variation. All specifications are shown with and without individual covariates.

Wartime casualties: joint estimates (cross-municipality)

Table 3.4 enters civilian and armed forces casualty shares jointly. Two patterns emerge. First, civilian casualties are positively associated with perceived corruption. For central institutions, coefficients are large and precise (4.34–4.57; $p < 0.01$). For local institutions, coefficients are smaller but mostly positive and statistically different from zero (1.77–2.42; $p < 0.05$). Second, armed forces casualties are negatively associated with perceived corruption, especially for local institutions (–8.84 to –5.04; all significant). For central institutions, the simple-average estimates are negative and significant (about –5.6), while PCA/FA are negative but imprecise.[14].

[14] Casualty regressors are municipality-level shares of the 1990 population; coefficients are interpreted per one-percentage-point (0.01) increase.

Table 3.2: Balance Table: Civilian and Armed Forces Casualties (Median Split)

Variable	Control Mean	Treatment Mean	Difference	p-value	SMD
Civilian Casualties					
Albanian ethnicity	0.661	0.700	0.039	0.000	0.084
Serb ethnicity	0.151	0.168	0.018	0.000	0.048
Other ethnicity	0.188	0.131	-0.057	0.000	-0.157
Age (years)	40.3	41.0	0.69	0.000	0.044
Female	0.498	0.489	-0.010	0.460	-0.019
Urban residence	0.486	0.470	-0.016	0.006	-0.033
Household income (index)	73.2	95.2	22.0	0.000	0.106
Unemployed	0.243	0.292	0.049	0.000	0.111
Homeowner	0.819	0.848	0.029	0.000	0.078
Pensioner	0.089	0.093	0.004	0.310	0.012
Years of schooling	11.5	11.3	-0.133	0.001	-0.039
Observations	14058	13859			
Armed Forces Casualties					
Albanian ethnicity	0.703	0.656	-0.047	0.000	-0.100
Serb ethnicity	0.166	0.153	-0.013	0.003	-0.035
Other ethnicity	0.131	0.189	0.059	0.000	0.161
Age (years)	40.2	41.1	0.916	0.000	0.058
Female	0.492	0.497	0.005	0.715	0.010
Urban residence	0.478	0.478	0.000	0.976	0.000
Household income (index)	82.4	85.7	3.31	0.188	0.016
Unemployed	0.263	0.273	0.010	0.061	0.022
Homeowner	0.818	0.848	0.030	0.000	0.080
Pensioner	0.089	0.093	0.004	0.282	0.013
Years of schooling	11.6	11.2	-0.345	0.000	-0.101
Observations	14096	13821			

Note: Treatment and control groups are defined by a median split of municipality-level exposure (per 1990 population). Observations equal to the median are included in the control group. Reported p-values come from two-sample t-tests of mean equality. Standardized mean differences (SMDs) are computed as the mean difference between treatment and control divided by the pooled standard deviation across groups. An SMD below 0.25 is conventionally interpreted as evidence of acceptable balance (Rosenbaum and Rubin, 1985; Austin, 2009).

These patterns are consistent with differential attribution across institutional tiers: conditional on the other casualty type, civilian fatalities map onto higher central-level perceived corruption, whereas armed forces fatalities map onto lower local-level perceived corruption[15].

The HRV specifications use municipality and year fixed effects and therefore identify from within-municipality, individual-level variation; they are not directly comparable to the cross-municipality casualty estimates above.

[15]Interpretation (non-causal): communities may valorize armed forces deaths as sacrifice linked to local defense networks (increasing forbearance toward proximate institutions), whereas civilian victimization is attributed to central institutions, lowering trust. The study does not test these channels.

Table 3.3: Balance Table: HRV During and After the War

Variable	Control Mean	Treatment Mean	Difference	p-value	SMD
HRV During the War					
Albanian ethnicity	0.589	0.820	0.231	0.000	0.524
Serb ethnicity	0.226	0.053	-0.174	0.000	-0.518
Other ethnicity	0.185	0.127	-0.058	0.000	-0.160
Age (years)	39.8	42.2	2.40	0.000	0.157
Female	0.486	0.500	0.015	0.246	0.030
Urban residence	0.473	0.456	-0.017	0.122	-0.034
Household income (index)	6.06	4.85	-1.22	0.003	-0.066
Unemployed	0.200	0.238	0.038	0.000	0.091
Homeowner	0.871	0.915	0.044	0.000	0.141
Pensioner	0.074	0.108	0.034	0.000	0.118
Years of schooling	11.8	11.6	-0.19	0.008	-0.057
Observations	3400	5652			
HRV After the War					
Albanian ethnicity	0.719	0.702	-0.017	0.208	-0.037
Serb ethnicity	0.118	0.175	0.056	0.000	0.159
Other ethnicity	0.163	0.123	-0.040	0.000	-0.114
Age (years)	41.4	40.1	-1.30	0.003	-0.085
Female	0.494	0.493	0.000	0.983	-0.001
Urban residence	0.464	0.411	-0.053	0.000	-0.106
Household income (index)	4.93	5.86	0.933	0.084	0.052
Unemployed	0.226	0.214	-0.012	0.296	-0.030
Homeowner	0.903	0.882	-0.021	0.026	-0.066
Pensioner	0.097	0.082	-0.015	0.061	-0.053
Years of schooling	11.6	12.1	0.533	0.000	0.161
Observations	8330	1408			

Note: Treatment is defined at the individual level as reporting (1) vs. not reporting (0) a HRV during or after the war. Reported p-values come from two-sample t-tests of mean equality. Standardized mean differences (SMDs) are computed as the mean difference between treatment and control divided by the pooled standard deviation across groups. SMDs above 0.25 indicate substantial imbalance; values above 0.5, as observed for ethnicity in the HRV-during specification, highlight ethnicity as a crucial potential confounder.

Human Rights Violations (HRV)

Tables 3.5 and C.9 estimate the relationship between exposure to HRV during and after the conflict and perceptions of corruption. All models include municipality and year fixed effects, and corruption is measured using three aggregation methods: Simple Average, Principal Component Analysis (PCA), and Factor Analysis (FA). Specifications with and without individual-level covariates are reported.

Table 3.5 shows that exposure to HRV during the conflict is consistently associated with lower perceived corruption in both local and central institutions. For local institutions (Panel A), coefficients range from -0.108 to -0.104 in PCA/FA (rescaled) and from

-0.096 to -0.090 in the Simple Average (all $p < 0.01$). For central institutions (Panel B), coefficients are likewise negative and precise (about -0.063 to -0.055 in Simple Average; similar magnitudes in PCA/FA). This stands in contrast to the positive casualty associations and reflects different identifying variation: the casualty models exploit cross-municipality exposure largely borne by Albanians, whereas the HRV models identify from within-municipality individual variation after fixed effects, where residual variation is disproportionately among minorities who report lower perceived corruption.

Post-conflict HRV is positively associated with local perceived corruption (Panel A), but small and statistically insignificant for central institutions (Panel B).

Interaction of wartime and postwar HRV. For local institutions (Panel A), the wartime×postwar HRV interaction is positive and statistically significant across all models (0.086–0.100), indicating that higher perceived corruption arises primarily among respondents exposed to both shocks; exposure to either period in isolation is associated with neutral or negative effects. For central institutions (Panel B), the interaction is likewise positive and statistically significant in most specifications (0.038–0.063), offsetting the weakly negative postwar main effect and yielding a net positive association. This pattern restores consistency with the casualty regressions: cumulative victimization across wartime and postwar periods maps into higher perceived corruption, whereas one-off HRV exposure (which, after fixed effects, is disproportionately reported by minorities) tends to attenuate measured perceptions[16].

Impact on Perceptions of Local Institutions

For local institutions, the joint cross-municipality estimates (Table 3.4, Panel A) show a positive association with civilian casualties (1.77–2.42; significant in most specifications) and a negative association with armed forces casualties (–8.84 to –5.04; significant throughout).

In the fixed-effects HRV models (Table 3.5, Panel A), wartime HRV is associated with lower local perceived corruption (–0.108 to –0.104 in the rescaled PCA/FA; –0.096 to –0.090 in the simple average; all $p < 0.01$), whereas post-conflict HRV is positively associated with local perceived corruption (+0.034 to +0.042; $p < 0.01$). The interaction models (Table C.9, Panel A) indicate that cumulative exposure (wartime × post-war HRV)

[16]Interpretation (non-causal): Possible channels include rally-around-the-flag dynamics, lower evaluative standards after trauma, or attribution of isolated postwar abuses to disorder rather than institutional malpractice. A full test of these mechanisms is beyond the scope of the design.

Table 3.4: Regression Results - Perceived Corruption and Wartime Casualties

	Simple Average		PCA		Factor Analysis	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Local Institutions						
Civilian casualties	2.416***	1.770**	2.341**	1.766*	2.254**	1.676
	(0.813)	(0.826)	(1.016)	(1.031)	(1.014)	(1.029)
Armed forces casualties	−8.836***	−7.707***	−5.652**	−5.080*	−5.619**	−5.037*
	(2.211)	(2.241)	(2.779)	(2.816)	(2.774)	(2.811)
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	27,491	27,055	18,264	18,034	18,264	18,034
Panel B: Central Institutions						
Civilian casualties	4.342***	4.361***	4.393***	4.533***	4.425***	4.574***
	(0.770)	(0.782)	(1.014)	(1.025)	(1.014)	(1.024)
Armed forces casualties	−5.676***	−5.599***	−3.519	−3.864	−3.593	−3.935
	(2.101)	(2.142)	(2.774)	(2.823)	(2.773)	(2.822)
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	27,472	27,034	17,154	16,953	17,154	16,953

Notes: Standard errors in parentheses. Statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Municipality-clustered standard errors in parentheses. Columns (1),(2),(3) exclude individual covariates; columns (4),(5),(6) include age, gender, urban residence, years of schooling, employment status (unemployed, public, private, student), and homeownership. No municipality or year fixed effects are included in these casualties specifications because exposure is time-invariant at the municipality level. PCA and FA indices are linearly rescaled to match the simple average's mean and standard deviation and re-oriented so higher values denote more perceived corruption.

is positive and precise (0.081–0.100), implying that elevated local corruption perceptions emerge primarily among respondents exposed in both periods.

Impact on Perceptions of Central Institutions

For central institutions, civilian casualties are positively and precisely associated with perceived corruption in the joint cross-section (Table 3.4, Panel B: 4.34–4.57; $p < 0.01$). Armed forces casualties are negative in the simple-average specification (about –5.6; significant) and negative but imprecise in PCA/FA.

In the HRV fixed-effects models (Table 3.5, Panel B), wartime HRV is consistently associated with lower perceived central corruption (about –0.063 to –0.055; $p < 0.01$). Post-conflict HRV is small and imprecise. The interaction models (Table C.9, Panel B) show a positive wartime×post-war HRV term (0.038–0.063; significant in most specifications), indicating that cumulative exposure restores a net positive association with central corruption perceptions despite the weakly negative post-war main effect.

Conditional Randomization Inference – Municipality-Level Placebos

Robustness is assessed using randomization inference that permutes one municipality-level exposure across municipalities while *holding the other exposure fixed* in the joint

Table 3.5: Regression Results – Impact of HRV During and After the Conflict on Corruption Measures (with Municipality and Year FE)

	Simple Average		PCA		Factor Analysis	
	(1)	(4)	(2)	(5)	(3)	(6)
Panel A: Local Institutions						
HRV during conflict	-0.096^{***}	-0.090^{***}	-0.108^{***}	-0.104^{***}	-0.108^{***}	-0.104^{***}
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
HRV after conflict	0.042^{***}	0.039^{***}	0.037^{***}	0.035^{***}	0.036^{***}	0.034^{***}
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,488	8,464	7,190	7,170	7,190	7,170
Panel B: Central Institutions						
HRV during conflict	-0.058^{***}	-0.055^{***}	-0.063^{***}	-0.059^{***}	-0.063^{***}	-0.059^{***}
	(0.008)	(0.009)	(0.009)	(0.030)	(0.009)	(0.009)
HRV after conflict	-0.005	-0.005	-0.013	-0.014	-0.012	-0.013
	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,514	8,489	7,089	7,070	7,089	7,070

Note: Standard errors in parentheses. Statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Regressions estimate the association between exposure to HRV during and after the 1998–99 conflict and perceptions of corruption in local (Panel A), central (Panel B). Corruption indices are constructed using Simple Average, PCA, and Factor Analysis (FA). Models include municipality and year fixed effects. Columns (1), (3), (5) are baseline models; Columns (2), (4), (6) include individual-level controls. Sample sizes vary due to item nonresponse and index-specific coverage. Positive postwar HRV effects are driven by respondents exposed to both wartime and postwar HRV (interaction term), while postwar HRV alone is associated with weak or negative effects at the central level.

Table 3.6: Regression Results - Impact of Wartime HRV on Corruption Measures with Municipality FE and Interaction

	Simple Average		PCA		Factor Analysis	
	(1)	(4)	(2)	(5)	(3)	(6)
Panel A: Local Institutions						
HRV during conflict	-0.106***	-0.100***	-0.121***	-0.155***	-0.120***	-0.115***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
HRV after conflict	-0.024	-0.023	-0.041*	-0.040	-0.041*	-0.040
	(0.022)	(0.022)	(0.025)	(0.025)	(0.025)	(0.026)
HRV during × after	0.086***	0.081***	0.100***	0.097***	0.099***	0.095***
	(0.025)	(0.025)	(0.028)	(0.028)	(0.028)	(0.028)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,488	8,464	7,190	7,170	7,190	7,170
Panel B: Central Institutions						
HRV during conflict	-0.063***	-0.060***	-0.070***	-0.066***	-0.070***	-0.066***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
HRV after conflict	-0.038*	-0.034	-0.061***	-0.059**	-0.061***	-0.059**
	(0.021)	(0.021)	(0.023)	(0.023)	(0.023)	(0.023)
HRV during × after	0.044*	0.038	0.062**	0.058**	0.063**	0.059**
	(0.024)	(0.024)	(0.026)	(0.026)	(0.026)	(0.026)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,514	8,489	7,089	7,070	7,089	7,070

Note: Standard errors in parentheses. Statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Regressions estimate the association between wartime HRV exposure, postwar HRV exposure, and their interaction on perceptions of corruption in local (Panel A), central (Panel B), and all institutions (Panel C). Corruption indices are constructed using Simple Average, PCA, and Factor Analysis. All models include municipality and year fixed effects. Covariates in columns (2), (4), and (6) include age, gender, education, employment status, urban residence, and homeownership. Sample sizes vary due to item nonresponse and index-specific coverage.

cross-sectional specification with individual covariates; standard errors in the true model are clustered by municipality. Because exposures are time-invariant at the municipality level, municipality or year fixed effects are not included in these cross-sectional placebo designs. Empirical p-values are computed from 10,000 permutations as the share of placebo statistics at least as extreme as the observed statistic.

Using binary indicators for above-median exposure shares, the conditional placebos reject the sharp null with signs consistent with the main results: permuting the civilian indicator yields positive associations with perceived corruption (local $p = 0.0039$, central $p = 0.0002$, all $p = 0.0001$), while permuting the armed-force indicator yields negative associations (local $p = 0.0017$, central $p = 0.0144$, all $p = 0.0014$). Figure 3.3 shows the placebo t-statistic distributions; summary estimates and p-values appear in Table 3.7[17]. By contrast, conditional placebos based on continuous exposure shares produce large p-values (0.22–0.58) across outcomes (Appendix Table C.5), suggesting either functional-form nonlinearity (threshold effects) or limited power for linear dose-response at the cross-municipality level. The main text, therefore, reports the binary design, which is less sensitive to heavy tails and outliers and aligns with the conceptual contrast between 'high' and 'low' exposure.

Table 3.7: Conditional Placebo Tests — Municipality-Level Exposure (Binary Indicators)

Outcome Variable	**Permuted Exposure (other held fixed)**	**True Coefficient**	**p-value**
Civilian casualties permuted (armed forces fixed)			
Corruption (Local Avg)	High Civilian Casualties	0.049	0.004
Corruption (Central Avg)	High Civilian Casualties	0.072	0.000
Corruption (All Avg)	High Civilian Casualties	0.063	0.000
Armed forces casualties permuted (civilian fixed)			
Corruption (Local Avg)	High Armed forces Casualties	-0.052	0.002
Corruption (Central Avg)	High Armed forces Casualties	-0.044	0.014
Corruption (All Avg)	High Armed forces Casualties	-0.051	0.001

Notes: Binary exposures equal one if the municipality's casualty share (civilian or armed forces) exceeds the sample median. Each row reports the coefficient from the original joint regression (including both binary exposures and individual covariates) and the permutation p-value from 10,000 replications that randomly reassign the *permuted* exposure across municipalities while keeping the *other* exposure fixed. Standard errors in the original model are clustered by municipality. Municipality or year fixed effects are not included in these cross-sectional placebo designs because exposures are time-invariant at the municipality level.

[17] Binary above-median indicators are reported in the main text for three reasons: (i) they are robust to heavy tails and outliers in municipality-level casualty shares; (ii) they directly capture the substantively relevant contrast between "high" and "low" exposure that underlies most theoretical mechanisms; and (iii) they deliver higher power when the true relationship is nonlinear or exhibits threshold effects. Per-capita shares (rather than raw counts) are used throughout to avoid mechanically confounding exposure with municipality size; counts disproportionately weight large municipalities and can induce spurious correlations. Continuous-share placebos are reported in the appendix for completeness; their large permutation p-values are consistent with limited linear dose-response at the cross-municipality level and do not overturn the binary-results pattern.

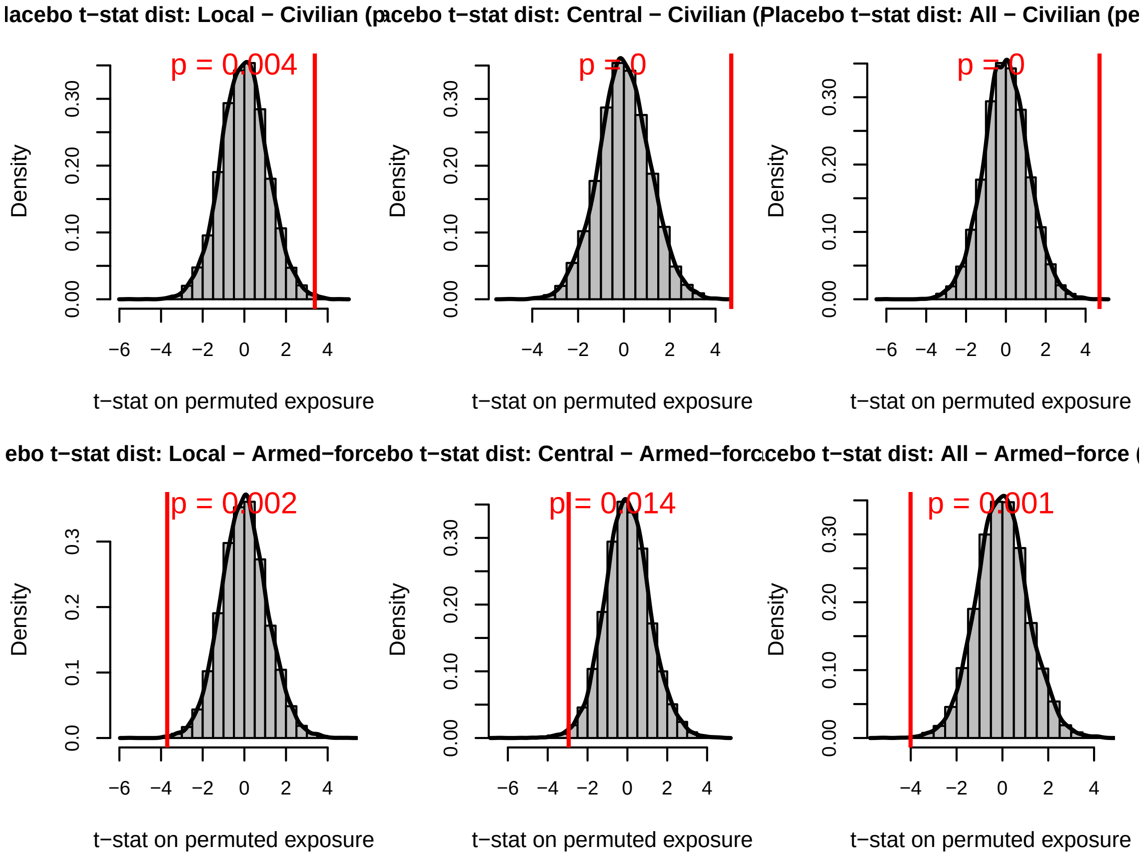


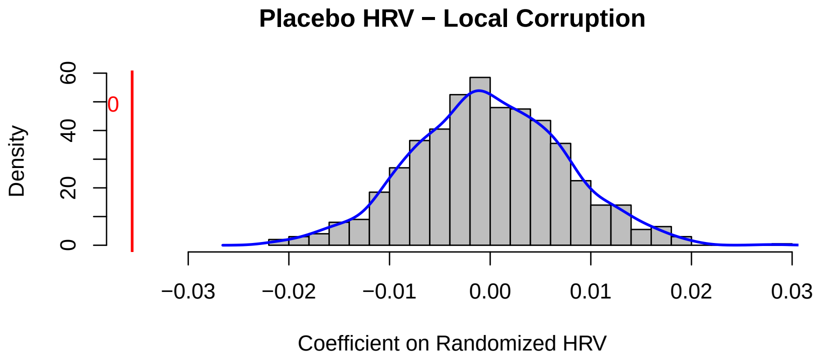
Figure 3.3: Conditional placebo distributions of t-statistics for municipality-level exposures (binary above-median indicators). Each panel shows the distribution from 10,000 permutations of one exposure across municipalities while holding the other fixed; the red line marks the observed t-statistic. The annotated p-value is the share of placebo statistics with absolute value at least as large as the observed statistic.

Table 3.8: Placebo Test Summary – Individual-Level HRV Exposure

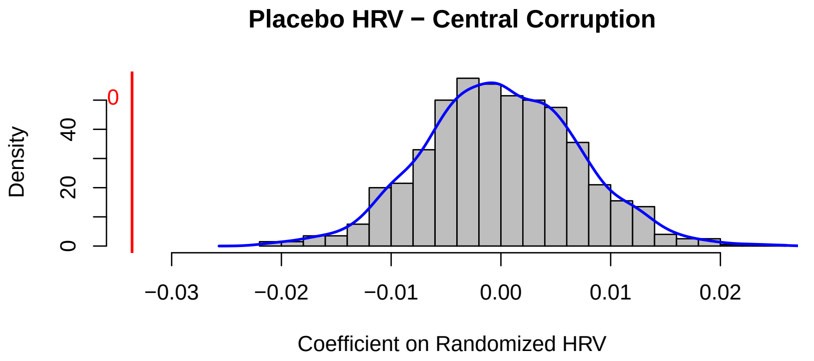
Outcome Variable	**True Coefficient**	***p*-value**
Corruption (Local Avg)	-0.031	0.000
Corruption (Central Avg)	-0.045	0.000
Corruption (All Avg)	-0.039	0.000

Note: This table reports the true coefficients from regressions of corruption perception indices on individual-level exposure to HRV, alongside placebo p-values based on 1,000 random permutations of the HRV variable. All models include individual-level controls and municipality and year fixed effects.

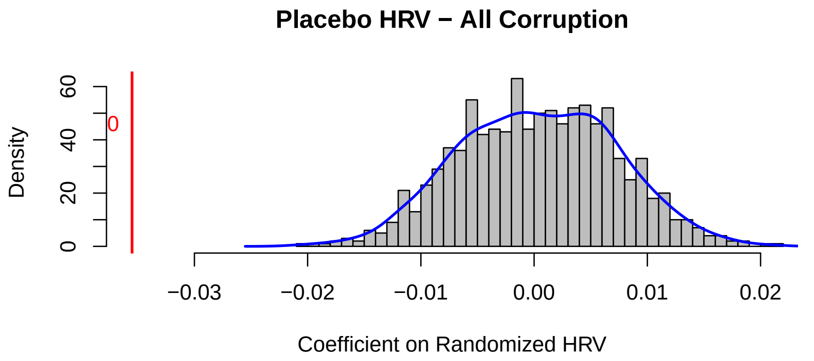
Figure 3.4: Placebo Distributions – Individual-Level HRV Exposure



(a) Local Corruption



(b) Central Corruption



(c) Overall Corruption

Note: Each panel shows the distribution of coefficients from 1,000 regressions using a randomly shuffled version of the HRV variable. The red vertical line represents the actual (true) coefficient from the data. The number printed next to the red line reports the empirical p-value, defined as the proportion of placebo estimates that are at least as extreme as the true coefficient.

Heterogeneity Analysis

To examine whether the association between conflict-related civilian casualties and corruption perceptions varies across population subgroups, interaction models are estimated with respect to education level, residential location, and gender. These models include interactions between the treatment variable (civilian casualties relative to 1990 municipal population) and subgroup indicators, and year fixed effects (no municipality fixed effects, since casualties are time-invariant). Marginal effects are then computed to illustrate differential impacts across categories.

Figure 3.5a presents results by education level. Among individuals with higher education, civilian casualties are associated with a positive and statistically significant increase in perceptions of central government corruption. For local corruption, the estimated marginal effect is negative in the high-education group and close to statistical significance, suggesting a potential divergence in how different tiers of government are perceived. Differences between low- and high-education groups are most pronounced for central corruption. These patterns are consistent with the possibility that individuals with greater educational attainment respond more strongly to national-level institutional shortcomings during and after conflict.

Figure 3.5b shows estimates disaggregated by residential area. For central government corruption, the effect of civilian casualties is positive and statistically significant in rural areas, whereas in urban settings, the estimate is smaller and not statistically distinguishable from zero. For local corruption, the effect is negative in rural areas but close to zero in urban areas, with both effects statistically insignificant. These results may reflect spatial variation in exposure intensity, institutional presence, or perceived state capacity.

Figure 3.5c examines heterogeneity by gender. Marginal effects on central corruption perceptions are similar for men and women, with overlapping confidence intervals indicating no statistically significant difference. For local corruption, both groups exhibit small negative effects, again with wide and overlapping confidence intervals. Ethnicity is not included in this section, as it is addressed separately in the confounding analysis (Section 3.5).

In addition to education, rural status, and gender, ethnicity significantly moderates the relationship between conflict exposure and corruption perceptions (Table C.6). For civilian casualties, the interaction term is negative and statistically significant for both local and central corruption perceptions, indicating that the positive association observed in the overall sample is attenuated—or even reversed—among Albanian respondents. In the case of armed forces casualties, interaction effects are large and negative, suggesting that Albanians in more heavily affected areas report substantially lower perceptions of corruption relative to non-Albanians, particularly at the central government level. For HRV, patterns are reversed: Albanian respondents who experienced wartime HRV report higher perceptions of corruption, while those exposed to post-conflict HRV report lower perceptions compared to non-Albanians. These results imply that ethnicity shapes not only the magnitude but also the direction of the association between conflict exposure and institutional trust, consistent with ethnically differentiated experiences of violence and post-conflict governance. Taken together, the heterogeneity analysis suggests that education and rural status moderate the relationship between conflict exposure and perceptions of institutional corruption, with stronger central government responses among more educated and rural respondents.

Interpretation

The findings reveal robust associations between wartime violence and long-term perceptions of corruption, with magnitudes that are substantively meaningful on the 0–1 perception scale.

Disaggregating violence types matters. Civilian victimization is associated with higher perceived corruption (especially for central institutions), while higher armed forces casualties are associated with lower perceived corruption at the local tier. Randomization-inference checks support these patterns for high-versus-low municipal exposure: conditional placebos with binary indicators reject the sharp null, whereas analogous placebos for continuous casualty shares yield large p-values. Accordingly, the evidence favors threshold-type contrasts over a linear dose–response. As the casualty models are cross-sectional, it is interpretable as a robust association rather than definitive causal effects, and explicitly assess their sensitivity to omitted-variable bias below.

By contrast, the results for individual-level exposure to HRV are robust to placebo checks and support a stronger causal interpretation. For HRV, wartime exposure is associated with lower perceived corruption, postwar exposure increases local perceived corruption, and combined exposure yields a net positive association.

Taken together, these results underscore the enduring legacy of conflict in shaping public perceptions of institutional legitimacy. The patterns highlight that the nature and target of violence, as well as individual characteristics such as education, rural residence, and ethnicity, condition the strength and direction of this relationship. Albanian respondents, in particular, exhibit attenuated or reversed effects for civilian and armed forces casualties, but heightened sensitivity to wartime and post-war HRV. These findings emphasize the need to consider both collective and individual experiences of violence—and the demographic and ethnic composition of affected communities—when designing governance reforms and rebuilding institutional trust in post-conflict societies.

Robustness and sensitivity to omitted variables

A remaining concern is that municipal exposure to wartime violence may proxy for predetermined unobservables—such as historical political alignment, pre-war institutional quality, or social capital—that also shape corruption perceptions. To assess the sensitivity of the main estimates to such omitted-variable bias, we implement the coefficient-stability approach proposed by Oster (2019). This method evaluates how strong selection on unobservables would need to be, relative to selection on observables, to eliminate the estimated effect, using changes in coefficients and explanatory power between restricted and fully controlled specifications.

Appendix Section C.9 reports the full set of Oster bounds. For each outcome–exposure pair, the table presents (i) the value of δ required to reduce the estimated coefficient to zero ($\delta \to 0$), and (ii) the bias-adjusted coefficient under $\delta = 1$, corresponding to equal selection on observables and unobservables. Larger values of $\delta \to 0$ indicate greater robustness, as they imply that unobserved factors would need to be at least as important as the included controls to overturn the estimated association.

Across specifications and index constructions, the sensitivity analysis reveals heterogeneous robustness. For local corruption perceptions, the estimated effect of civilian casualties is relatively stable: the value of δ required to drive the coefficient to zero is close to or above one across alternative index constructions, implying that unobservables would need to be at least as important as the included controls to fully explain away the association. For armed-force casualties, the corresponding δ values are smaller, indicating greater sensitivity; however, the bias-adjusted coefficients under equal selection on observables and unobservables remain negative and substantively meaningful.

By contrast, for central corruption perceptions, the estimates are more sensitive to omitted-variable concerns. In several specifications, the implied δ is close to zero or negative, reflecting limited additional explanatory power from controls and indicating that these results should be interpreted with greater caution. For the overall corruption index, the sensitivity measures lie between these two cases: while bias-adjusted coefficients preserve their sign, the required δ values suggest moderate sensitivity.

Taken together, the Oster bounds reinforce the interpretation that the relationship between wartime violence and local institutional perceptions—particularly for civilian victimization—is comparatively robust, whereas estimates for central institutions remain more exposed to omitted-variable bias. Full results and implementation details are reported in Appendix Section C.9.

The Moderating Role of Elections on the Conflict–Corruption Link

This subsection investigates how the relationship between wartime violence exposure and perceptions of local corruption evolves around electoral periods. Using an event study design, multiple parliamentary and local elections are pooled over time and construct event-time dummies that capture each municipality's relative position in the electoral cycle—specifically, one year before, the election year, and one year after the election. The models control for demographic and socioeconomic factors as well as municipality and year fixed effects to absorb unobserved heterogeneity and time shocks.

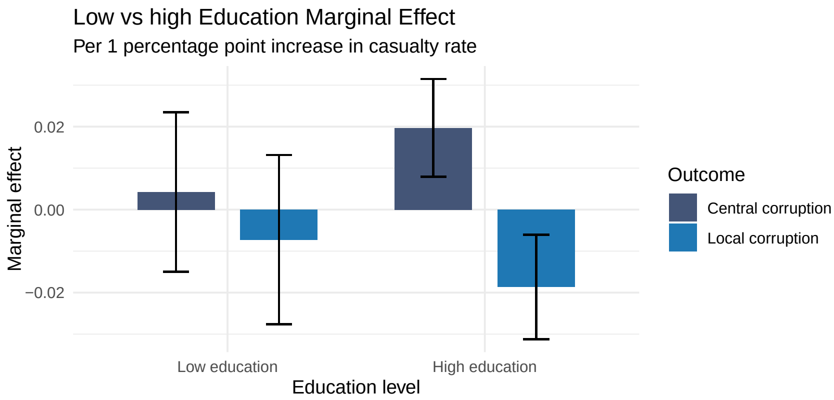
Table 3.9 reports the estimated coefficients. Because casualties are time-invariant, the event-study models identify from interactions between casualty exposure and election-timing indicators in the presence of municipality and year fixed effects. The interaction terms show that electoral periods attenuate the baseline association: positive and statistically significant interactions in the year before and during parliamentary elections, and in the year before and after local elections, indicate that perceived corruption rises more in high-exposure municipalities around elections. A notable exception is a negative interaction for combatant casualties in local-election years, suggesting a distinct dynamic where combatant losses are salient. Overall, elections appear to be inflection points that temporarily amplify corruption concerns where wartime exposure was high. In summary, the dynamics linking conflict exposure to corruption perceptions are cyclical and politically sensitive rather than linear. Post-conflict institutional trust appears fragile and subject to erosion during key political moments, underscoring the importance of accounting for political timing in studies of long-term governance perceptions.

Table 3.9: Event Study: Dynamic Effects of Conflict Exposure on Local Corruption Perception Around Elections

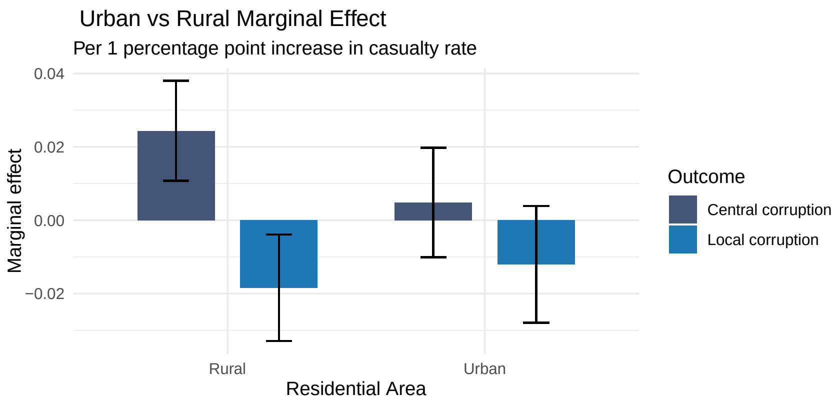
Variable	**Civilian Casualties**	**Armed Forces Casualties**
Casualties	-8.722***	-11.42$^{\cdot}$
	(2.393)	(6.475)
Parl. election: 1 year before	-0.0200	-0.0021
	(0.0134)	(0.0120)
Parl. election: election year	-0.0102	0.0027
	(0.0127)	(0.0113)
Parl. election: 1 year after	-0.0108	-0.0116
	(0.0122)	(0.0109)
Local election: 1 year before	-0.0673***	-0.0480***
	(0.0091)	(0.0081)
Local election: election year	-0.0552***	-0.0344***
	(0.0110)	(0.0098)
Local election: 1 year after	-0.0447***	-0.0234**
	(0.0124)	(0.0112)
Casualties × Parl. election: 1 year before	6.044**	6.025
	(2.279)	(6.171)
Casualties × Parl. election: election year	5.752**	8.247$^{\cdot}$
	(2.179)	(5.811)
Casualties × Parl. election: 1 year after	2.491	8.680$^{\cdot}$
	(2.059)	(5.394)
Casualties × Local election: 1 year before	7.482***	9.877**
	(1.577)	(4.122)
Casualties × Local election: election year	1.481	-11.33**
	(1.910)	(5.100)
Casualties × Local election: 1 year after	8.583***	11.70**
	(2.110)	(5.729)
N obs	26,462	26,462

Note: Standard errors are clustered at the municipality level. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, $^{\cdot}p < 0.15$. This table presents estimates from an event-study analysis investigating how the relationship between conflict exposure and perceptions of local corruption evolves around parliamentary and local election periods. The dependent variable is the local corruption perception index, where higher values indicate greater perceived corruption (lower trust in local government).
Conflict exposure is measured as civilian and armed forces casualties per capita, scaled using 1990 municipality population figures. The main coefficient on casualties reflects the association in non-election years: notably, higher civilian casualties are associated with significantly lower perceived corruption, while armed forces casualties exhibit an even stronger negative association.
Interaction terms capture how this relationship changes in proximity to election years. Positive and statistically significant interactions with parliamentary and local elections suggest that electoral periods weaken the baseline "trust effect" of conflict exposure. This pattern is especially pronounced in the years immediately before and after local elections, indicating that electoral cycles temporarily heighten corruption perceptions in municipalities heavily affected by wartime violence.
All regressions control for demographic and socioeconomic covariates (age, gender, schooling, urban residence, employment status, home ownership) as well as municipality and year fixed effects. Identification comes from interactions between time-invariant casualties and election-timing indicators.

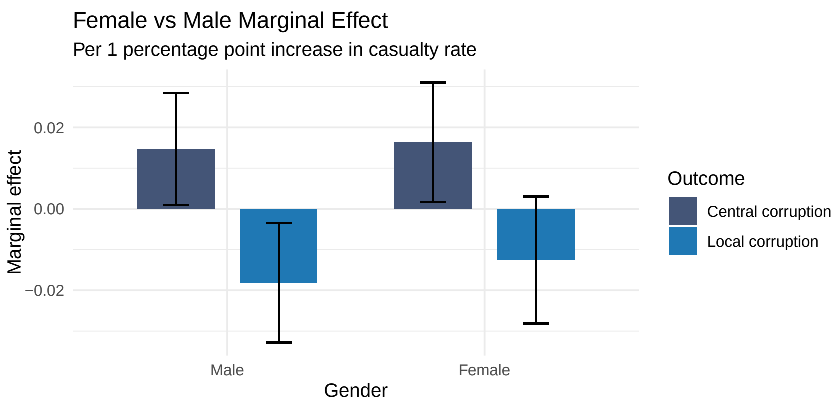
Figure 3.5: Heterogeneous effects by subgroup



(a) By education



(b) Urban vs rural



(c) By gender

3.7 Conclusion

This chapter examines the long-run relationship between wartime violence and perceptions of corruption in post-conflict Kosovo, combining municipality-level casualty data with individual reports of HRV. Casualty exposures are normalized by the 1990 municipal population and enter jointly; HRV models exploit within-municipality, individual-level variation with municipality and year fixed effects. The evidence points to differentiated associations across exposure types and institutional tiers, with implications for theories of institutional trust and for post-conflict governance.

Civilian casualties are positively associated with perceived corruption, especially for central institutions, with smaller but still positive associations for local institutions. By contrast, armed forces casualties are negatively associated with perceived corruption at the local level and weakly negative for central institutions once civilian victimization is held constant. Conditional randomization-inference placebos that permute one exposure while holding the other fixed yield small permutation p-values (typically below 0.02) across outcomes, indicating that the joint patterns are unlikely to arise under random spatial assignment of exposure. Given the cross-sectional design and potential residual spatial confounding, these estimates are interpreted as robust associations rather than definitive causal effects.

Results based on HRV, identified from within-municipality variation, display a complementary pattern. HRV during the conflict is negatively associated with perceived corruption for both local and central institutions. Post-conflict HRV is positively associated with local perceived corruption and close to zero for central institutions. The interaction between wartime and post-war HRV is positive and statistically significant, implying that cumulative exposure restores the positive association observed in the casualty regressions. Ethnicity operates as an effect modifier rather than a confounder: several estimates are concentrated among non-Albanian minorities, whereas combined main and interaction effects for Albanian respondents are often small.

These conclusions are robust to alternative index constructions. HRV findings are consistent across the simple average and PCA/FA indices (with PCA/FA rescaled to match the mean and standard deviation of the simple average), reinforcing measurement coherence. By contrast, municipality-level casualty estimates remain more sensitive to spatial structure, warranting cautious interpretation.

From a policy perspective, the results highlight the importance of interventions that combine trauma-informed services with institutional reforms and that are tailored to heterogeneity across institutional tiers and communities. Distinguishing between civilian and armed forces victimization—and recognizing their distinct legacies for local versus central institutions—appears critical for rebuilding trust.

Further work should investigate mechanisms—psychological trauma, attribution, civic engagement, and transitional justice—and dynamics using designs with sharper identification (e.g., quasi-experimental exposure and panels). Understanding how historical violence shapes public trust is central to the design of durable institutions in fragile settings.

Limitations

Several limitations merit emphasis. First, in the absence of exogenous variation in wartime exposure, municipality-level estimates are vulnerable to spatial confounding; randomization-inference placebos support the presence of systematic associations but do not establish causality. Second, both perceived corruption and HRV are self-reported and may reflect recall or reporting biases. Third, the analysis documents patterns rather than mechanisms; psychological, political, and economic channels are not identified separately. Fourth, external validity is constrained by the specificity of the Kosovo case. Finally, results can vary with index construction and covariate inclusion; robustness checks mitigate but cannot eliminate these concerns. Future research exploiting quasi-experimental designs or longitudinal data could isolate mechanisms and dynamics more sharply.

Conclusions

This thesis has examined how crime, corruption, and conflict—often interconnected and mutually reinforcing—shape institutional behaviour, public goods provision, and citizen trust. Across three distinct settings, the analyses show that violence, whether as a direct act or as a structural condition, operates as a persistent force that can distort governance long after the events themselves have passed.

The evidence from the United States illustrates that the design of administrative and legal frameworks can influence not only institutional performance but also the visibility of wrongdoing. In death investigation systems where elected sheriffs hold investigative authority, police killings are substantially more likely to be misclassified or go unreported. This form of strategic bureaucratic opacity demonstrates that even in established democracies, institutional rules can be configured in ways that undermine accountability and transparency in crime measurement.

The analysis of post-war Kosovo reveals how the physical and institutional destruction caused by conflict can shape service delivery and perceptions of governance for years. Heavily bombed municipalities hosting post-war waste landfills experienced both higher infant mortality near these sites and greater perceptions of local corruption. These findings highlight the dangers of reconstruction processes that prioritise physical rebuilding over institutional integrity, and they illustrate how governance weaknesses can persist and compound the human costs of conflict.

The study of wartime legacies for institutional trust in Kosovo shows that the type of violence matters. Civilian-targeted violence was associated with higher perceived corruption, particularly toward local government, while armed group casualties were linked to lower perceived corruption. Personal experiences of human rights violations emerged as especially powerful in shaping long-run perceptions, suggesting that the political consequences of violence are mediated by both its form and its proximity to the individual.

Three common themes emerge. First, institutional design matters for accountability: the rules that govern administrative processes can either enable or constrain misconduct.

Second, post-conflict governance requires more than physical reconstruction; without safeguards, it risks perpetuating the very weaknesses that fuelled instability. Third, the legacies of violence are heterogeneous, with effects that depend on both the nature of the violence and the roles of the actors involved.

These findings have implications for policy in both developed and post-conflict settings. Administrative frameworks should be structured to prevent conflicts of interest and to ensure transparency in reporting and oversight. Post-conflict reconstruction must embed institutional safeguards into infrastructure and service delivery projects. Efforts to rebuild trust should account for the differentiated experiences of violence across communities, recognising that uniform approaches are unlikely to succeed.

The lessons extend beyond the cases studied here. Current conflicts in Ukraine, Gaza, Sudan, and other regions will leave behind institutional fragility, contested legitimacy, and fractured public trust. The design of post-conflict institutions and the governance of reconstruction will determine whether peace endures or whether instability re-emerges.

The arguments developed here echo long-standing insights from political thought. Hobbes warned that without strong institutions, life becomes 'nasty, brutish, and short'. Hume emphasised that authority depends on conventions that can be undone by mistrust. Machiavelli observed that stability requires both the cunning of the fox and the strength of the lion. Kadare, chronicling the Balkans, reminds us that 'the past is never dead; it is not even past'. These insights remain as relevant today as in their time: the challenge is not only to end violence, but to ensure that the institutions emerging from it constrain its recurrence rather than perpetuate its harms.

Bibliography

Abrahams, F. (2001). *Under Orders*. Human Rights Watch. 180, 181

Agjencia për Mbrojtjen e Mjedisit të Kosovës (2018). Menaxhimi i mbeturinave komunale në kosovë: Raport mbi gjendjen. Accessed: 2024-06-26. 42, 45, 181, 182

Aiello, F. and V. Pupo (2012). Structural funds and the economic divide in italy. *Journal of Policy Modeling 34*(3), 403–418. 40

Alesina, A. and E. La Ferrara (2005). Ethnic diversity and economic performance. *Journal of economic literature 43*(3), 762–800. 88, 92

Alite, M., H. Abu-Omar, M. T. Agurcia, M. Jácome, J. Kenney, A. Tapia, and M. Siebel (2023). Construction and demolition waste management in kosovo: a survey of challenges and opportunities on the road to circular economy. *Journal of Material Cycles and Waste Management 25*(2), 1191–1203. 182

Ang, D. (2021). The effects of police violence on inner-city students. *The Quarterly Journal of Economics 136*(1), 115–168. 10

Ang, D., P. Bencsik, J. Bruhn, and E. Derenoncourt (2021). Police violence reduces civilian cooperation and engagement with law enforcement. Technical report, HKS Working Paper No. RWP21-022. 10

Ang, D. and J. Tebes (2020). Civic responses to police violence. Technical report, HKS Working Paper No. RWP20-033. 10

Arceo, E., R. Hanna, and P. Oliva (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from mexico city. *The Economic Journal 126*(591), 257–280. 39

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine 28*(25), 3083–3107. 105, 107

Baicker, K. and M. Jacobson (2007). Finders keepers: Forfeiture laws, policing incentives, and local budgets. *Journal of Public Economics 91*(11-12), 2113–2136. 9

Baker, A., B. Callaway, S. Cunningham, A. Goodman-Bacon, and P. H. Sant'Anna (2025). Difference-in-differences designs: A practitioner's guide. *arXiv preprint arXiv:2503.13323*. 57

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2021). Improving police performance in rajasthan, india: Experimental evidence on incentives, managerial autonomy, and training. *American Economic Journal: Economic Policy 13*(1), 36–66. 9

Becker, G. S. and G. J. Stigler (1974). Law enforcement, malfeasance, and compensation of enforcers. *The Journal of Legal Studies 3*(1), 1–18. 9

Bellows, J. and E. Miguel (2009). War and local collective action in sierra leone. *Journal of public Economics 93*(11-12), 1144–1157. 85, 86, 91

Benoît, J.-P. and J. Dubra (2004). Why do good cops defend bad cops? *International Economic Review 45*(3), 787–809. 11

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics 119*(1), 249–275. 25

Blattman, C. (2009). From violence to voting: War and political participation in uganda. *American political Science review 103*(2), 231–247. 91

Bor, J., A. S. Venkataramani, D. R. Williams, and A. C. Tsai (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. 10

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of econometrics 225*(2), 200–230. 58, 59, 61, 188

Campbell, T. (2022). Black lives matter's effect on police lethal use of force. *Available at SSRN 3767097*. 11, 174

Chen, J. and J. Roth (2024). Logs with zeros? Some problems and solutions. *The Quarterly Journal of Economics 139*(2), 891–936. 21

Cheng, C. and W. Long (2018). Improving police services: Evidence from the french quarter task force. *Journal of Public Economics 164*, 1–18. 9

Collier, P. et al. (2003). *Breaking the conflict trap: Civil war and development policy*, Volume 41181. World Bank Publications. 85

Comey, M. L., A. R. Eng, and Z. Pei (2022). Supercompliers. *arXiv preprint arXiv:2212.14105*. 8, 28, 30

Conner, A., D. Azrael, V. H. Lyons, C. Barber, and M. Miller (2019). Validating the national violent death reporting system as a source of data on fatal shootings of civilians by law enforcement officers. *American journal of public health 109*(4), 578–584. 11, 12

Cook, S. J. and D. Fortunato (2023). The politics of police data: State legislative capacity and the transparency of state and substate agencies. *American Political Science Review 117*(1), 280–295. 6, 9, 13, 23

Currie, J. and M. Neidell (2005). Air pollution and infant health: what can we learn from california's recent experience? *The quarterly journal of economics 120*(3), 1003–1030. 39

Currie, J. and R. Walker (2011). Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal: Applied Economics 3*(1), 65–90. 48, 49

Currie, J., J. G. Zivin, J. Mullins, and M. Neidell (2014). What do we know about short- and long-term effects of early-life exposure to pollution? *Annu. Rev. Resour. Econ. 6*(1), 217–247. 39

Cutts, M. (2000). The state of the world's refugees, 2000: Fifty years of humanitarian action. 42

De Angelis, I., G. de Blasio, and L. Rizzica (2020). Lost in corruption. evidence from eu funding to southern italy. *Italian Economic Journal 6*(2), 355–377. 40

De Juan, A. and J. H. Pierskalla (2016). Civil war violence and political trust: Microlevel evidence from nepal. *Conflict Management and Peace Science 33*(1), 67–88. 91

De Luca, G. and M. Verpoorten (2015). Civil war, social capital and resilience in uganda. *Oxford economic papers 67*(3), 661–686. 86

Desmond, M., A. V. Papachristos, and D. S. Kirk (2016). Police violence and citizen crime reporting in the black community. *American Sociological Review 81*(5), 857–876. 10

Devi, T. and R. G. Fryer Jr (2020). Policing the police: The impact of "pattern-or-practice" investigations on crime. Technical report, National Bureau of Economic Research. 9, 11, 174

Dorff, C. (2017). Violence, kinship networks, and political resilience: Evidence from mexico. *Journal of Peace Research 54*(4), 558–573. 85, 91

Dube, A., T. W. Lester, and M. Reich (2010). Minimum wage effects across state borders: Estimates using contiguous counties. *The Review of Economics and Statistics 92*(4), 945–964. 19

Edwards, F., M. H. Esposito, and H. Lee (2018). Risk of police-involved death by race/ethnicity and place, united states, 2012–2018. *American Journal of Public Health 108*(9), 1241–1248. 9, 10

Edwards, F., H. Lee, and M. Esposito (2020). Risk of being killed by police use-of-force in the us by age, race/ethnicity, and sex. *Journal of Proceedings of the National Academy of Sciences 116*(34), 1–12. 9

Euroepan Commission (2020). Commission staff working document — kosovo* 2020 report (swd(2020) 356 final, brussels, 6.10.2020). *Commission Staff Working Document*. 183, 184

European Commission (1999). European agency for reconstruction set up for kosovo. `https://europa.eu/rapid/press-release_IP-99-411_en.htm`. 44

European Court of Auditors (2001). Special report on the european agency for reconstruction (kosovo), with the agency's and commission's replies. `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C:2001:355:FULL`. 44

European Environment Agency (2020, October). Municipal waste management in kosovo: Country fact sheet. Technical report. 42

European Environment Agency (2021). Municipal waste management in western balkans countries - country profile kosovo. *Report of European Environment Agency*. 42, 45, 182, 183, 184

European Parliament (2002). Texts adopted, 10 april 2002 — european agency for reconstruction. `https://www.europarl.europa.eu/doceo/document/TA-5-2002-04-10_EN.html`. 44

European Union (2001). Official journal of the european communities, c 355, 2001. Accessed: 2024-06-24. 44

European Union (2006). Summary of eu legislation: European agency for reconstruction. Accessed: 2024-06-24. 44

European Union News Bulletin Oct (2002). European Agency for Reconstruction Kosovo News Bulletin, Oct 2002. Accessed: 2024-06-24. 44

Fearon, J. D. and D. D. Laitin (2003). Ethnicity, insurgency, and civil war. *American political science review 97*(1), 75–90. 85

Feldman, J. M., S. Gruskin, B. A. Coull, and N. Krieger (2017). Quantifying underreporting of law-enforcement-related deaths in united states vital statistics and news-media-based data sources: A capture–recapture analysis. *PLoS medicine 14*(10), e1002399. 10, 12, 13

Fernandes, P. (2011). Transitional justice in balance: Comparing processes, weighing efficacy. *Prisma Juridico 10*, 527. 85, 93

Fernandez, J. M. (2019). The political economy of death: Do coroners perform as well as medical examiners in determining suicide? *Available at SSRN 3360656*. 13

Frangakis, C. E. and D. B. Rubin (2002). Principal stratification in causal inference. *Biometrics 58*(1), 21–29. 8, 28, 30

Fryer Jr, R. G. (2019). An empirical analysis of racial differences in police use of force. *Journal of Political Economy 127*(3), 1210–1261. 9

Gallego, J. (2018). Civil conflict and voting behavior: Evidence from colombia. *Conflict Management and Peace Science 35*(6), 601–621. 85, 91

Gates, S. and M. K. Justesen (2020). Political trust, shocks, and accountability: Quasi-experimental evidence from a rebel attack. *Journal of Conflict Resolution 64*(9), 1693–1723. 92

Gavazza, A. and A. Lizzeri (2007). The perils of transparency in bureaucracies. *American Economic Review 97*(2), 300–305. 11, 174

GBD et al. (2021). Fatal police violence by race and state in the usa, 1980–2019: a network meta-regression. *The Lancet 398*(10307), 1239–1255. 9, 10, 13, 15, 170

Goodman, S., A. BenYishay, Z. Lv, and D. Runfola (2019). Geoquery: Integrating hpc systems and public web-based geospatial data tools. *Computers & geosciences 122*, 103–112. 52

Grant, R. (1999). *The Kosovo campaign: aerospace power made it work.* Air Force Association. 38

Greenstone, M. and R. Hanna (2014). Environmental regulations, air and water pollution, and infant mortality in india. *American Economic Review 104*(10), 3038–3072. 39

Guarnieri, E. (2025). Cultural distance and ethnic civil conflict. *American Economic Review 115*(4), 1338–1368. 86

Guriev, S. and D. Treisman (2019). Informational autocrats. *Journal of economic perspectives 33*(4), 100–127. 9

Hadzic, D. and M. Tavits (2019). The gendered effects of violence on political engagement. *The Journal of Politics 81*(2), 676–680. 85, 91

Hanzlick, R. (2007). The conversion of coroner systems to medical examiner systems in the united states: a lull in the action. *The American Journal of Forensic Medicine and Pathology 28*(4), 279–283. 12, 13, 17, 22

Hanzlick, R. L. (2014). A synoptic review of the 1954 "model postmortem examinations act". *Academic Forensic Pathology 4*(4), 451–454. 13, 142, 143

He, G. and J. M. Perloff (2016). Surface water quality and infant mortality in china. *Economic Development and Cultural Change 65*(1), 119–139. 39

Henkin, L. (1999). Kosovo and the law of "humanitarian intervention". *American Journal of International Law 93*(4), 824–828. 179

Hobbes, T. (1651). *Leviathan.* London: Andrew Crooke. 1

House of Lords European Union Committee (2002). Eu economic and monetary union: Government response to the committee's report. Accessed: 2024-06-24. 44

Hui, C., Q. Yu, B. Liu, M. Zhu, Y. Long, and D. Shen (2023). Microbial contamination risk of landfilled waste with different ages. *Waste Management 170*, 297–307. 65

Hume, D. (1739). *A Treatise of Human Nature.* London: John Noon. 1

ICTY, U. (2000). Final report to the prosecutor by the committee established to review the nato bombing campaign against the federal republic of yugoslavia. *Pre-2008 Archives*. 38

Ishiyama, J., F. C. Betancourt Higareda, A. Pulido, and B. Almaraz (2018). What are the effects of large-scale violence on social and institutional trust? using the civil war literature to understand the case of mexico, 2006–2012. *Civil Wars 20*(1), 1–23. 92

Judah, T. (2002). *Kosovo: War and revenge*. Yale University Press. 98, 103

Jusufi, I. (2021). The death of the european agency for reconstruction: A peculiar case of termination of the agencies of the european union. *Hrvatska i komparativna javna uprava: časopis za teoriju i praksu javne uprave 21*(4), 675–706. 44

Kadare, I. (1990). *Broken April*. New York: New Amsterdam Books. Originally published in 1978. 3

Kadare, I. (1993). *The Palace of Dreams*. New York: Arcade Publishing. Originally published in 1981. 3

Kadare, I. (2000). *Elegy for Kosovo*. Arcade Publishing. 38

Kijewski, S. and M. Freitag (2018). Civil war and the formation of social trust in kosovo: Posttraumatic growth or war-related distress? *Journal of conflict resolution 62*(4), 717–742. 86

Kosovo Environmental Protection Agency (2019). Report on municipal waste 2019. Accessed: 2024-06-24. 181

Kosovo Environmental Protection Agency (2021). Questionnaire to western balkan countries for providing information regarding municipal waste management and waste prevention follow-up interview kosovo. 45, 183, 184

Kreutz, J. and E. Nussio (2019). Destroying trust in government: Effects of a broken pact among colombian ex-combatants. *International Studies Quarterly 63*(4), 1175–1188. 92

Krieger, H. (2001). *The Kosovo conflict and international law: an analytical documentation 1974-1999*. Cambridge University Press. 39, 98

Kritsiotis, D. (2000). The kosovo crisis and nato's application of armed force against the federal republic of yugoslavia. *International and Comparative Law Quarterly 49*(2), 330–359. 179

Ledeneva, A. (2017). Where does informality stop and corruption begin? informal governance and the public/private crossover in mexico, russia and tanzania. *Slavonic and East European Review 95*(1), 49–75. 85

Leflar, R. (1955). Drafting the model post-mortem examinations act. *ABAJ 41*, 266. 13

Loyle, C. E. and B. J. Appel (2017). Conflict recurrence and postconflict justice: Addressing motivations and opportunities for sustainable peace. *International Studies Quarterly 61*(3), 690–703. 85, 93

Lum, C., C. S. Koper, D. B. Wilson, M. Stoltz, M. Goodier, E. Eggins, A. Higginson, and L. Mazerolle (2020). Body-worn cameras' effects on police officers and citizen behavior: A systematic review. *Campbell Systematic Reviews 16*(3). 172

Luzi, S. A., J. Melinek, and W. R. Oliver (2013). Medical examiners' independence is vital for the health of the american legal system. *Academic Forensic Pathology 3*(1), 84–92. 6, 8, 14, 36

Machiavelli, N. (1532). *The Prince*. Florence: Antonio Blado d'Asola. 1

Martinez, L. R. (2022). How much should we trust the dictator's gdp growth estimates? *Journal of Political Economy 130*(10), 2731–2769. 9

Mastrorocco, N., A. Ornaghi, et al. (2020). *Who watches the watchmen? Local news and police behavior in the United States*. University of Warwick, Department of Economics. 9

Melinek, J., L. C. Thomas, W. R. Oliver, G. A. Schmunk, V. W. Weedn, and N. A. of Medical Examiners Ad Hoc Committee on Medical Examiner Independence (2013). National association of medical examiners position paper: Medical examiner, coroner, and forensic pathologist independence. *Academic Forensic Pathology 3*(1), 93–98. 6, 8, 14, 36

Ministry of Environment and Spatial Planning (2013). Strategy of the republic of kosovo on waste management 2013–2022. Technical report, Government of Kosovo, Pristina. 42

Ministry of Environment and Spatial Planning (2019). Ministry of environment and spatial planning, 2013, strategy of the republic of kosovo on waste management: 2013-2022. *Kosovo Ministry of Environment and Spatial Planning, Pristina*. 45, 182, 183, 184

Montalvo, J. G. and M. Reynal-Querol (2005). Ethnic polarization, potential conflict, and civil wars. *American economic review 95*(3), 796–816. 88, 92

Morina, I., N. Bajraktari, R. Morina, S. Shala, and T. Veselaj (2017). The research of illegal dumps in some parts of the territory of kosovo. *WSEAS Transactions on Environment and Development 13*, 2016–224. 182

Morris, K. and K. Shoub (2023). Contested killings: The mobilizing effects of community contact with police violence. *American Political Science Review*. 10

Newman, E. and G. Visoka (2024). Nato in kosovo and the logic of successful security practices. *International Affairs 100*(2), 631–653. 179

Nix, J., B. A. Campbell, E. H. Byers, and G. P. Alpert (2017). A bird's eye view of civilians killed by police in 2015: Further evidence of implicit bias. *Criminology & Public Policy 16*(1), 309–340. 9

Olzak, S. (2021). Does protest against police violence matter? evidence from us cities, 1990 through 2019. *American Sociological Review 86*(6), 1066–1099. 10

OSCE (1999). *Kosovo/Kosova as Seen, as Told: An Analysis of the Human Rights Findings of the OSCE Kosovo Verification Mission October 1998 to June 1999*, Volume 2. OSCE Office for Democratic Institutions and Human Rights. 180

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204. 118, 212

Owens, E. and B. Ba (2021). The economics of policing and public safety. *Journal of Economic Perspectives 35*(4), 3–28. 9, 10

Paulozzi, L., J. Mercy, L. Frazier, and J. Annest (2004). Cdc's national violent death reporting system: background and methodology. *Injury Prevention 10*(1), 47–52. 11, 37

Prados, M. J., T. Baker, A. N. Beck, D. B. Burghart, R. R. Johnson, D. Klinger, K. Thomas, and B. K. Finch (2022). Do sheriff-coroners underreport officer-involved homicides? *Academic Forensic Pathology 12*(4), 140–148. 9, 13

Prendergast, C. (2003). The limits of bureaucratic efficiency. *Journal of Political Economy 111*(5), 929–958. 11, 174

Rad, A. N., D. S. Kirk, and W. P. Jones (2023). Police unionism, accountability, and misconduct. *Annual Review of Criminology 6*, 181–203. 9, 11

Rahman, D. (2012). But who will monitor the monitor? *American Economic Review 102*(6), 2767–2797. 9, 11

Ramboll, E. and M. Insight (2009). Evaluation of the eu decentralised agencies in 2009. 44

Robinson, A. L. (2016). Nationalism and ethnic-based trust: Evidence from an african border region. *Comparative Political Studies 49*(14), 1819–1854. 92

Rose-Ackerman, S. (2008). Corruption and government. *International peacekeeping 15*(3), 328–343. 86, 92

Rosenbaum, P. R. and D. B. Rubin (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician 39*(1), 33–38. 105, 107

Ruiz, L., B. Posey, M. Neuilly, M. Stohr, and C. Hemmens (2018). Certifying death in the united states. *Journal of Forensic Sciences 63*(4), 1138–1145. xiii, 12, 14, 17, 19, 23, 149

Schwartz, G. L. and J. L. Jahn (2020). Mapping fatal police violence across us metropolitan areas: Overall rates and racial/ethnic inequities, 2013-2017. *PloS one 15*(6), e0229686. 9, 10

Sherman, L. W. (1978). *Scandal and reform: Controlling police corruption*. Univ of California Press. 5, 9

Shi, L. (2008). Does oversight reduce policing? evidence from the cincinnati police department after the april 2001 riot. *Journal of Public Economics*. 11, 174

Spector, B. I. (2005). *Fighting corruption in developing countries: Strategies and analysis*. Kumarian Press Bloomfield, CT. 85

Strezhnev, A. (2023). Decomposing triple-differences regression under staggered adoption. *arXiv preprint arXiv:2307.02735*. 57

Tennenbaum, A. N. (1994). The influence of the garner decision on police use of deadly force. *J. Crim. L. & Criminology 85*, 241. 151

Threedeach, S., W. Chiemchaisri, T. Watanabe, C. Chiemchaisri, R. Honda, and K. Yamamoto (2012). Antibiotic resistance of escherichia coli in leachates from municipal solid waste landfills: comparison between semi-aerobic and anaerobic operations. *Bioresource technology 113*, 253–258. 65

Trako, I. (2018). Returning home after conflict displacement: Labor supply and schooling outcomes among kosovar households. *Available at SSRN 2838200*. 38

Treisman, D. (2000). The causes of corruption: a cross-national study. *Journal of public economics 76*(3), 399–457. 86, 92

Tyler, T. R., J. Fagan, and A. Geller (2014). Street stops and police legitimacy: Teachable moments in young urban men's legal socialization. *Journal of Empirical Legal Studies 11*(4), 751–785. 10

UNHCR (2000). Human rights watch, kosovo: Civilian deaths in the nato air campaign. *https://www.refworld.org/reference/countryrep/hrw/2000/en/32567*. 39, 98

UNICEF (2017). Safely managed drinking water. Technical report, UNICEF and World Health Organization. 65

UNICEF (2024, September). Geographic data in mics: Anonymisation and dissemination. 75

UNICEF Kosovo Programme (2020). Kosovo multiple indicator cluster survey 2020. Accessed: 2024-06-24. 65

UNODC (2022). Corruption in kosovo: Thematic report. 40, 101

U.S. Government (1999). Ethnic cleansing in kosovo, fact sheet based on information from u.s. government sources. *https://1997-2001.state.gov/regions/eur/rpt_990422_ksvo_ethnic.html*. 41

Von der Goltz, J. and P. Barnwal (2019). Mines: The local wealth and health effects of mineral mining in developing countries. *Journal of Development Economics 139*, 1–16. 39

Voytas, E. and B. Crisman (2024, November). State violence and participation in transitional justice: Evidence from Colombia. *Journal of Peace Research 61*(6), 1069–1084. 85, 86, 91

Wang, Y. (2021). The political legacy of violence during china's cultural revolution. *British Journal of Political Science 51*(2), 463–487. 91

Williams Jr, M. C., N. Weil, E. A. Rasich, J. Ludwig, H. Chang, and S. Egrari (2021). Body-worn cameras in policing: Benefits and costs. Technical report, National Bureau of Economic Research. 172

World Bank (2018). Water security outlook for kosovo. Technical report. 65

Youngs, Oakes, Bowers, and M. Hillyard (1999). Kosovo: Operation" allied force". *Parliament, House of Commons Library*. 180

Zaiour, R. and M. Mikdash (2023). The impact of police shootings on gun violence and civilian cooperation. Technical report, Available at SSRN 4390262. 10

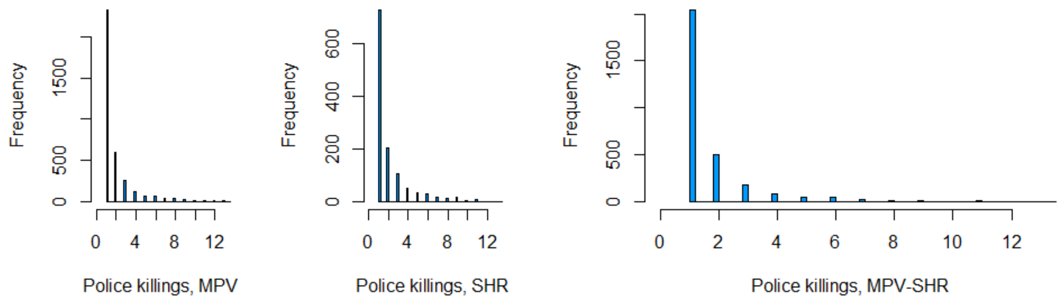
Zimring, F. E. (2017). *When police kill.* Harvard University Press. 6, 16

Appendix A

Appendix to Chapter 1

APPENDIX

A.1 Histograms: MPV and SHR Police Killings



Note: only positive values displayed for MPV-SHR

A.2 Maps: MPV and SHR Police Killings

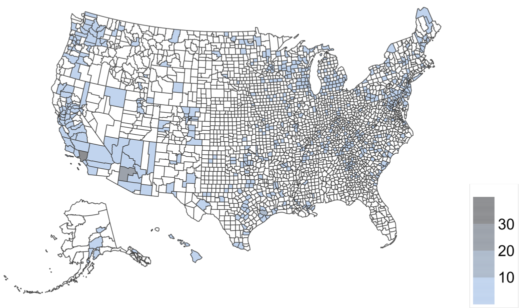


Figure A.1: SHR mean police killings

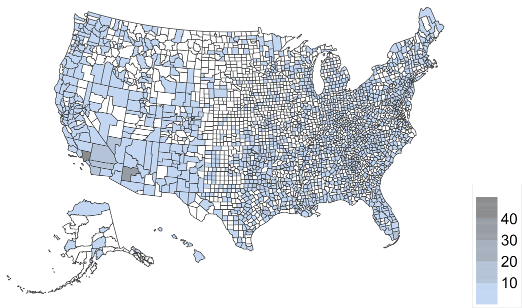


Figure A.2: MPV mean police killings

A.3 Maps: Additional Differences in Death Investigation Systems

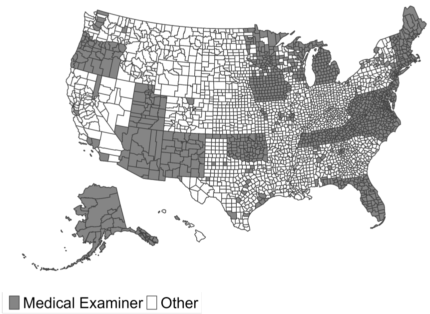


Figure A.3: Medical Examiner System

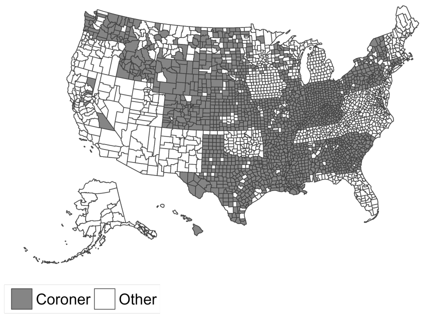


Figure A.4: Coroner System

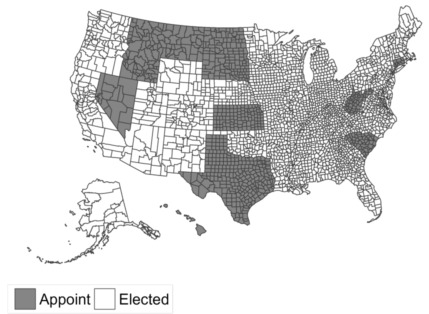


Figure A.5: Appointed vs Elected

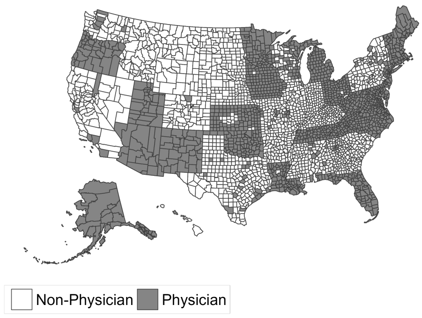


Figure A.6: Physician vs non-Physician

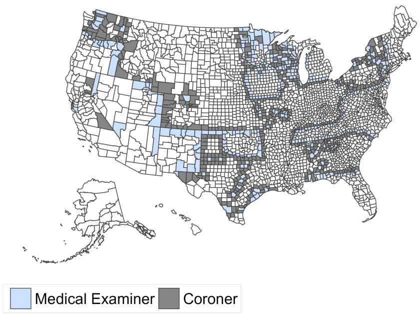


Figure A.7: Coroner vs Medical Examiner analysis sample

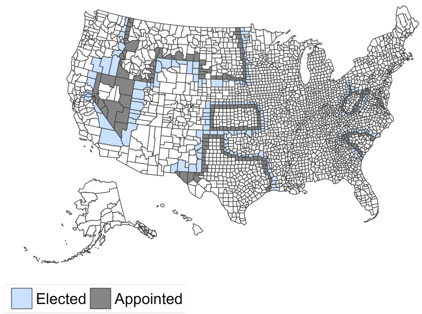


Figure A.8: Appointed vs Elected analysis sample

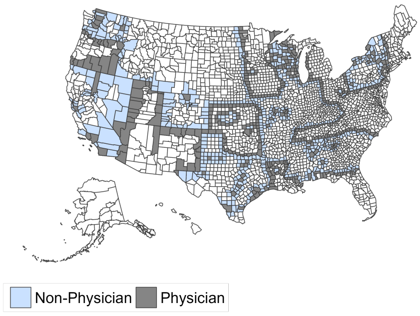


Figure A.9: Physician vs non-Physician analysis sample

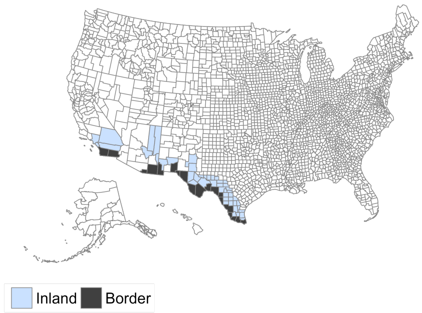


Figure A.10: US-Mexico border analysis sample

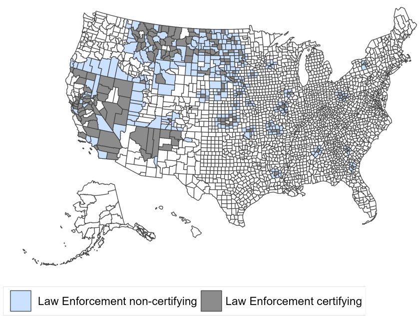


Figure A.11: Sh-Coroner or Native vs Other

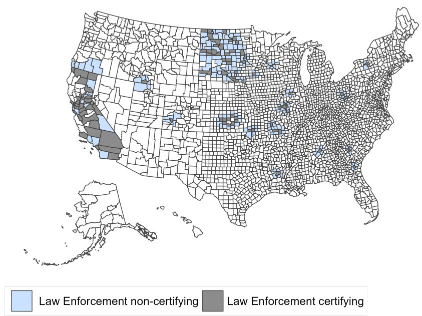


Figure A.12: Sh-Coroner subsample vs Other

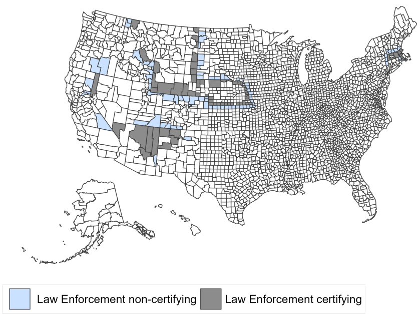


Figure A.13: non-Sh-Coroner subsample vs Other

A.4 Death Investigation System: History and Current Distribution

Year	State
Before 1954	Maryland, Virginia, Vermont, Rhode Island
1961	Tennessee*
1963	Oklahoma
1964	Oregon
1967	North Carolina, Utah
1968	Maine
1969	Connecticut
1970	Delaware, Iowa*
1976	West Virginia , New Mexico
1977	Kentucky*, Alabama*
1978	New Jersey*
1979	Arkansas*
1980	Montana*
1983	Massachusetts
1986	Mississippi*, New Hampshire
1990	Georgia*
1994	North Dakota*
1996	Alaska

List of states that have implemented a state medical examiner.
* States that have implemented the state medical examiner but chose to keep a coroner or chose a medical examiner system when the mixed system was an option.
Source: Hanzlick (2014)

State	OUR DATASET	CDC Systems	CDC State ME	Hanzlick '07*
AL	ME and Cor	ME and Cor	No	-
AK	ME	ME	Yes	State ME
AZ	ME	ME	No	ME
AR	Coroner	Coroner	Yes	Coroner
CA	ME and Cor	ME and Cor	No	ME and Cor
CO*	ME and Cor	Coroner	No	ME and Cor
CT	ME	ME	Yes	State ME
DE	ME	ME	Yes	State ME
FL	ME	ME	No	ME
GA	ME and Cor	ME and Cor	Yes	ME and Cor
HI	ME and Cor	ME and Cor	No	ME and Cor
ID	Coroner	Coroner	No	Coroner
IL	ME and Cor	ME and Cor	No	ME and Cor
IN	Coroner	Coroner	No	Coroner
IA	ME	ME	Yes	State ME
KS	Coroner	Coroner	No	Coroner
KY	Coroner	Coroner	Yes	Coroner
LA	Coroner	Coroner	No	Coroner
ME	ME	ME	Yes	State ME
MD	ME	ME	Yes	State ME
MA	ME	ME	Yes	State ME
MI	ME	ME	No	ME
MN	ME and Cor	ME and Cor	No	ME and Cor
MS*	Coroner	ME and Cor	Yes	Coroner
MO	ME and Cor	ME and Cor	No	ME and Cor
MT	Coroner	Coroner	Yes	Coroner
NE	Coroner	Coroner	No	Coroner
NV**	ME and Cor	Coroner	No	Coroner
NH	ME	ME	Yes	State ME
NJ	ME	ME	Yes	State ME
NM	ME	ME	Yes	State ME
NY	ME and Cor	ME and Cor	No	ME and Cor
NC	ME	ME	Yes	State ME
ND	Coroner	Coroner	Yes	Coroner
OH	ME and Cor	ME and Cor	No	ME and Cor
OK	ME	ME	Yes	State ME
OR	ME	ME	Yes	State ME
PA	ME and Cor	ME and Cor	No	ME and Cor
RI	ME	ME	Yes	State ME
SC	Coroner	Coroner	No	Coroner
SD	Coroner	Coroner	No	Coroner
TN	ME	ME	Yes	State ME
TX	ME and Cor	ME and Cor	No	ME and Cor
UT	ME	ME	Yes	State ME
VT	ME	ME	Yes	State ME
VA	ME	ME	Yes	State ME
WA	ME and Cor	ME and Cor	No	ME and Cor
WV	ME	ME	Yes	State ME
WI	ME and Cor	ME and Cor	No	ME and Cor
WY	Coroner	Coroner	No	Coroner
DC	ME	ME	No	State ME

Note: Following our inquiries, *CO, MS and NV of our dataset differ from CDC information, NV is also different from Hanzlick (2014)

Table A.1: Number of counties per Death Investigation System by State

State	Coroner	MedExam	Sheriff
AK	0	29	0
AL	57	9	1
AR	75	0	0
AZ	0	15	0
CA	7	5	46
CO	61	2	1
CT	0	8	0
DC	0	1	0
DE	0	3	0
FL	0	67	0
GA	152	5	2
HI	2	1	1
IA	0	99	0
ID	43	0	1
IL	96	1	5
IN	92	0	0
KS	92	0	13
KY	120	0	0
LA	63	0	1
MA	0	14	0
MD	0	24	0
ME	0	16	0
MI	0	83	0
MN	19	66	2
MO	96	15	4
MS	82	0	0

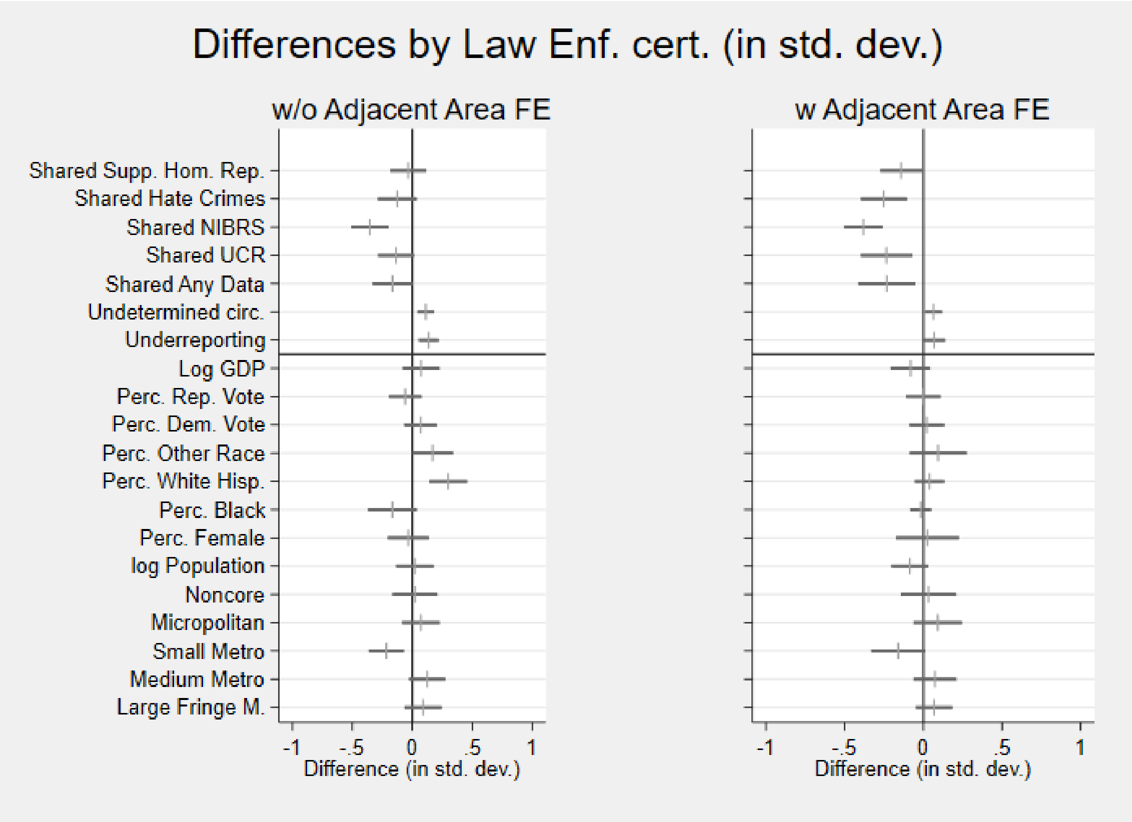
(a)

State	Coroner	MedExam	Sheriff
MT	19	0	37
NC	0	100	0
ND	35	0	18
NE	84	0	9
NH	0	10	0
NJ	0	21	0
NM	0	33	0
NV	0	2	15
NY	39	22	1
OH	84	2	2
OK	0	77	0
OR	0	36	0
PA	64	3	0
RI	0	5	0
SC	46	0	0
SD	55	0	11
TN	0	95	0
TX	228	21	0
UT	0	29	0
VA	0	133	0
VT	0	14	0
WA	49	6	0
WI	30	41	1
WV	0	55	0
WY	22	0	1
Total	1796	1168	172

(b)

A.5 Balancing Tables and Summary Statistics

Figure A.14: Balance Tables



Note: We regress each standardized variable on the law enforcement certifying dummy, with (right panel) and without (left panel) adjacency fixed effects.

Table A.2: Summary statistics: death investigation systems and outcomes

Dep. Var.:	Full population					Analysis sample	
	US	Coroner	Med.Exam.	Sh.-Coroner	LE certify	LE certify	Control
LE certify	0.1 (0.3)	0.07 (0.25)	0.02 (0.14)	1 (0)	1 (0)	1 (0)	0 (0)
Sh.-Coroner	0.05 (0.23)	0 (0)	0 (0)	1 (0)	0.53 (0.5)	0.63 (0.48)	0 (0)
Coroner	0.57 (0.49)	1 (0)	0 (0)	0 (0)	0.39 (0.49)	0.27 (0.44)	0.75 (0.43)
Med.-Exam.	0.37 (0.48)	0 (0)	1 (0)	0 (0)	0.08 (0.27)	0.1 (0.3)	0.25 (0.43)
police kill. MPV-SHR	0.21 (0.8)	0.13 (0.49)	0.32 (1.1)	0.3 (0.88)	0.24 (0.76)	0.33 (0.93)	0.15 (0.81)
police kill. SHR	0.14 (1.09)	0.04 (0.31)	0.26 (1.66)	0.3 (1.32)	0.2 (1.13)	0.34 (1.49)	0.23 (2.27)
police kill. MPV	0.35 (1.54)	0.17 (0.59)	0.58 (2.27)	0.6 (1.94)	0.44 (1.63)	0.67 (2.12)	0.38 (2.96)
SHR part.	0.87 (0.27)	0.84 (0.29)	0.91 (0.22)	0.89 (0.23)	0.84 (0.29)	0.86 (0.25)	0.89 (0.22)
NIBRS part.	0.48 (0.46)	0.42 (0.46)	0.57 (0.46)	0.45 (0.47)	0.45 (0.46)	0.42 (0.45)	0.59 (0.45)
Tot. Homicides SHR	4.8 (25.06)	2.32 (9.01)	8.43 (38.58)	6.26 (19.91)	4.79 (18.38)	7.43 (24.02)	4.25 (36.57)
Circ. undet. SHR	1.73 (10.4)	0.72 (4.78)	3.18 (15.5)	2.56 (9.27)	2.01 (9.28)	3.05 (11.83)	1.06 (8.61)
Appointed	0.22 (0.42)	0.29 (0.45)	0.07 (0.25)	0.56 (0.5)	0.36 (0.48)	0.36 (0.48)	0.44 (0.5)
Physician	0.45 (0.5)	0.14 (0.35)	1 (0)	0 (0)	0.08 (0.27)	0.1 (0.3)	0.42 (0.49)

Table A.3: Summary statistics: covariates

Dep. Var.:	Full population					Analysis sample	
	US	Coroner	Med.Exam.	Sh.-Coroner	LE certify	LE certify	Control
Large central Metro	0.02	0	0.05	0.02	0.02	0.02	0.03
	(0.15)	(0.06)	(0.22)	(0.13)	(0.12)	(0.15)	(0.16)
Large fringe Metro	0.12	0.1	0.15	0.06	0.05	0.08	0.06
	(0.32)	(0.3)	(0.36)	(0.23)	(0.21)	(0.27)	(0.23)
Medium Metro	0.12	0.1	0.15	0.09	0.09	0.11	0.08
	(0.32)	(0.31)	(0.35)	(0.29)	(0.28)	(0.32)	(0.27)
Micropolitan	0.2	0.22	0.18	0.17	0.19	0.15	0.23
	(0.4)	(0.42)	(0.38)	(0.38)	(0.39)	(0.36)	(0.42)
Non-core	0.42	0.46	0.35	0.55	0.57	0.54	0.52
	(0.49)	(0.5)	(0.48)	(0.5)	(0.5)	(0.5)	(0.5)
Small Metro	0.11	0.11	0.12	0.1	0.09	0.09	0.08
	(0.32)	(0.31)	(0.33)	(0.31)	(0.29)	(0.28)	(0.27)
Population (per 10,000)	10.27	5.48	17.37	12.66	10.75	16.17	10.83
	(32.9)	(10.8)	(49.32)	(38.79)	(33.48)	(43.72)	(63.73)
Fem. perc	0.5	0.5	0.5	0.49	0.49	0.5	0.5
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
White (non-Hisp.) perc.	0.77	0.77	0.76	0.76	0.8	0.77	0.83
	(0.2)	(0.2)	(0.2)	(0.22)	(0.2)	(0.22)	(0.17)
Black perc.	0.09	0.1	0.09	0.02	0.02	0.02	0.04
	(0.14)	(0.16)	(0.12)	(0.05)	(0.04)	(0.05)	(0.08)
Hisp. perc.	0.09	0.09	0.08	0.12	0.1	0.11	0.07
	(0.13)	(0.14)	(0.11)	(0.15)	(0.13)	(0.14)	(0.09)
Asian-PacIsl. perc.	0.02	0.01	0.02	0.03	0.02	0.02	0.02
	(0.03)	(0.02)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Native perc.	0.02	0.02	0.03	0.05	0.05	0.06	0.04
	(0.08)	(0.06)	(0.09)	(0.12)	(0.13)	(0.14)	(0.12)
REP. Perc. votes	0.62	0.65	0.57	0.62	0.64	0.62	0.63
	(0.15)	(0.14)	(0.15)	(0.17)	(0.17)	(0.18)	(0.15)
DEM. Perc. votes	0.35	0.32	0.39	0.33	0.31	0.33	0.32
	(0.15)	(0.15)	(0.15)	(0.16)	(0.17)	(0.17)	(0.15)
Other Perc. votes	0.04	0.04	0.04	0.05	0.05	0.05	0.05
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
GDP (per 100,000)	56.31	25.55	105.29	61.9	57.39	83.73	71.04
	(242.59)	(67.95)	(380.71)	(212.87)	(198.96)	(252.1)	(437.89)

A.6 From a discrete choice model to the Poisson specification

Suppose that police killings $k \in 0, 1$ in each county follow a Poisson process with a daily rate λ_i. Each year, the number of times the event happens $K_i = 365\lambda_i$ is Poisson-distributed. When a killing occurs, law enforcement must decide whether to report the killing or not, which will depend on a cost-benefit analysis of the two alternatives.

Assuming a random value function V, which is the sum of observable and (X) and unobservable factors (ϵ_r, ϵ_u), law enforcement will underreport if $V_r = X'\beta_r + \epsilon_r$ is greater than $V_u = X'\beta_u + \epsilon_u$, or if $X'(\beta_r - \beta_u) > -(\epsilon_r - \epsilon_u)$.

Assuming that the unobservable factors that enter the analysis follow a type-I generalized extreme value distribution, as is common in these types of models, the probability to underreport is:

$$p = \frac{1}{1 + \exp(-X'(\beta_r - \beta_u))} = \frac{1}{1 + \exp(\Delta)}.$$

Assuming that conditional on potential confounders unobservables are independent across time, the number of underreported killings is going to be distributed Binomial(K,p). It is known that when K is large and p is small, this binomial distribution can be approximated by a Poisson distribution with mean Kp. Moreover, for small differences Δ, it can be shown that $-\log(1 + \exp(\Delta)) \approx -\log(2) - \Delta/2$.[1] Thus, controlling for the exposure K_i, the Poisson model is linear in $X'(\beta_r - \beta_u)$, and our treatment effects $\beta_u - \beta_r$ are proportional to the differences in the values of under-reporting.

[1] In particular, for small Δ, we have that $\log(1 + \exp(\Delta)) \approx \log(2 + \Delta)$, and with a first order Taylor approximation $\log(2 + \Delta) \approx \log(2) + \Delta/2$.

A.7 Robustness Results, Placebo Evaluations and Leverage of Each State

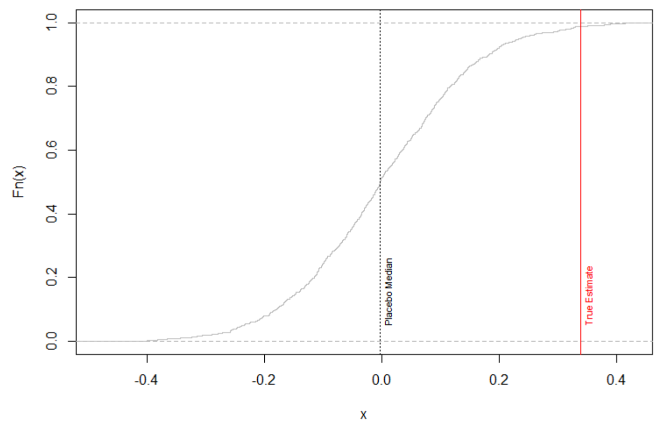


Figure A.15: Main Effect Placebo Point Estimate CDF

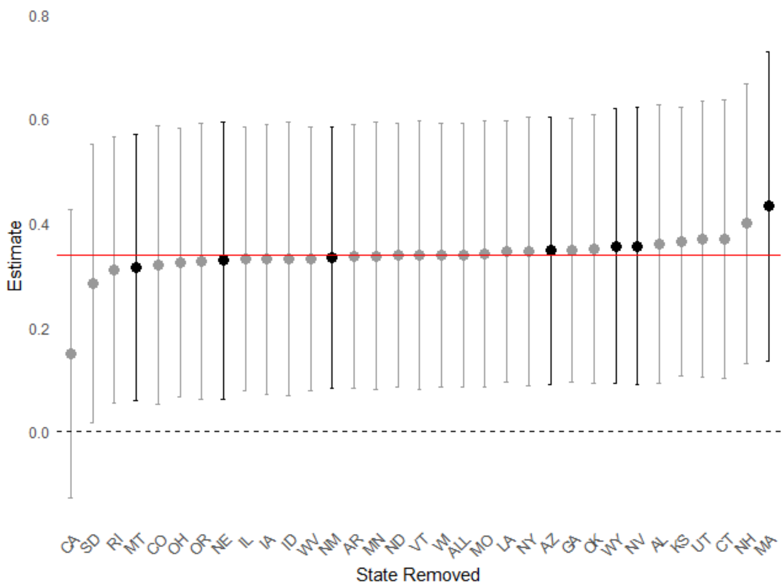


Figure A.16: Leverage of Each State on Main Effect
This Figure presents the non-transformed point estimates and 95% confidence intervals from our main specification when removing each state listed on the x-axis. In black are highlighted the removed states which are listed in Ruiz et al. (2018) as allowing law enforcement to certify the cause of death. For comparison, we include our main effect highlighted in red.

Table A.4: Robustness checks of authorizing law enforcement to certify cause of death

Dep. Var.:	Diff. law enforcement homicides														
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
LE certify	0.605**	0.639***	0.948***	0.392**	0.416**	0.309**	0.397***	0.455***	0.383**	0.480***	0.616***	0.134***	0.357**	0.090**	0.606**
	0.473	0.494	0.667	0.331	0.348	0.269	0.334	0.375	0.324	0.392	0.480		0.305		0.474
	(0.214)	(0.170)	(0.186)	(0.140)	(0.168)	(0.124)	(0.129)	(0.132)	(0.132)	(0.127)	(0.164)	(0.022)	(0.103)	(0.038)	(0.191)
	[0.027]	[0.004]	[0.000]	[0.018]	[0.038]	[0.029]	[0.010]	[0.005]	[0.014]	[0.002]	[0.003]	[0.000]	[0.003]	[0.019]	[0.013]
All cov.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Spec.	Pois.	Pois.	Pois.	Pois.	Pois.	Pois.	Pois.	Pois.	Pois.	Pois.	Pois.	Lin.	Pois.	Lin.	Pois.
N_{tot}	3289	1889	2156	3247	929	3289	3289	3289	3289	3289	3289	21993	21993	21993	21993
N_{eff}	1904	1370	1197	2589	821	2561	2561	2561	2561	2561	2561	21993	21993	4506	4506
N_{treat}	553	440	462	833	436	874	874	874	874	874	874	2253	2253	2253	2253
$N_{counties}$	272	196	171	370	182	366	366	366	366	366	366	3142	3142	1165	1165
$\mu_{outc.}$	0.296	0.181	0.311	0.258	0.777	0.286	0.286	0.286	0.277	0.259	0.286	0.217	0.22	0.149	0.194

Note: Table displays transformed coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by estimated coefficient, its standard error, and its p-value. (1) includes adjacent county cluster fixed effects and state fixed effects, (2) includes adjacent county cluster fixed effects and only includes bordering state counties, (3) includes adjacent county cluster fixed effects and matches treated and control on same urbanization level, (4) consider as 'treated' sheriff-coroner counties and counties with large tribal territories in Arizona and New Mexico allowing 'controls' to include counties in states which allow law enforcement to certify the cause of death, (5) conditions on counties that have positive reported SHR homicides, (6) includes a covariate indicating whether death investigator was a Physician (which is a potential bad control) (7) drops the potential bad control total homicide variable, (8) adds a covariate capturing the stringency of gun and ammunition laws by state, (9) Outcome is MPV police killings - MPV off-duty police killings - SHR police killings, (10) Outcome is MPV police killings by firearm (+ other) - SHR police killings, (11) extrapolated effects weighting by propensity to be in sample. Poisson estimation with year fixed effects and standard errors clustered at the county level. Set of controls described in Appendix, (12-13) show pooled OLS and Poisson results on all counties, (14-15) show propensity score nearest-neighbor matching OLS and Poisson results on all counties. For columns 12 and 14 $\mu_{outc.}$ represents the outcome mean for the control group A.5.

A.8 Subcircumstances effects

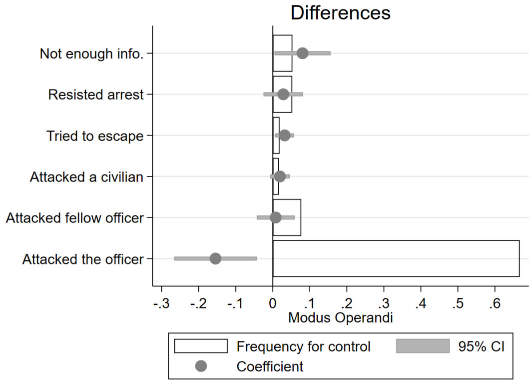


Figure A.17: Effects on SHR sub-circumstances of death

Sub-circumstances are listed for 89% of SHR police killings. Estimating linear probability models, the results indicate that police killings in law enforcement certifying counties occur less often in response to attacks by victims and more often when a victim was fleeing or under undisclosed reasons. If taken at face value, these results are insightful. Since the Supreme Court decision in Tennessee v. Garner (1985), deadly force should only be used on a fleeing suspect if that suspect committed egregious crimes or poses an immediate threat to others.[2] Our results indicate that officers in law enforcement certifying counties apply a less strict threshold when choosing to use lethal violence against fleeing offenders. The result showing fewer justified reasons for lethal violence ('Not enough info') may also suggest that officers assume they are likely to be questioned when filing police homicide reports.

[2]There is some correlational evidence that this led to a slight reduction in reported police killings (Tennenbaum, 1994), though in light of our results, without crowed-sourced 'true' numbers of police killings that go so far back in time, it is unclear whether those changes were due to increased underreporting.

A.9 Placebo checks on other SHR categories

To verify the claim of strategic re-classification in Table 1.2, and check on the robustness of our main death certification law effects, we also carry out placebo checks on all other 29 FBI death categories aside from 'circumstances undetermined'. These results, not presented in tables, find that only three categories, 'circumstances undetermined', 'Other negligent handling of gun', and 'child killed by babysitter' appear significant in our main specification and in that same specification when including state fixed-effects. These effects are all positive. The effect on 'Other negligent handling of gun' may be linked to re-classification of police killings but occurs relatively little in our analysis sample. We only observe, on average, 24 instances of 'Other negligent handling of gun' a year, compared to 1545 instances of 'circumstances undetermined' homicides and 263 instances of SHR police killings. We consider the effect on 'child killed by babysitter' as a statistical outlier, as it only includes on average 10 deaths a year in our adjacency sample. These placebo tests offer further support to the robustness of our main underreporting effects and to the claim that the effect on 'circumstances undetermined' is strategically relevant.

A.10 Principle Stratification Effect Decomposition

A.10.1 Identification of Effect Decomposition

Table A.5: Extended County Types

Type	W^1	W^0	Y^1	Y^0
$\mathbf{a^W a^Y}$	**1**	**1**	**1**	**1**
$\mathbf{a^W n^Y}$	**1**	**1**	**0**	**0**
$\mathbf{a^W c^Y}$	**1**	**1**	**1**	**0**
$a^W d^Y$	1	1	0	1
$\mathbf{n^W a^Y}$	**0**	**0**	**1**	**1**
$\mathbf{n^W n^Y}$	**0**	**0**	**0**	**0**
$\mathbf{n^W c^Y}$	**0**	**0**	**1**	**0**
$n^W d^Y$	0	0	0	1
$\mathbf{c^W a^Y}$	**1**	**0**	**1**	**1**
$\mathbf{c^W n^Y}$	**1**	**0**	**0**	**0**
$\mathbf{c^W c^Y}$	**1**	**0**	**1**	**0**
$c^W d^Y$	1	0	0	1
$d^W a^Y$	0	1	1	1
$d^W n^Y$	0	1	0	0
$d^W c^Y$	0	1	1	0
$d^W d^Y$	0	1	0	1

To reduce notational burden, we exclude in the proof the subscript i. Assumption

A.II fixes all probabilities of non-bold types in Table A.5 to 0. As a result, and using assumption A.I, we can derive the following shares,

$$
\begin{aligned}
&\Pr(\mathrm{W}^1 = 1, \mathrm{W}^0 = 1) \equiv \Pr(n^Y) = \Pr(\mathrm{W} = 1 |\, \mathrm{D} = 0) \\
&\Pr(\mathrm{W}^1 = 1, \mathrm{W}^0 = 0) \equiv \Pr(c^W) = \Pr(\mathrm{W} = 1 |\, \mathrm{D} = 1) - \Pr(\mathrm{W} = 1 |\, \mathrm{D} = 0) \\
&\Pr(\mathrm{W}^1 = 0, \mathrm{W}^0 = 0) \equiv \Pr(n^W) = 1 - \Pr(\mathrm{W} = 1 |\, \mathrm{D} = 1) = \Pr(\mathrm{W} = 0 |\, \mathrm{D} = 1) \\
&\Pr(\mathrm{Y}^1 = 1, \mathrm{Y}^0 = 1) \equiv \Pr(a^Y) = \Pr(\mathrm{Y} = 1 |\, \mathrm{D} = 0) \\
&\Pr(\mathrm{Y}^1 = 1, \mathrm{Y}^0 = 0) \equiv \Pr(c^Y) = \Pr(\mathrm{Y} = 1 |\, \mathrm{D} = 1) - \Pr(\mathrm{Y} = 1 |\, \mathrm{D} = 0) \\
&\Pr(\mathrm{Y}^1 = 0, \mathrm{Y}^0 = 0) \equiv \Pr(n^Y) = 1 - \Pr(\mathrm{Y} = 1 |\, \mathrm{D} = 1) = \Pr(\mathrm{Y} = 0 |\, \mathrm{D} = 1)
\end{aligned}
\tag{A.1}
$$

In addition, we can relate the nine remaining county type probabilities to observed outcomes as follows:

$$
\begin{aligned}
&\Pr(\mathrm{W} = 1, \mathrm{Y} = 1 |\, \mathrm{D} = 0) = \Pr(a^W a^Y) \\
&\Pr(\mathrm{W} = 1, \mathrm{Y} = 1 |\, \mathrm{D} = 1) = \Pr(a^W a^Y) + \Pr(a^W c^Y) + \Pr(c^W a^Y) + \Pr(c^W c^Y) \\
&\Pr(\mathrm{W} = 1, \mathrm{Y} = 0 |\, \mathrm{D} = 0) = \Pr(a^W c^Y) + \Pr(a^W n^Y) \\
&\Pr(\mathrm{W} = 1, \mathrm{Y} = 0 |\, \mathrm{D} = 1) = \Pr(a^W n^Y) + \Pr(c^W n^Y) \\
&\Pr(\mathrm{W} = 0, \mathrm{Y} = 1 |\, \mathrm{D} = 0) = \Pr(n^W a^Y) + \Pr(c^W a^Y) \\
&\Pr(\mathrm{W} = 0, \mathrm{Y} = 1 |\, \mathrm{D} = 1) = \Pr(n^W a^Y) + \Pr(n^W c^Y) \\
&\Pr(\mathrm{W} = 0, \mathrm{Y} = 0 |\, \mathrm{D} = 1) = \Pr(n^W n^Y)
\end{aligned}
$$

The above effectively presents us with a system of five equations and seven unknowns.

Simple algebra shows that under the additional Assumption A.III, with $\Pr(c^W n^Y) = 0$ and $\Pr(n^W c^Y) = 0$ all remaining joint probabilities in equation A.2 are identified,

$$
\begin{aligned}
\Pr(c^W c^Y) &= \Pr(\mathrm{W} = 1, \mathrm{Y} = 1 |\, \mathrm{D} = 1) - \Pr(\mathrm{W} = 1, \mathrm{Y} = 1 |\, \mathrm{D} = 0) \\
&\quad + \Pr(\mathrm{W} = 1, \mathrm{Y} = 0 |\, \mathrm{D} = 1) - \Pr(\mathrm{W} = 1, \mathrm{Y} = 0 |\, \mathrm{D} = 0) \\
&\quad + \Pr(\mathrm{W} = 0, \mathrm{Y} = 1 |\, \mathrm{D} = 1) - \Pr(\mathrm{W} = 0, \mathrm{Y} = 1 |\, \mathrm{D} = 0) \\
&= \Pr(\mathrm{W} = 0, \mathrm{Y} = 0 |\, \mathrm{D} = 0) - \Pr(\mathrm{W} = 0, \mathrm{Y} = 0 |\, \mathrm{D} = 1) \qquad \textit{upper} \\
\Pr(a^W c^Y) &= \Pr(\mathrm{W} = 1, \mathrm{Y} = 0 |\, \mathrm{D} = 0) - \Pr(\mathrm{W} = 1, \mathrm{Y} = 0 |\, \mathrm{D} = 1) \qquad \textit{lower} \\
\Pr(a^W n^Y) &= \Pr(\mathrm{W} = 1, \mathrm{Y} = 0 |\, \mathrm{D} = 1) \qquad \textit{upper} \\
\Pr(c^W a^Y) &= \Pr(\mathrm{W} = 0, \mathrm{Y} = 1 |\, \mathrm{D} = 0) - \Pr(\mathrm{W} = 0, \mathrm{Y} = 1 |\, \mathrm{D} = 1) \qquad \textit{lower} \\
\Pr(n^W a^Y) &= 1 - \Pr(\mathrm{W} = 1, \mathrm{Y} = 1 |\, \mathrm{D} = 1) \qquad \textit{upper}
\end{aligned}
\tag{A.2}
$$

In addition, if either of Assumptions A.IIIi and A.IIIii does not hold, we can still infer potentially informative upper and lower bounds on the above type probabilities. From the joint probabilities in equations A.2 and the share probabilities in equations A.1, we can also apply Bayes' rule to obtain all conditional probabilities of interest such as $\Pr(\mathrm{W}^1 = 1, \mathrm{W}^0 = 0 |\, \mathrm{Y}^1 = 1\, \mathrm{Y}^0 = 0) = \Pr(c^W c^Y)/\Pr(c^Y)$.

A.10.2 Principle Stratification placebo evaluations

We can further inspect whether the joint effect observed in columns 1-3 is special relative to the other 29 SHR crime circumstances categories. Notice that the joint probability of being a complier of any type in our setting is $\Pr(\mathrm{W}=1,\mathrm{Y}=1|\,\mathrm{D}=1)-\Pr(\mathrm{W}=1,\mathrm{Y}=1|\,\mathrm{D}=0)=\Pr(a^W c^Y)+\Pr(c^W a^Y)+\Pr(c^W c^Y)$, which for 'circumstances undetermined' equals 0.044. Figure A.18 compares this overall complier effect for all SHR categories. We find that only two other SHR crime categories produce joint significant effects. These are the categories 'other manslaughter' and 'other non-specified' with joint effects half as large as the ones on 'circumstances undetermined'.[3]

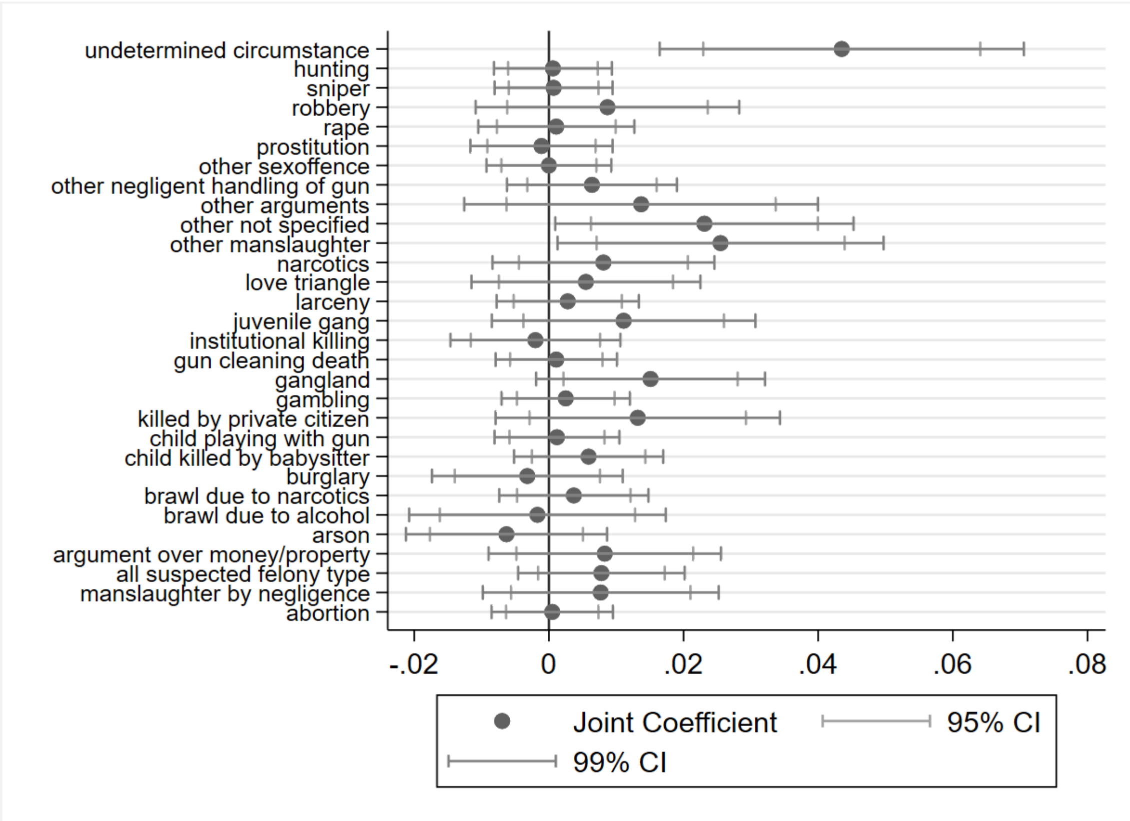


Figure A.18: Joint Estimate of complier effects on other SHR crime categories

[3]As indicated in the main text, these results may indicate that underreported police killings are also reclassified into these 'other etc.' categories. Although we cannot exclude this possibility, some of the 'other etc.' categories are associated to a set of conditions (e.g. felony vs non-felony), which is not the case for 'circumstances undetermined'. To be sure, we ran our joint estimate combining the 'other manslaughter' and 'other non-specified' categories with circumstances undetermined and obtained almost the same result (0.042 (0.010)) as that presented in column 2 of Table 1.3. We also estimated our results in column 3 of Table 1.2 with the combined measure and still find significant positive effects.

A.10.3 Testing Assumption A.III

The results also partially test Assumption III. Recall that if the treatment effect from column 3, $\Pr(\text{W} = 0, \text{Y} = 1 | \text{D} = 1) - \Pr(\text{W} = 0, \text{Y} = 1 | \text{D} = 0) = \Pr(n^W c^Y) - \Pr(c^W a^Y)$, and of column 4, $\Pr(\text{W} = 1, \text{Y} = 0 | \text{D} = 1) - \Pr(\text{W} = 1, \text{Y} = 0 | \text{D} = 0) = \Pr(c^W n^Y) - \Pr(a^W c^Y)$, were positive (and significant), this would imply that $Pr(n^W c^Y) > 0$ and $Pr(c^W n^Y) > 0$. Given that we find a zero treatment effect, it is unlikely for counties to be simultaneously compliers and never-takers, or compliers and always-takers, unless it so happens that $Pr(c^W n^Y) = Pr(a^W c^Y) \neq 0$ and $Pr(n^W c^Y) = Pr(c^W a^Y) \neq 0$.

A.11 Estimation and Additional Results of Effect Decomposition

Table 1.3 presents robust estimates of complier effects accounting for adjacency cluster fixed effects. These linear specifications are, however, not best suited to estimate type probabilities for probability types with no compliers. While the focus of this paper is not to develop the econometric theory of estimators, we offer here a view on some tradeoffs in estimation looking at logit estimators.

Specific to our setting, we must also address a relevant choice in the estimation. When calculating these effects, some ratios $\underline{\text{W}}_i$ are undefined if no UCR total homicides are observed in a given county-year. Even if we relax the definition of $\underline{\text{W}}_i$ to represent non-zero occurrences of 'circumstances undetermined' homicides in a given county, we must address the question of whether it makes sense to consider re-classification of police killings if no homicides are observed. Such instances are non-negligible in the data. They appear in 61% of county-year observations. A priori, it is unclear how to redefine W_i in these cases. In our main results, we address this question by presenting results in Table A.4 conditioning on positive homicides in a given year-county. These results are very close to the ones from our main specification in Table 1.1. The same cannot be said for our estimation of joint effects. As a result, in our main results, we opt to randomly impute a value of 0 or 1 to W_i generated from a binomial distribution when no homicides are observed. In effect, this implies that these undefined instances do not contribute to any effects. We find this imputation to work better than imputing values of 0 to all undefined instances of W_i or throwing them out of the sample, both of which result in any probabilities which lie below 0 or above 1.

Table A.6 presents a full set of results with this random imputation calculating each probability type from its bootstrap mean. More precisely, we first estimate three models: a logit estimation of Y on D, a logit estimation of W on D conditioning on the sample with $Y = 0$, and a logit estimation of W on D conditioning on the sample with $Y = 1$. We also present bootstrap 95% confidence intervals in brackets. Based on our potential outcome model, we find that $\Pr(a^Y) = 8.5\%$ of counties show underreporting of police killings regardless of the leniency of certification laws. The majority of counties, $\Pr(n^Y) = 82.9\%$, never underreport police killings. We also find that $\Pr(c^W) = 7.8\%$ of counties report excess undetermined homicides only when facing permissive certification laws.

We further find that $\Pr(c^Y) = 8.6\%$ and $\Pr(c^W c^Y) = 7.0\%$ are higher than those estimated in our more robust main specification. They still, however, convey the same conclusion, namely that the difference in underreporting between law enforcement certifying and non-certifying counties is largely driven by counties in which law enforcement underreport police killings and simultaneously over-report excess undetermined homicides. From Table A.6, we can also see that, conditional on underreporting police killings, a county is very likely to display excess undetermined homicides, $\Pr(c^W|c^Y) = 81.9\%$. Similarly, conditional on reporting excess undetermined homicides, a county is likely to underreport police killings, $\Pr(c^Y|c^W) = 88.1\%$.

Finally, we also notice that a small share of counties, $\Pr(a^W a^Y) = 3.4\%$, may always be reclassifying underreported police killings in the circumstances undetermined category regardless of death certification laws. This last effect does not depend on assumption A.III for identification. In contrast, if assumption A.III were incorrect, allowing $\Pr(c^W n^Y) > 0$ and $\Pr(n^W c^Y) > 0$, then $\Pr(c^W|c^Y)$ and $\Pr(c^Y|c^W)$ would be upper bounds and $\Pr(c^W a^Y)$ would be a lower bound.

Table A.7 includes covariates in the logit estimation. We find this specification to be less robust producing joint and conditional probability estimates below 0 and above 1, although the 95% confidence intervals cover the 0 to 1 bound.

Table A.6: Re-classification Type Probabilities: Logit no-covariates

$a^W a^Y$	0.046 [0.038;0.055]	$c^W a^Y$	0.009 [-0.003;0.022]	$n^W a^Y$	0.030 [0.020;0.040]	a^Y	0.085 [0.074;0.097]	$c^W\|a^Y$	0.105 [-0.038;0.244]
$a^W c^Y$	0.016 [-0.020;0.047]	$c^W c^Y$	0.070 [0.033;0.103]	$n^W c^Y$	0.000 [-]	c^Y	0.086 [0.064;0.109]	$c^W\|c^Y$	0.819 [0.421;1.272]
$a^W n^Y$	0.403 [0.377;0.432]	$c^W n^Y$	0.000 [-]	$n^W n^Y$	0.426 [0.398;0.457]	n^Y	0.829 [0.808;0.849]	$c^W\|n^Y$	0.000 [-]
a^W	0.465 [0.445;0.486]	c^W	0.079 [0.043;0.111]	n^W	0.456 [0.428;0.485]				
$c^Y\|a^W$	0.034 [-0.044;0.097]	$c^Y\|c^W$	0.881 [0.683;1.045]	$c^Y\|n^W$	0.000 [-]				

Note: Table displays logit estimates with bootstrap 95% confidence intervals in brackets, $N_{tot} = 3,464$, $N_{counties} = 495$.

Table A.7: Re-classification Type Probabilities: Logit with covariates

$a^W a^Y$	0.046 [0.036;0.058]	$c^W a^Y$	0.012 [-0.006;0.032]	$n^W a^Y$	0.012 [0.090;0.117]	a^Y	0.103 [-0.060;0.288]	$c^W\|a^Y$	0.119 [-0.030;0.253]
$a^W c^Y$	-0.011 [-0.047;0.021]	$c^W c^Y$	0.045 [0.009;0.082]	$n^W c^Y$	0.000 [-]	c^Y	0.033 [0.012;0.055]	$c^W\|c^Y$	1.574 [0.396;4.125]
$a^W n^Y$	0.430 [0.404;0.458]	$c^W n^Y$	0.000 [-]	$n^W n^Y$	0.433 [0.407;0.462]	n^Y	0.863 [0.845;0.882]	$c^W\|n^Y$	0.000 [-]
a^W	0.465 [0.444;0.487]	c^W	0.057 [0.022;0.094]	n^W	0.478 [0.447;0.507]				
$c^Y\|a^W$	-0.025 [-0.103;0.045]	$c^Y\|c^W$	0.774 [0.361;1.133]	$c^Y\|n^W$	0.000 [-]				

Note: Table displays probabilities generated from fitted values of logit estimations with bootstrap p-values in brackets. $N_{tot} = 3,464$, $N_{counties} = 495$. Bootstrap p-values estimated as share of 399 bootstrap estimates smaller or equal to 0.

A.12 Histograms and Yearly Mean Maps: Threats to Police

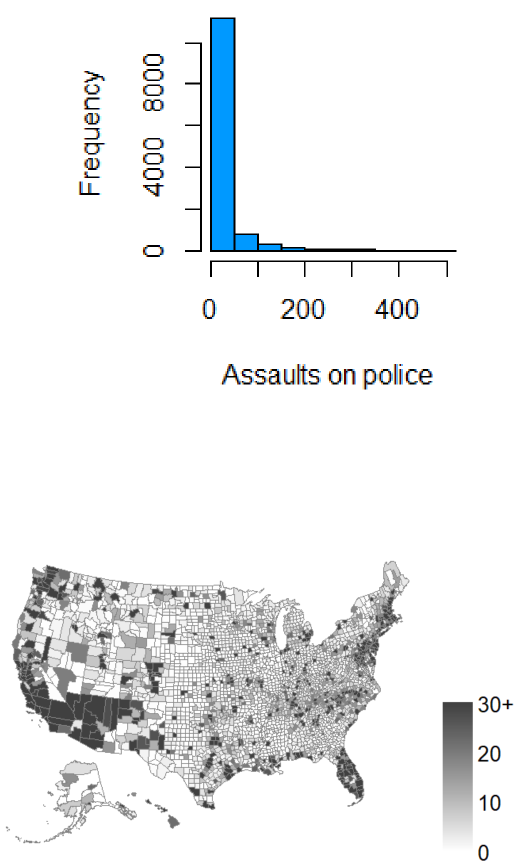


Figure A.19: Assaults on police

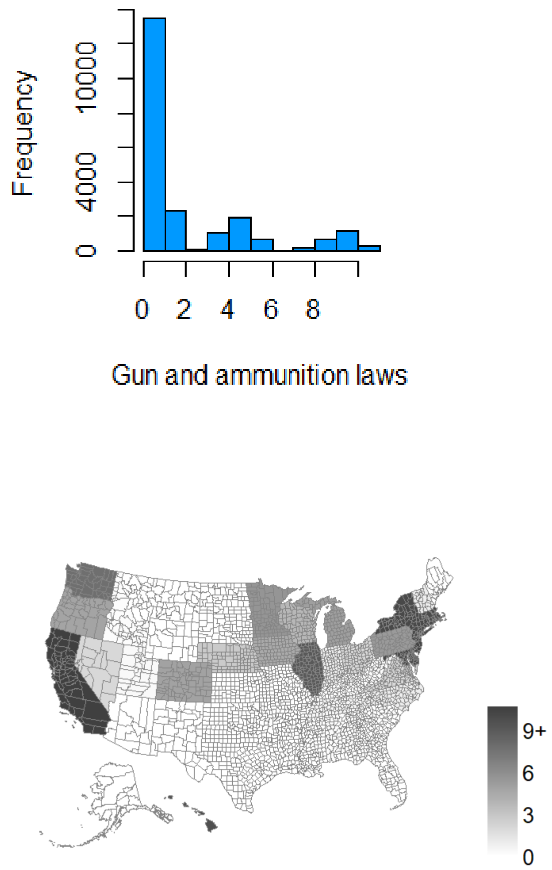
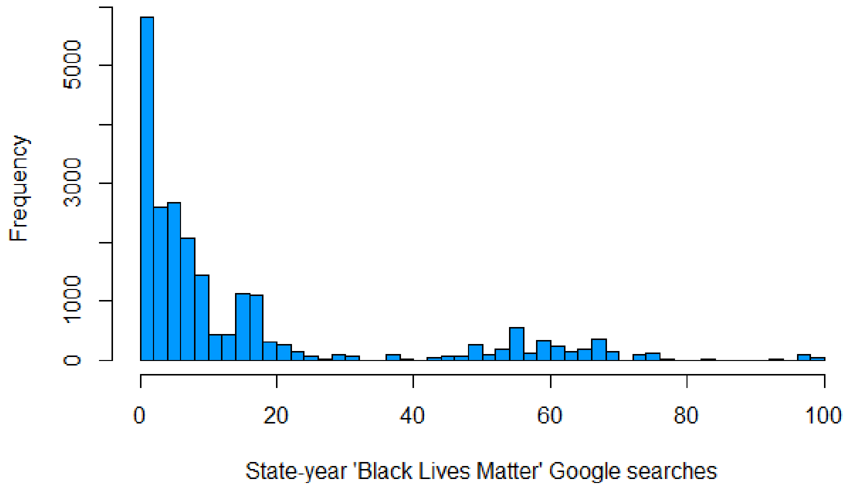


Figure A.20: Gun regulations (11=A, 0=F)

A.13 Black Lives Matter Google Trends



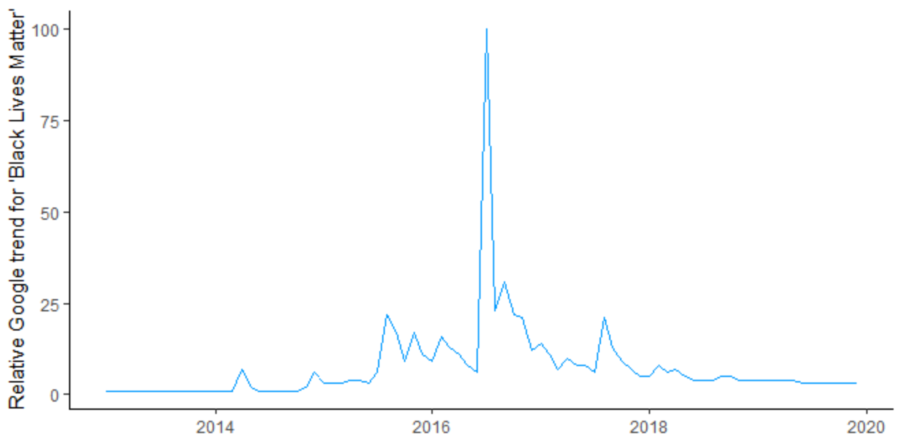


Table A.8: Description of Variables

	Description
Outcome Variables	
map_kill_pol	The killing of a felon by a law enforcement officer in the line of duty. This is the Official number of police killings reported by the FBI .
mpv_kill_pol	Total incidents where a law enforcement officer (off- duty or on-duty) applies, on a civilian, lethal force resulting in the civilian being killed whether it is considered "justified" or "unjustified" by the US Criminal Legal System. We consider MPVthis to be our True value of police killings.
diff_kill_pol	Difference between mpv_kill_pol and map_kill_pol. When the difference is negative, namely when the official killings are more than the true police killings, the variable is set to zero
diff_kill_black	Difference between mpv_kill_black and map_kill_black. When the difference is negative, namely when the official killings are more than the true police killings, the variable is set to zero
diff_kill_hisp	Difference between mpv_kill_hisp and map_kill_hisp. When the difference is negative, namely when the official killings are more than the true police killings, the variable is set to zero. The variable includes Black people and White people with Hispanic origins. Black's Hispanic origin has been retrieved with Forebears analyzing their names and accounting for a percentage of Hisponosphere above 50%. While the information for White Hispanic people is already embedded in the datasets.
diff_kill_white	Difference between mpv_kill_white and map_kill_white. When the difference is negative, namely when the official killings are more than the true police killings, the variable was set to zero
diff_kill_nwhite	Difference between mpv_kill_nwhite and map_kill_nwhite, namely all the Non-White categories of race and ethnicity

Table A.8 continued from previous page

	Description
diff_kill_nblack	Difference between mpv_kill_nblack and map_kill_nblack, namely all the Non-Black categories of race and ethnicity
map_hom_tot	Total Homicides on the murder accountability project dataset which is based on SHR
mental_illness	A Person who died after an interaction with the police was allegedly showing mental health issues, as registered in the Mapping Police Violence database.
circumstance_undetermined	Person died after an interaction with the police under circumstances undetermined as reported in the Supplementary Homicide Report (SHR).
lemas_body_cam	law enforcement agencies in a county claiming to operate with body-worn cameras, Body-Worn Camera Supplement (LEMAS-BWCS) census conducted in 2016
leoka_homicides	Total homicides of law enforcement officers from yearly numbers of law enforcement officers who have been victims of felony homicides while on duty in each county are gathered from the Law Enforcement Officers Killed and Assaulted, LEOKA
leoka_assault	Total homicides of law enforcement officers from yearly numbers of law enforcement officers who have been victims of felony assaults while on duty in each county are gathered from the Law Enforcement Officers Killed and Assaulted, LEOKA
UCR	Agencies participate voluntarily and submit their crime data either through a state UCR program or directly to the FBI's UCR Program. cde.ucr.cjis.gov/LATEST
NIBRS	National Incident-Based Reporting System (NIBRS) Participation Status Indicates if the agency has submitted NIBRS data. Uniform Crime Reporting (UCR) Program - NIBRS Participation by State cde.ucr.cjis.gov/LATEST
Treatments	

Table A.8 continued from previous page

	Description
LE_cert	Law Enforcement is allowed to certify the cause of death in the counties where it takes value one, zero otherwise.
SherCoroner*	Takes value one when this is the Death Investigation System implemented in the county of Interest. Namely, the Sheriff-Coroner is the person entitled to determine the cause of death of a deceased person
Coroner*	Takes value one when this is the Death Investigation System implemented in the county of Interest. Namely the Coroner is the person entitled to determine the cause of death of a deceased person
MedExaminer*	Takes value one when this is the Death Investigation System implemented in the county of Interest. Namely the Medical Examiner is the person entitled to determine the cause of death of a deceased person
Physician	Counties where the death investigator is required to be a physician.
Appointed	Counties where the death investigator is appointed and not elected.
Covariates	
Large Central Metro	Large central metro counties are counties in metropolitan statistical areas (MSA) of 1 million or more population. Urbanization size of the county according to the 2013 urbanization. www.cdc.gov
Large Fringe Metro	Large fringe metro counties are counties in MSAs of 1 million or more population that do not qualify as large central. Urbanization size of the county according to the 2013 urbanization. www.cdc.gov
Medium Metro	Medium metro counties are counties in MSAs of 250,000 to 999,999 population. Urbanization size of the county according to the 2013 urbanization. www.cdc.gov
Small Metro	Small metro counties are counties in MSAs of less than 250,000 population. Urbanization size of the county according to the 2013 urbanization.www.cdc.gov
Microp	Micropolitan counties are counties in micropolitan statistical areas. Urbanization size of the county according to the 2013 urbanization. www.cdc.gov
Noncore	Non-core counties are nonmetropolitan counties that are not in a micropolitan statistical area.

Table A.8 continued from previous page

	Description
	Urbanization size of the county according to the 2013 urbanization. www.cdc.gov
Population	Population of the selected county in a determined year, data from census www.census.gov
fem_perc	Percentage of the female population according the census data www.census.gov
Black_perc	Percentage of the Black population according the census data www.census.gov
WHisp_perc	Percentage of the White Hispanic population according the census data www.census.gov
WNonHisp_perc	Percentage of the White population according the census data www.census.gov
Other_perc	Percentage of the Asian-Pacific Islanders and Native American population according to the census data www.census.gov
perc_votes.REP	Voting percentages during the Mayor elections in counties' major cities - Republican party, Presidential election data
perc_votes.DEM	Voting percentages during the Mayor elections in counties' major cities - Democratic party, Presidential election data
perc_votes.OTHER	Voting percentages during the Mayor elections in counties' major cities - Other parties, Presidential election data
GDP	Gross Domestic Product levels per county during the seven-year period. Data from Bureau of Economic Analysis Bureau of Economic Analysis [CAGDP1:GDP Summary by County and MSA]

*Note:**The variables have been created following the Census of Medical Examiners' and Coroners' Offices (MECO) Series published in December 2021. The missing values have been added one by one by searching through the CDC website (www.cdc.gov) and for each state indicating a mixed system we have gone through all the counties' websites looking whether they had a Coroner, Sheriff-Coroner or a Medical Examiner.

A.14 Strategic Withholding of Data

A.14.1 Law Enforcement Reporting of UCR and NIBRS Data

In addition to measures of homicide, we also have yearly information on whether each law enforcement agency shared crime reports to the UCR and NIBRS databases of the FBI.[4] In general, the distribution of law enforcement agencies per county selected in our data is right skewed and varies considerably, with an average of 5.6 law enforcement agencies per county across the US and a maximum of 131 for Cook county, Illinois.[5] The NIBRS, which has been proposed to replace the UCR program, is designed to include a wider range and more detailed accounts of crimes.[6] Participation rates in the UCR database are high in the years covered by our data, at 86.6%. This UCR information can be seen as a close proxy to the SHR data since only 3.3% of UCR reports do not have associated SHR homicide circumstances. Data sharing to the NIBRS is noticeably lower, at 47.8% and spread unevenly across state lines as a result of some states making the NIBRS mandatory to obtain subsidies in our observation period. For both measures of agency data sharing, we still notice considerable variation within states. Figures A.21-A.22 show the share reporting to the UCR and NIBRS. In general, medical examiner counties have a higher percentage of agencies sharing to both databases.

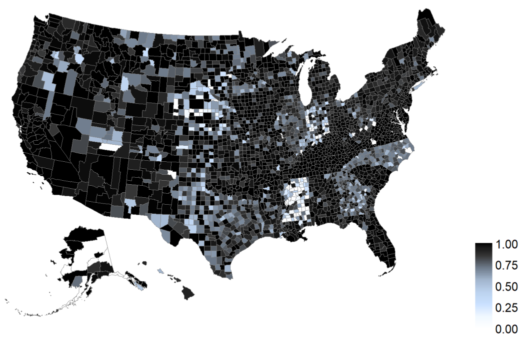


Figure A.21: Share reporting UCR data

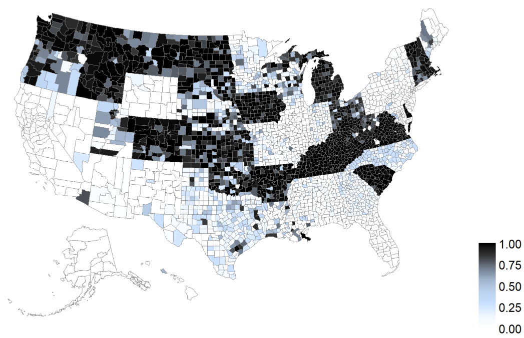


Figure A.22: Share reporting NIBRS data

[4] At the time of writing, these data could be downloaded at the following link on the FBI crime explorer website.

[5] To obtain county shares we exclude any agency - such as state troopers, park rangers, university security - which are not defined at the county level. 92% of remaining law enforcement agencies operate only in one county. Of those operating in more than one county, 86% operate in only two counties. For any law enforcement agency spanning multiple counties we divide their participation rate in proportion to the population weight of each county covered.

[6] Despite being proposed to replace the UCR by Jan. 2021, and some federal programs and grants requiring NIBRS reporting as a condition for funding, the uptake of the NIBRS has been slow. In 2021 only 65% of police departments submitted homicide details in crime reports to the FBI (The Marshall Project, 2022).

A.14.2 Certification Law Effects on Data Reporting

Besides reclassifying police killings into other homicide categories, an additional or alternative strategy would be for law enforcement agencies to avoid sharing homicide information altogether with the FBI. Table A.9 explores this possibility using the data on agency reporting. Column 1 considers whether county-level participation rates of law enforcement agencies in the FBI's UCR program, a prerequisite to the SHR participation analyzed in previous tables, differ between law enforcement certifying counties and their controls.

Table A.9: Law Enforcement Agency Data Sharing to UCR and NIBRS

Dep. Var.:	UCR (1)	NIBRS (2)	UCR (3)	NIBRS (4)	UCR (5)	NIBRS (6)
LE certify	-0.044**	-0.168***				
	(0.018)	(0.029)				
	[0.015]	[0.000]				
Sher-Coroner			-0.028	-0.039		
			(0.025)	(0.025)		
			[0.275]	[0.115]		
LE certify					-0.089***	-0.353***
No Sher.-Cor.					(0.029)	(0.054)
					[0.004]	[0.000]
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes	Yes	Yes
Spec.	Lin.	Lin.	Lin.	Lin.	Lin.	Lin.
N_{tot}	3289	3289	2128	2128	1056	1056
N_{eff}	3289	3289	2128	2128	1056	1056
N_{treat}	1224	1224	644	644	454	454
$N_{counties}$	470	470	304	304	151	151
$\mu_{outc.}$	0.878	0.527	0.882	0.495	0.867	0.591

Note: Table displays coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by standard error, and p-value. Data covers 2013-2019. Dependent variables: UCR: percentage of law enforcement agencies sharing data to the UCR program. NIBRS: percentage of law enforcement agencies sharing data to the NIBRS program. Panel linear model with year and adjacency county cluster fixed effects. Standard errors clustered at the county level. Set of controls described in Appendix A.5.

A priori, the expected effect on UCR participation is unclear. When facing restrictive death certification laws, some law enforcement agents and agencies may be less able to reclassify police killings as 'circumstances undetermined' homicides. As a result, they may decide to avoid scrutiny by not sharing UCR data entirely, which would predict a negative effect of permissive death certification laws on UCR participation. In contrast, it may be that some law enforcement agencies that reclassify police killings as documented in previous tables will additionally cover their tracks by not sharing UCR data. The results in column 1 suggest that this second theory prevails on aggregate. We find that law enforcement-certifying counties are 4.4 percentage points less likely to share their crime data with the UCR.

Column 2 looks further into differences in the sharing of detailed homicide data by considering submissions of reports to the NIBRS. The results indicate that law enforcement agencies in counties with more lenient death certification laws are 16.8 percentage points less likely to submit a report of the detailed circumstances of death.[7]

Columns 3-6 examine these differences in UCR and NIBRS participation rates for sheriff-coroner and other law enforcement certifying counties separately. Column 7 adds to these results by including state fixed effects in the estimation of NIBRS reporting for non-sheriff law enforcement certifying counties. Our results indicate that differences in participation to both the UCR and NIBRS are mainly driven by counties which permit law enforcement to certify the cause of death but are not sheriff-coroner counties. Considering these results in hand with those of Tables 1.2 and 1.3, we surmise that the main cover-up method in sheriff-coroner counties is to reclassify police killings into the 'circumstances undetermined' homicide category. In addition, law enforcement agencies in sheriff-coroner counties do so without additionally resorting to hiding their homicide data from the public. This less cautious behavior is consistent with the theory that sheriff-coroners, united with police departments behind the 'blue wall of silence', are more likely to stand behind the reclassified cause of death in the event of an external inquiry.

A.15 Underreporting Effects by Race

Table A.10 considers whether the effects uncovered for counties that allow law enforcement to certify the cause of death in Table 1.1 are different by race and ethnicity. The race-ethnicity groupings are chosen by necessity as they are the only ones which allow a direct comparison of the MPV and SHR police killings data. In the top frame, we present results from our adjacent county fixed-effect Poisson specification. Fixed effects Poisson has a drawback when it comes to analyzing effects for different race groups. In our data, officer involved homicides of any specific race, besides White people, are relatively low in absolute terms. This means that many counties will be excluded in the Poisson estimation which drops any fixed effect cluster of adjacent counties with 0 outcomes in all years. To account for this low-observation problem, we also present results from the Poisson specification without fixed effects in our lower panel. The latter will not account for possible unobserved heterogeneity at the county adjacency cluster.

Column 1 presents effects for White people of non-Hispanic ethnicity and column 2

[7] When considering whether law enforcement agencies in counties report the UCR or NIBRS, our results are practically identical to the UCR results of column 1.

presents effects for all other minorities.[8] The results indicate that misreporting of officer-involved fatalities for White people are 50% higher in counties which allow law enforcement to certify the cause of death relative to their controls. The effects in column 2 for minorities are close to 0 and not significant.

Table A.10: Race-ethnicity Group Effects

Dep. Var.:	Diff. L.E. homicides			
Victims:	White	Non-White	Black	Hispanic
	(1)	(2)	(3)	(4)
LE certify	0.498***	0.063	0.213	0.183
	0.404	0.061	0.193	0.168
	(0.145)	(0.211)	(0.313)	(0.312)
	[0.005]	[0.774]	[0.537]	[0.591]
N_{tot}	3289	3289	3289	3289
N_{eff}	2288	1638	735	987
N_{treat}	776	595	287	420
$N_{counties}$	327	234	105	141
$\mu_{outc.}$	0.133	0.250	0.137	0.260
Covariates	Yes	Yes	Yes	Yes
Share_{sample}	0.409	0.356	0.123	0.240
Share_{US}	0.485	0.348	0.209	0.149

Note: Table displays transformed coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by estimated coefficient, its standard error, and its p-value. The dependent variable is *MPV police killings - FBI-SHR police killings* for each race-ethnicity grouping over 2013-2019. We use a Poisson specification with year and adjacency county cluster fixed effects, and standard errors clustered at the county level. Set of controls described in Appendix A.5.

The following two columns further consider the two main minority groups of race-ethnicity, with column 3 showing results for Black people and column 4 showing results for people of Hispanic origin.[9] We do not see any significant difference in the underreporting of police killings between law enforcement certifying counties and their adjacent controls for both race-ethnic groups. Although surprising, the absence of effects for Black people may simply reflect the low share of Black people and the lower share of underreported police killings for Black people in the analysis sample relative to the US population as a whole.

[8] The share of underreported police killings for White people and minority people do not sum up to 1 due to some deaths being classified as Unknown race.

[9] In order to ensure the race and ethnicity categories are comparable in the MPV and SHR data, we exclude Hispanic people from the Black people category by cross-checking names through a genealogy website, https://forebears.io/, and reclassifying any Black person of Spanish or Portuguese name into the Hispanic people category. We similarly reclassify Unknown race people with Spanish or Portuguese names into the Hispanic people category. The Hispanic people category will therefore also include Black race people.

A.16 Underreporting in US-Mexico Border Counties

One consequence of institutions which can be exploited to circumvent accountability is that they will affect those with fewer rights. In particular, families of illegal immigrants killed by police may be less likely to bring a civil case against law enforcement. As such, law enforcement may be more likely to underreport killings of illegal immigrants. To explore, somewhat informally, this question, we use our adjacent sample approach comparing the underreporting of police killings in counties along the US-Mexico border, to their nearest inland neighboring counties.

Table A.11: US-Mexico border effects

Dep. Var.:	Hisp. MPV-SHR	Hisp. SHR	Hisp. MPV	Non-Hisp. MPV-SHR	All SHR
LE certify	0.742 ***	-0.466 ***	-0.014	-0.387	0.062
	0.555	-0.628	-0.014	-0.490	0.060
	(0.164)	(0.185)	(0.139)	(0.158)	(0.150)
	[0.001]	[0.001]	[0.922]	[0.002]	[0.689]
Covariates	Yes	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes	Yes
Spec.	Pois.	Pois.	Pois.	Pois.	Pois.
N_{tot}	546	546	546	546	546
N_{eff}	532	308	539	546	448
N_{treat}	203	189	203	203	196
$N_{counties}$	76	44	77	78	64
$\mu_{outc.}$	0.400	1.013	0.970	0.332	1.605

Note: Table displays transformed coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by estimated coefficient, its standard error, and its p-value. Dependent variable: (1) *MPV police killings - FBI-SHR police killings* for Hispanic people, (2) Reported SHR police killings for Hispanic people, (3) Reported MPV police killings for Hispanic people, (4) *MPV police killings - FBI-SHR police killings* for non-Hispanic people (5) Reported SHR police killings for all race-ethnicity groups. Poisson estimation with year and adjacency county cluster fixed effects, and standard errors clustered at the county level. Set of controls described in Appendix A.5.

Table A.11 presents these US-Mexico border results. To increase sample size, we include second degree adjacency counties in control counties, as presented in map A.10 of Appendix A.3. Column 1 of Table A.11 shows that counties bordering Mexico are 74% more likely to underreport police killings of people of Hispanic origin. Columns 2 and 3 show that this effect is driven by lower law enforcement agency reporting of Hispanic victims to the SHR, rather than higher true police killings in the MPV data. Column 4 adds to these results showing that, for non-Hispanic people, US-Mexico bordering counties have fewer underreported police killings. This difference cannot be explained by population differences, since Mexico bordering counties display very similar average populations of Hispanic people relative to their controls and, in fact, have slightly higher populations of non-Hispanic people.

One explanation, consistent with previous results and the results of column 5, showing that there is no difference in overall reported SHR police killings, is that police agencies are aware and careful to not report excessive police killings relative to agencies in their neighboring counties. They do, however, use discretion, possibly based on their chances of coming under scrutiny, to decide which police killings to hide from their SHR reports.

A.17 Additional Differences in Death Investigation Systems

As discussed previously, different death investigation systems and lead examiners may differ along several dimensions. As statistically described in Table A.3 of Appendix A.5 and represented in maps in Appendix A.3, they may be coroner or medical examiner counties, the death investigator may or may not be required to be a physician, the coroners may be appointed or elected, and they may be required to hold a high-school degree or not. Table A.12 explores whether these partially overlapping differences in laws and systems lead to differences in the underreporting of police killings. Each analysis restricts the sample to adjacent treated and control counties with different death investigation systems and excludes counties in which law enforcement can certify the cause of death.

Column 1 compares the underreporting of police killings between coroner and medical examiner counties but finds no significant difference. Speaking again to the question of competence, column 2 considers differences in underreporting between counties which require a physician to conduct an autopsy and counties allowing non-physicians to perform autopsies. Although the effects are marginally significant at the 10% level, we do not find these robust to specifications which match physician and non-physician counties on urbanization levels, nor robust to the inclusion of state fixed effects. Neither of these results provide convincing evidence that higher medical competence shields death investigators from intimidation and external pressure to alter their autopsy report or change the cause of death. These results are in line with the nationwide comparison of underreported police killings in medical examiner and coroner counties in GBD et al. (2021) using NVSS data.

Column 3 considers whether counties in which the death investigator is appointed rather than elected have higher police killings. Theoretically, the expected effect of the manner of selection is unclear. Elected officials may be more willing to uprightly serve their voting constituents by opposing any outside pressure to change death reports. They may, however, also want to conform to pressures by local politicians and police if they want to receive political support during elections. Ultimately, the results indicate no difference in the underreporting of police killings between elected and appointed counties. This may be because both stated effects cancel out, or simply because the election vs appointment selection process of the death investigator is not of primary importance.

Column 4 considers whether state laws requiring the death investigator to hold a high school diploma influence the reporting of police killings. Again, the results do not appear to show any difference on this binary measure.

Taken together, the results from Table A.12 do not indicate that our main results on underreporting depending on law enforcement death certification laws are driven by other underlying differences in death investigator characteristics or systems.

Table A.12: Coroner vs Medical Examiner, Appointed vs Elected, Physician vs Non-Physician, HS diploma vs No HS diploma

	Cor vs ME	Phys vs non-Phys	App vs Elec	HS vs no-HS
Treatment	-0.038	0.164*	-0.006	0.116
	-0.039	0.152	-0.006	0.110
	(0.089)	(0.083)	(0.096)	(0.103)
	[0.663]	[0.066]	[0.950]	[0.286]
Covariates	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes
Spec.	Pois.	Pois.	Pois.	Pois.
N_{tot}	6,223	7315	2,764	3400
N_{eff}	5,271	6034	2,190	2867
N_{treat}	3010	2247	356	243
$N_{counties}$	753	862	313	410
$\mu_{outc.}$	0.318	0.337	0.215	0.272

Note: Table displays transformed coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by estimated coefficient, its standard error, and its p-value. Analysis over years 2013-2019. (1) Coroner vs Medical Examiner adjacency sample, Treatment=Coroner, (2) Appointed vs Elected adjacency sample, Treatment=Appointed, (3) Physician vs non-Physician adjacency sample, Treatment=Physician, (4) High School diploma vs no High School diploma sample, Treatment=High School diploma. Poisson estimation with year and adjacency county cluster fixed effects, and standard errors clustered at the county level. Set of controls described in Appendix A.5.

A.18 Responses to Monitoring and Threats

In this section, we explore whether law enforcement agencies and agents in counties permitting them to certify the cause of death are more likely to resist the introduction of monitoring and accountability measures, in particular the use of body-worn cameras and the threat of charging police officers. The analysis further considers the effect of certification laws on homicide clearance rates, with a view to understanding whether additional scrutiny through stricter certification laws lowers police effectiveness.

We then move on to examine more generally, and not causally, whether modern-day perceived and actual threats to law enforcement are related to differences in underreporting. We first consider whether police in counties with permissive death certification laws are more likely to be assaulted or killed while on duty. Then, looking at nationwide correlations, we assess whether the permissiveness of gun and ammunition laws is related to the underreporting of police killings. In a last step, we consider whether awareness and positive or negative concern for issues raised by the BLM movement, as proxied by Google search trends and their changes, are associated with the underreporting of police killings.

Body-worn Cameras

Body-worn cameras have been proposed as an important tool to help monitor police activity. In the context of underreported police killings, body-cameras may prevent otherwise unverifiable cover-ups. Despite this promise, experimental and non-experimental results on the effects of body-worn cameras are mixed (Lum et al., 2020; Williams Jr et al., 2021). One stated problem is the apparent resistance to adopting body-worn cameras. Here, we explore whether this resistance is correlated to law enforcement death certification laws. In general, US-wide county-level information on body-worn cameras is limited. The most complete county-level source is the Law Enforcement Management and Administrative Statistics Body-Worn Camera Supplement (LEMAS-BWCS) census conducted in 2016. This census asks a random sample of law enforcement agencies when they began using body-worn cameras. It also asks questions regarding when these cameras are required to be worn.

The relation between the underreporting of police killings, whether law enforcement can certify death, the adoption and subsequent utilization of body cameras, and census response rates, can operate in many potentially endogenous ways. Laws pertaining to death certification may have, over years, contributed to cultures of impunity, rendering law enforcement agencies less likely to adopt cameras. As previously documented with the NIBRS, these law enforcement agencies may also be less willing to share detailed information on their operations. In addition, we only have a measure of whether and how many body-worn cameras were purchased, but agents may selectively choose when to turn them on. Despite these issues, which we make no claim to address adequately in this paper, we offer some first insights on how death certification laws may influence the adoption of body-cameras.

Column 1 of Table A.13 considers whether permissive death certification laws are correlated with the probability that law enforcement agencies within a county adopt body-worn cameras. We take as the county outcome variable for body-worn cameras an indicator equal to 1 if the share of law enforcement agencies in a county that claims to operate with body-worn cameras is above the national mean of 6%, and 0 otherwise.[10] Because the census was administered in 2016, we also drop any year thereafter from the analysis.

The results in column 1 do not suggest that lenient death-certification laws influence the adoption of body-worn cameras. Although not presented in tables, we also explored heterogeneous effects of law enforcement certification laws, interacted with our body-worn camera variable, on the underreporting of police killings, but found no significant heterogeneous effects. These stated results pertain to a noisy measure of body-worn camera availability at the extensive margin. Further inquiries into the intensive use of body-worn cameras may produce additional insights.

Table A.13: Monitoring and Threat Effects

	LEMAS	UCR		LEOKA
Dep. Var.:	Body-cam	Clearances	Clearance rate	Assaults
	(1)	(2)	(3)	(4)
LE certify	0.012	-0.023	0.020	0.016
		-0.023		0.016
	(0.033)	(0.052)	(0.099)	(0.101)
	[0.714]	[0.653]	[0.838]	[0.878]
Covariates	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes
Spec.	Lin.	Pois.	Quasi-bin.	Pois.
N_{tot}	1880	3289	3289	3289
N_{eff}	1880	3044	1292	3163
N_{treat}	700	1077	506	1161
$N_{counties}$	470	435	317	452
$\mu_{outc.}$	0.219	3.479	0.558	31.226

Note: Table displays in columns 2 and 4 the transformed coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by estimated coefficient, its standard error, and its p-value. Analysis over years 2013-2019. Dependent variables: (1) LEMAS data: 1 if the share of law enforcement agencies in a county claiming to operate with body-worn cameras is above the national mean of 6%, 0 otherwise, (2) UCR data: Total clearances of UCR homicides, (3) UCR data: Clearance rate of UCR homicides (4) LEOKA data: Total assaults on law enforcement officers. Set of controls described in Appendix A.5.

Charging Officers

Speaking to another form of accountability, we also considered whether charging an officer had any effect on the underreporting of police killings in subsequent years. Unfortunately, an officer was charged in only 3.5% of true MPV police killing cases in our neighboring county sample, and one-third of those resulted in an acquittal or dropped charge(s). As a

[10] Approximately 23% of law enforcement agencies responded to the LEMAS-BWCS (only 30% were sampled) which explains the relatively low share of law enforcement agencies operating with body-worn cameras.

result, we only observe 18 charged or charged-convicted officers in our adjacency analysis sample over the years 2013-2019, too few for any rigorous analysis.[11]

Police Effectiveness

Raising the expected penalty of an officer's errors through more stringent monitoring laws may, in theory, backfire and reduce the effectiveness of police (Prendergast, 2003; Gavazza and Lizzeri, 2007). Some empirical evidence exists supporting this view (Shi, 2008; Devi and Fryer Jr, 2020; Campbell, 2022). We consider here whether more lenient law enforcement certifying laws affect police performance by looking at total homicide clearances and clearance rates obtained from the UCR. Our results in Table A.13 columns 2-3 do not offer any evidence that more stringent death certification laws influence police effectiveness.

Assaults on Police Officers

Our base independence assumption is that prior to any changes in death investigation systems and laws in the 1960s-1980s, adjacent treatment and control counties in our analysis sample were comparable. However, after changing the system, different policing and reporting approaches may have influenced the general hostility in a county. Homicide cover-ups by police in the late 1980s and early 1990s in counties permitting law enforcement to certify the cause of death may have incited more violent behavior towards police, which itself induced police to respond with more unwarranted lethal violence, requiring additional cover-ups of police killings.

Table A.13, column 4, assesses evidence of such a cyclical process by looking at whether police in counties that permit law enforcement to certify the cause of death face more hostile environments, in the form of violent assaults. Yearly numbers of law enforcement officers who have been victims of assaults while on duty in each county are gathered from the Law Enforcement Officers Killed and Assaulted (LEOKA) dataset.[12] Column 4 shows that officers in law enforcement certifying counties are no more likely to be assaulted while on duty than officers in their adjacent control counties. These results also suggest that higher lethal police action, resulting in higher unjustified killings and their cover-ups, are not a response to more hostile environments in law enforcement certifying counties.

[11] Overall in the US over that same period, we only observe 91 charged or charged-convicted officers out of 3611 police killings, or 2.5%.

[12] This outcome is right skewed, as shown in Appendix A.12. We also show maps of the averages by county over all years for these two variables.

Nationwide Threats and Underreporting

Departing from any causal interpretation, our data also offer a chance to describe the nationwide relation between underreported police killings and other policy relevant variables. In particular, assuming that underreported police killings are a close proxy to unwarranted police killings, our data can offer a first look into the relation between unwarranted police killings and threats on police, both actual and potential.

Table A.14 column 1 presents Poisson model estimates of underreported police killings on the arcsinh transformed value of violent assaults on police, the same variable described in the previous section. The result, controlling for our demographic, cultural, and economic covariates, does not seem to suggest that there is a US-wide correlation between these threats to police and the underreporting of police killings.[13]

Table A.14: Threats on Police and Underreporting

Dep. Var.:	MPV-SHR	MPV-SHR	SHR	MPV
	(1)	(2)	(3)	(4)
Assaults	-0.081 -0.085 (0.090) [0.347]			
Gun laws		-0.171*** -0.187 (0.031) [0.000]	0.050 0.049 (0.061) [0.423]	-0.079*** -0.082 (0.026) [0.002]
Covariates	Yes	Yes	Yes	Yes
FE	No	No	No	No
Spec.	Pois.	Pois.	Pois.	Pois.
N_{tot}	21993	21993	21993	21993
N_{eff}	21993	21993	21993	21993
$N_{counties}$	3142	3142	3142	3142
$\mu_{outc.}$	0.220	0.220	0.138	0.347

Note: Table displays transformed coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by estimated coefficient, its standard error, and its p-value. Analysis over years 2013-2019. Dependent variable: (1 and 2) *MPV police killings - FBI-SHR police killings*, (3) *FBI-SHR police killings*, (4) *MPV police killings*. Poisson estimation with year-fixed effects and standard errors clustered at the county level. Set of controls described in Appendix A.5.

Columns 2-4 of Table A.14 examine whether potential threats, proxied by state-level gun and ammunition laws and regulations, correlate with underreported police killings. In theory, operating in an environment where citizens have a higher probability of carrying a gun can lead to premature and excessive use of lethal force. As a result, we would expect to see less underreporting of police killings in states with stricter gun regulation laws.

Our proxy for state-level gun and ammunition regulations comes from Giffords Law Center's generalized gun regulation scorecard. This is a yearly measure for each state of

[13] Although assaults on police do correlate with higher reported SHR and true MPV police killings.

the strictness of gun and ammunition regulation ranging from 0-11 where 11 corresponds to a score of A and 0 to a score of F. The yearly state score weighs the many laws pertaining to ammunition and gun possession, distribution, and right-to-carry within a state. We describe and map the average state score in Appendix A.12.

The result in column 2 of Table A.14 indicates that stricter gun and ammunition laws are correlated to lower underreporting of police killings. Columns 3 and 4 further show that this relation is due to strict gun and ammunition laws being associated with lower true police killings rather than higher reported killings. Even if only correlations, these results are concerning if we interpret the underreporting outcome as a proxy to unwarranted police killings. More formal causal evaluations would be necessary to confirm whether law enforcement agents respond to the potential threat of working in permissive gun law state by using excess lethal violence, resulting in unwarranted police killings which they subsequently hide.

Black Lives Matter and Underreporting

Another question of interest when it comes to correlation patterns is how the underreporting of police killings has been affected by the Black Lives Matter movement. A variety of measures could be used to assess the salience of the BLM movement in particular areas. Among these are the proximity to highly publicized police-killings, the scale of local BLM protests, or the general awareness and concern for issues raised by the BLM movement. The current discussion only considers this last question of awareness and concern, which we proxy by the trends in Google searches for the topic 'Black Lives Matter' in each state-year. We construct this measure by standardizing the Google trend to range from $[0, 100]$, with 100 being the highest state-year per capita search record.[14]

Taking the continuous measure of Google trends, Column 1 of Table A.15 shows that, nationwide, states with higher searches for 'Black Lives Matter' topics show no correlation with the underreporting of police killings. However, these null results may not fully describe the relationship between concern for the BLM movement and the underreporting of police killings. In fact, the Google trends measure shows a bimodal distribution as presented in Appendix A.13. It may be that the BLM movement only influences the underreporting of police killings beyond a certain threshold of public concern, whether favorable or unfavorable.

[14] More precisely, we download each year's Google trend by state ranging from $[0, 100]$, and multiply each year-state Google trend value by the ratio of that year's Google trend searches to total searches over 2013-2019, as presented in Appendix A.13.

We explore this possibility in columns 2-5. Defining a low-high binary value for BLM concern, we find that high BLM concern in a given state[15] is associated with more underreporting. Column 3 includes state-fixed effects, thereby considering the relation between changes from low-to-high or high-to-low BLM concern on changes in underreporting. There does not appear to be any relation in terms of changes. Among other explanations, the correlations from columns 2-3 may indicate that the higher concern for the BLM movement puts pressure on police to reduce visible police killings more than it reduces actual killings. Another possibility we certainly do not exclude is that states with high per capita Google searches for 'Black Lives Matter' are distinct in several dimensions affecting the underreporting of police killings.

Columns 3-4 further explore the heterogeneous effects of BLM concern in our adjacent county sample. Column 4, without state-fixed effects, and column 5, with state fixed effects, indicate that counties permitting law enforcement to certify the cause of death and with high concern for the BLM movement display higher underreporting of police killings compared to the interaction effect for counties in states with low concern for the BLM movement.

These effects are consistent with previous findings of strategic manipulation of death records. Assuming Google trends are not only a proxy for the concern for the BLM movement, but also a credible measure of additional scrutiny on police, the results suggest that police departments in counties with permissive certification laws and under additional scrutiny are even more strategic in hiding police killings than those facing less scrutiny. If instead the BLM proxy reflects searches from people opposed to the BLM movement, or results in a more negative view of the BLM movement, a negative view known by the police, then the results may indicate that law enforcement agencies take advantage of the sentiment against the BLM movement to further manipulate death records. Further investigations on this matter are important, but beyond the scope of this paper.

[15]This is defined as a value of the Google trend above 20.

Table A.15: Black Lives Matter and Underreporting

	MPV-SHR	MPV-SHR	MPV-SHR	MPV-SHR	MPV
Dep. Var.:	MPV-SHR	MPV-SHR	MPV-SHR	MPV-SHR	MPV
	(1)	(2)	(3)	(4)	(5)
BLM_{con}	0.004 0.004 (0.005) [0.444]				
BLM_{20+}		0.269* 0.238 (0.125) [0.057]	0.090 0.086 (0.111) [0.438]	-0.136 -0.146 (0.310) [0.637]	-0.120 -0.128 (0.316) [0.686]
LE certify if $BLM_{20+} = 0$				0.324** 0.281 (0.134) [0.036]	0.504** 0.408 (0.184) [0.026]
LE certify if $BLM_{20+} = 1$				0.831*** 0.605 (0.224) [0.007]	1.085*** 0.735 (0.267) [0.006]
Covariates	Yes	Yes	Yes	Yes	Yes
FE	No	No	St	Cl	Cl+St
Spec.	Pois.	Pois.	Pois.	Pois.	Pois.
N_{tot}	21993	21993	21993	3289	3289
N_{eff}	21,433	21,433	21,433	2561	2561
N_{treat}	2526	2526	2526	874	874
$N_{counties}$	3142	3142	3142	366	366
$\mu_{outc.}$	0.22	0.22	0.22	0.286	0.286

Note: Table displays the transformed coefficient with *p<0.1; **p<0.05; ***p<0.01, followed by estimated coefficient, its standard error, and its p-value. Dependent variable: (1 to 4) *MPV police killings - FBI-SHR police killings*, (5) *MPV police killings*. Year fixed effects included in all specifications. *BLM* specified as continuous variable in $[0, 100]$ in column 1 and as a binary indicator equal to 1 when the continuous measure exceeds 20, and 0 otherwise in columns 2-5. Heterogeneity terms are specified as separate, not cumulative, effects depending on whether the binary *BLM* is equal to 0 or 1. Standard errors clustered at the county level. Set of controls described in Appendix A.5.

Appendix B

Appendix to Chapter 2

APPENDIX

B.1 Contemporaneous accounts and displacement patterns

This appendix provides additional historical and qualitative context on the 1998–1999 conflict. It first outlines the broader political chronology and international legal debate around NATO's intervention, and then collects contemporaneous reports documenting expulsions and abuses.

When the Socialist Federal Republic of Yugoslavia (SFRY) began to dissolve in the early 1990s, wars broke out as republics declared independence. Serbia and Montenegro later formed the Federal Republic of Yugoslavia (FRY) in 1992.[1]

The Kosovo case was particularly unusual. Unlike Bosnia-Herzegovina, where NATO had a clearer mandate from the UN, the Kosovo intervention lacked explicit authorization from the UN Security Council. Operation Allied Force marked NATO's first military intervention without UN approval, justified on humanitarian grounds in response to ethnic violence, human rights abuses, and a growing humanitarian crisis (Henkin, 1999; Kritsiotis, 2000; Newman and Visoka, 2024). NATO framed the operation as necessary to prevent further atrocities and to protect civilian populations.

Despite the intention to prevent further violence, the intervention had heavy repercussions on the pace and modalities through which Serbian forces drove Kosovar-Albanians into neighboring countries. Human Rights Watch and contemporaneous observers documented mass expulsions, village burnings, and the destruction of identity records. Hun-

[1]SFRY refers to the pre-1990 socialist federation (six republics and two autonomous provinces, Kosovo and Vojvodina). After secessions in 1991–1992, Serbia and Montenegro constituted the FRY.

dreds of thousands fled to neighboring Albania and Macedonia.

UNHCR (30 March 1999)– Reported approximately 94,000 Kosovar-Albanians had fled since 24 March 1999 and described patterns of forced removals. Following a year of persistent ethnic tensions and violent clashes between Albanians and Serbs, coupled with failed diplomatic efforts, the UN High Commissioner for Refugees (UNHCR) announced on 30 March 1999 that approximately 94,000 ethnic Albanians had been forced to flee Kosovo since NATO military action began on 24 March 1999. According to a UNHCR representative cited in The Guardian, the emerging pattern involved paramilitary forces arriving, forcibly gathering civilians at gunpoint, and ordering them to leave. The situation was characterized as 'officially sanctioned ethnic cleansing of the Albanian population'[2].

NATO also accused the authorities in Belgrade of trying to erase the identities of thousands of ethnic Albanians by destroying their property deeds, birth and marriage certificates, and other records[3].

OSCE (1999)– a report from the Organization for Security and Co-operation in Europe (OSCE, 1999) on patterns of human rights and humanitarian law violations in Kosovo confirms the entity of the issue: "After the start of the NATO bombing on the FRY on 24 March, Serbian police and/or VJ (Yugoslav Army), often accompanied by paramilitaries, went from village to village and, in the towns, from area to area threatening and expelling the Kosovar-Albanian population. Others who were not directly forcibly expelled fled as a result of the climate of terror created by the systematic beatings, harassment, arrests, killings, shelling, and looting carried out across the province. Kosovar-Albanians were clearly targeted for expulsion because of their ethnicity. [...] Large numbers of civilians were also deliberately targeted and killed because of their ethnicity. No one, it seems, was immune, as people of all ages, including women and children, were killed in large numbers." These testimonies suggest that soldiers were forcibly displacing Kosovar-Albanians from their homes under the pretext of humanitarian aid allegedly awaiting them at the border. Considering these factors, the rationale for implementing Operation Allied Force becomes evident.

Human Rights Watch– A witness statement cited in the Human Rights Watch report (Abrahams, 2001) recounts how Serbian forces attacked a village using tanks and trucks,

[2]Guardian, 30 March 1999

[3]In the United Kingdom's House of Commons the Defence Secretary said that the actions of Belgrade's forces in Kosovo were not just murder but premeditated murderYoungs et al. (1999), The Guardian 3 April 1999

indiscriminately shooting, setting homes on fire, and physically assaulting residents. Civilians were reportedly told to evacuate to Albania under threat of death. Refugees described harrowing experiences as they fled: "I left with only the clothes on my back, with my two children. We walked for hours, hiding in the woods. The Serbian soldiers took everything from us, even the little food we had. It was a nightmare. We just wanted to survive." (Abrahams, 2001).

Press reporting – See contemporaneous coverage summarizing UNHCR statements and patterns of expulsions. For example, according to a UNHCR representative cited in The Guardian, the emerging pattern involved paramilitary forces arriving, forcibly gathering civilians at gunpoint, and ordering them to leave. The situation was characterized as 'officially sanctioned ethnic cleansing of the Albanian population'.

B.2 Kosovo waste management historic development

Kosovo's waste management system faces significant challenges, representing one of the country's foremost issues. The inadequate collection and treatment of waste accumulated over the past decade have resulted in poor aesthetics and environmental pollution. Furthermore, the waste management infrastructure has yet to extend its coverage to encompass the entire nation. In the reporting year of 2019, the waste collection service coverage rate for households was 78.5%, marking an increase of 2.9% at the national level compared to 2018 (Kosovo Environmental Protection Agency, 2019; Agjencia për Mbrojtjen e Mjedisit të Kosovës, 2018). The path to an up to standard waste infrastructure is still long and tortuous. Before the 1999 conflict, waste management in Kosovo was part of the centralized Yugoslav system, with Kosovo relying on Serbia for key infrastructural and policy decisions. However, Kosovo was underfunded and neglected, resulting in poor waste management services and significant environmental degradation, especially in rural and industrial areas. This situation worsened during the 1990s as political tensions escalated between Kosovo and Serbia. The mining and heavy metal industries played a significant role in the poor management of waste in Kosovo, especially before the 1999 conflict. Kosovo's economy historically relied heavily on mining, particularly through the Trepča mining complex, one of the largest industrial complexes in the region, which extracted lead, zinc, silver, and other valuable minerals. In 2000 the European Union built the European Agency for Reconstruction (EAR) to manage EU-funded reconstruction efforts in Kosovo. One of its focuses was on environmental rehabilitation and industrial clean-up, especially addressing the pollution caused by decades of poor waste management and industrial activities. Despite being built or renovated between 2003 and 2007 from the European Agency for Reconstruction, says the Ministry of Environment and Spatial Planning (2019), the infrastructure is in poor condition. Focusing on the population level, less than 60% of the population receives a regular waste collection service, which is even lower in rural where only 40% of the generated waste is collected and disposed.[4] The lack of a waste management system testified by Morina et al. (2017); Alite et al. (2023), is the result of several factors such as the lack of network management, lack of technical and financial means, low-rate payment, and lack of waste law implementation. The European Environment Agency (2021) report explains why Kosovo did not meet most of the targets related to waste management set for 2020 and was not on track to meet draft targets set for 2021/2022[5]. The key challenges of waste management in Kosovo are related to insufficient cross-institutional cooperation, budget deficiencies, a lack of (trained) staff, extensive informal sector activities, the poor financial viability of waste management operations, lack of investment in infrastructure, low levels of public awareness and poor enforcement of laws. These challenges are all to some extent interlinked. Despite efforts by the Ministry of Environment and Spatial Planning and the Kosovo Environmental Protection Agency (KEPA) to monitor the annual creation of illegal landfills, mandated by the Waste Law which requires municipalities to compile a registry of such sites within their jurisdictions, the illegal dumps are just one of the issues of the waste management system. Something that is understudied is the maintenance and up keeping of the current municipal, regional, transit and industrial landfills. Out of the ten landfills (five municipal, two regional and three transit deposits) constructed after the war, just five are managed by a state-owned company (Kompania per menaxhimin e deponive ne kosove) while for the other ones the management is still quite ambiguous. For example, the Municipal Landfill of Mitrovica, was established in 1998 by the Serb authorities and was legalised in 2002 by the Danish KFOR[6]. Despite the post-war legalization, no effort was made whatsoever to regulate the dump. It was supposed to have a 15-year lifespan but in 2020 it was still operative. Despite the numerous requests from residents to sensitize authorities[7] on the water pipes serving

[4]Recycling rate is very low, Agjencia për Mbrojtjen e Mjedisit të Kosovës (2018) says then about 15% of the total waste is recycled. The recycling infrastructure is underdeveloped, limiting the potential for material recovery and reuse.

[5]Agjencia për Mbrojtjen e Mjedisit të Kosovës (2018): The waste sector contributes to approximately 4% of Kosovo's total CO2 emissions, indicating the need for improved waste management practices to mitigate climate impact

[6]KFOR is the Kosovo Force, NATO-led international peacekeeping force in Kosovo

[7]Residents of the village of Koshtovë in Mitrovica have been facing problems related to an illegal waste dump since 1998. Despite opposition from the local residents, the Serbian authorities at the time established this landfill - Telegrafi

200 families passing through the area and the emission affecting the air quality, nothing was done. Similar stories refer to other waste landfills operating in the country. The Ministry of Environment and Spatial Planning (2019) states that there are seven sanitary landfills that receive an estimated share of less than 40% of the municipal waste. These were funded by the European Agency for Reconstruction between 2003 and 2007 and were supposed to operate in accordance with the EU standards. But of these seven sanitary landfills, almost all have reached their maximum capacity, so they have to be closed urgently. There are concrete plans to close the Mirash landfill in the Pristina region with the support of the German state-owned development bank KfW (Kosovo Environmental Protection Agency, 2021). Other landfills, such as the ones in Peja and Prizren, are planned to be extended, as stipulated under the Integrated Waste Management Strategy 2021-2030.

It is also worth considering the price of this inadequate system is pretty costly for the population. As reported by Kosovo Environmental Protection Agency (2021); European Environment Agency (2021), the fee is a rate that varies across the municipalities but usually lies at EUR 60 per household per year, roughly 1% of the yearly net salary of the average population. In addition to the household fees, there are fees for waste collection services for businesses and institutions, in fact, some landfills request gate fees, usually around EUR 7 per tonne of waste. This covers only the basic cost of managing the landfill, without making any provisions for upgrading or aftercare. These revenues cover only the collection service and the remaining disposal costs have to be obtained from other funding sources. Half of the funds come from the Ministry of Economy and Environment and the other half is received from project grants, mainly from GIZ-supported projects[8] funded by the European Commission. In 2017, performance grants were introduced to stimulate municipalities with financial incentives to provide better data. Although there are no official estimates of uncollected waste, it is estimated that currently 70% of the total waste generated is actually collected (Euroepan Commission, 2020). Furthermore, as no other treatment infrastructure exists yet, all collected waste ends up in landfills (Ministry of Environment and Spatial Planning, 2019).

[8]Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) is the German development agency. GIZ is a federal enterprise owned by the German government, working in international cooperation for sustainable development. It operates globally, providing expertise in areas such as economic development, environmental protection, and governance, often partnering with organizations, governments, and NGOs to implement development projects.

B.2.1 Current waste system: coverage, assets, and finance

Asset governance and maintenance. Of the ten landfills constructed after the war (five municipal, two regional, three transit deposits), only five are managed by the state-owned landfill company (Kompania për Menaxhimin e Deponive në Kosovë); governance of the remainder is ambiguous (Ministry of Environment and Spatial Planning, 2019). Maintenance and operational standards vary across sites and over time.

Illustrative case: Mitrovica/Koshtovë. The Mitrovica municipal landfill, established in 1998 and legalized in 2002 by Danish KFOR, remained in operation in 2020 beyond its planned 15-year lifespan, amid local concerns about water infrastructure and air emissions.[9] Similar issues are reported at other sites.

Financing and cost recovery. Fees vary by municipality but are typically about EUR 60 per household per year, with additional charges for businesses and institutions; some landfills apply gate fees near EUR 7 per tonne (Kosovo Environmental Protection Agency, 2021; European Environment Agency, 2021). These revenues generally cover collection and basic operations but rarely fund upgrades or aftercare; the residual is financed via central budget allocations and externally funded projects (e.g., EU/GIZ programmes). Performance grants introduced in 2017 incentivize improved municipal reporting (Euroepan Commission, 2020).

Challenge taxonomy (EEA). European Environment Agency (2021) highlight: (i) insufficient cross-institutional cooperation, (ii) budget deficiencies and low cost recovery, (iii) staffing shortages and training needs, (iv) extensive informal sector activity, (v) underinvestment in infrastructure, (vi) low public awareness, and (vii) weak enforcement. These constraints are mutually reinforcing.

B.3 Bombing randomness

Whether pre-existing municipal differences could confound the estimated bombing effects was assessed via a two-sample t-test assuming equal variances. The results, presented in Table B.1 indicate that the difference in mean night light intensity between treated and untreated groups is not statistically significant at conventional levels. This suggests that, prior to the bombing, there were no systematic differences in night light intensity

[9] Telegrafi report on the Koshtovë/Mitrovica site.

Figure B.1: Pre- and post-trend of bombing

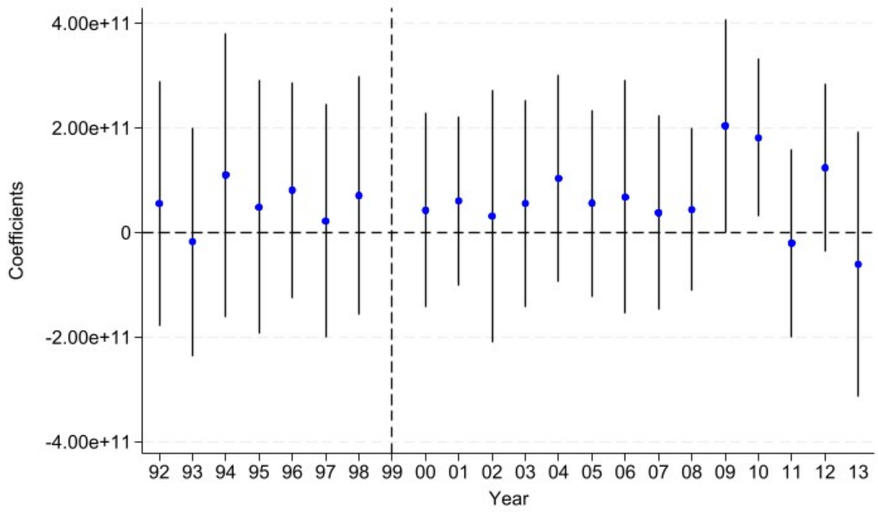


Table B.1: Night Lights pre and after conflict

	(1) Baseline	(2) Year FE	(3) Year-Municipality FE	(4) Balanced panel	(5) Unbalanced panel
High bomb	6.56	6.56		6.56	6.56 ***
	(4.56)	(4.62)		(4.61)	(2.21)
After	1.37			1.37	
	(2.69)			(4.13)	
High bomb × After	1.15	1.15	1.15	1.15	1.15
	(5.44)	(5.50)	(5.50)	(7.97)	(2.77)
N	814	814	814	74	44
N clust	37	37	37		

Notes: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

that could have biased our estimates. The model accounts for unobserved time-invariant heterogeneity at the cluster level while also controlling for time-fixed effects. Standard errors are clustered at the municipality level. Overall, the results show no significant differences in night light intensity in the years leading up to the bombing, reinforcing the validity of the identification strategy. Although there is some evidence of increased light intensity in treated areas only after a decade, the lack of significance in earlier periods suggests that any observed post-bombing effects are unlikely to be driven by pre-existing differences in economic activity or infrastructure development.

B.4 Statistical differences

Table B.2: Statistical Differences

	Panel A: Bombing intensity (All)				Panel B: Proximity to landfill				Panel C: Combined (close × Bombing)			
	Low Bomb	High Bomb	Diff	N	Far	close	Diff	N	Low Bomb*	High Bomb*	Diff	N
Infant mort	0.028	0.030	0.001	5373	0.029	0.028	-0.001	5373	0.026	0.031	0.005	5373
			(0.25)				(-0.23)				(1.09)	
Close	0.386	1.000	0.614***	5373					0.594	0.506	-0.087***	5373
			(47.51)								(-6.43)	
High Bomb	0.350	1.000	0.650***	5373	0.570	0.482	-0.088***	5373				
			(51.35)				(-6.43)					
Postdump	0.448	0.420	-0.028*	5373	0.472	0.415	-0.057***	5373	0.425	0.455	0.030**	5373
			(-1.85)				(-4.17)				(2.23)	
Rural	0.556	0.419	-0.138***	5373	0.671	0.396	-0.275***	5373	0.491	0.546	0.055***	5373
			(-8.98)				(-20.84)				(4.05)	
Albanian	0.924	0.884	-0.040***	5373	0.925	0.904	-0.022***	5373	0.936	0.893	-0.043***	5373
			(-4.58)				(-2.80)				(-5.66)	
Serb	0.030	0.000	-0.030***	5373	0.016	0.026	0.010**	5373	0.040	0.005	-0.035***	5373
			(-6.61)				(2.49)				(-8.92)	
Other ethn	0.046	0.116	0.070***	5373	0.058	0.070	0.012*	5373	0.023	0.102	0.079***	5373
			(9.23)				(1.71)				(11.90)	
No educ	0.027	0.025	-0.001	5373	0.024	0.028	0.004	5373	0.029	0.024	-0.005	5373
			(-0.24)				(0.81)				(-1.12)	
Primary educ	0.051	0.054	0.003	5373	0.058	0.046	-0.013**	5373	0.042	0.060	0.018***	5373
			(0.44)				(-2.15)				(2.98)	
Second educ	0.537	0.569	0.032**	5373	0.572	0.523	-0.049***	5373	0.507	0.581	0.073***	5373
			(2.11)				(-3.61)				(5.42)	
Up Second educ	0.265	0.257	-0.008	5373	0.242	0.280	0.038***	5373	0.281	0.246	-0.035***	5373
			(-0.56)				(3.12)				(-2.94)	
Poor	0.213	0.222	0.009	5373	0.225	0.207	-0.018	5373	0.191	0.238	0.047***	5373
			(0.69)				(-1.58)				(4.20)	
Middle class	0.202	0.178	-0.023*	5373	0.216	0.179	-0.037***	5373	0.194	0.197	0.003	5373
			(-1.90)				(-3.43)				(0.30)	
Upper middle class	0.186	0.217	0.031**	5373	0.201	0.189	-0.012	5373	0.188	0.200	0.012	5373
			(2.53)				(-1.12)				(1.07)	
Rich	0.195	0.189	-0.006	5373	0.191	0.195	0.004	5373	0.203	0.184	-0.019*	5373
			(-0.48)				(0.35)				(-1.75)	

Notes: Panel A compares municipalities by bombing intensity (low vs. high). Panel B compares survey clusters by proximity to landfill (far vs. close). Panel C combines both dimensions, comparing clusters that are both close to a landfill and in high-bombing municipalities. "Other ethn" includes Bosnians, Ashkali, Egyptians, Gorani, Roma, and other smaller ethnic groups.

B.5 Callaway–Sant'Anna DiD: Subgroup and Negative-Control Evidence

Figures B.2 and B.3 illustrate average treatment effect on the treated (ATT) estimates by proximity to the landfill. For 'close = 1" (Figure B.2), pre-treatment ATT estimates are negative, indicating no upward pre-trend in infant mortality and supporting parallel trends. Post-treatment, ATT estimates turn positive and remain sizable, suggesting a sustained increase in infant mortality. In contrast, Figure B.3 ('close = 0") serves as a negative-control subgroup: ATT estimates are consistently close to zero in both pre- and post-treatment periods, indicating no meaningful effect among those living farther from the landfill. For 'close = 1," the ATT is 0.0391 ($p = 0.043$), with a marginally significant post-treatment average (0.030, $p = 0.062$) and a downward pre-treatment average (-0.024, $p = 0.017$). In the negative-control subgroup ('close = 0"), ATT and period averages are not statistically significant (e.g., ATT $p = 0.520$, Pre_avg $p = 0.323$, Post_avg $p = 0.767$). Consistent with terminology, 'close = 0" is interpreted as a negative-control subgroup (expected to show no effect), rather than a placebo based on fake treatment timing or outcomes.

Figures B.5 and B.4 shift the focus to bombing intensity. Figure B.5 ('high bomb = 0") shows no significant pre-treatment ATT and negligible changes post-treatment, suggesting little impact on municipalities with lower bombing intensity. Conversely, Figure B.4 ('high bomb = 1") shows a notable positive trend in post-treatment ATT, indicating a stronger effect in bomb-intense municipalities.

Table B.3 complements these figures by reporting ATT estimates across subgroups. Proximity to the landfill ("close = 1") and high bombing intensity ("high bomb = 1") each matter, and their combination yields a larger but imprecisely estimated ATT (0.024, $p = 0.105$), consistent with amplification by bombing intensity. Pre-treatment values are insignificant (-0.012), while post-treatment effects are marginally significant (0.023, $p = 0.121$). For municipalities close to landfills but not heavily bombed, the ATT is smaller (0.020, $p = 0.047$), with significant pre-treatment trends (-0.028, $p = 0.014$) and a modest post-treatment effect (0.020, $p = 0.012$), indicating a reduced impact absent high bombing intensity.

In conclusion, neither proximity to landfills nor bombing intensity alone explains the full heterogeneity in treatment effects. The interaction between the two—being close to a landfill and heavily bombed—reveals the strongest and most significant effects on infant mortality, as shown by the robustness checks in Table B.3.

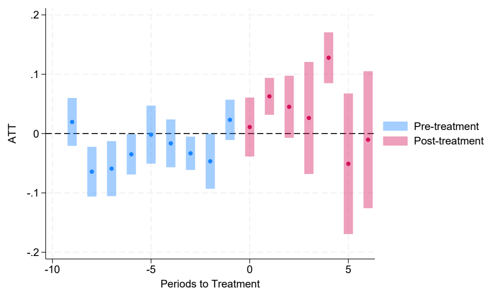


Figure B.2: close = 1

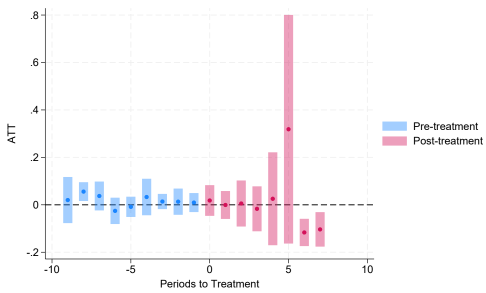


Figure B.3: close = 0

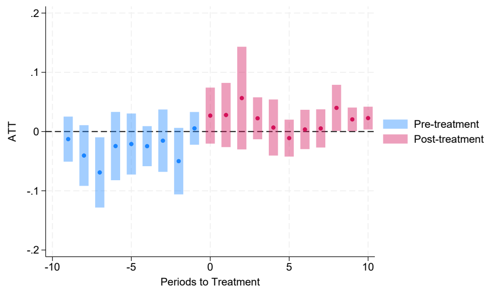


Figure B.4: high bomb = 0

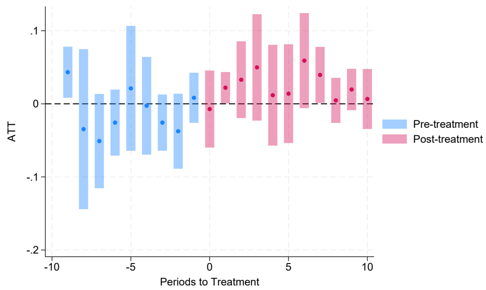


Figure B.5: high bomb = 1

Table B.3: Average Treatment Effect on Treated (ATT) and Event Study: Dynamic Effects

	Coefficient	**Std. Err.**	**z**	**P>|z|**	**[95% Conf. Interval]**	
			Close = 1			
ATT	0.0391	0.0193	2.02	0.043	0.001	0.077
Pre avg	-0.024	0.010	-2.40	0.017	-0.043	-0.004
Post avg	0.030	0.016	1.87	0.062	-0.002	0.062
			Negative control: Close = 0			
ATT	0.038	0.059	0.64	0.520	-0.077	0.152
Pre avg	0.016	0.017	0.99	0.323	-0.016	0.049
Post avg	0.016	0.055	0.30	0.767	-0.091	0.124
			Robustness: High bomb = 1 ifCclose			
ATT	0.024	0.015	1.62	0.105	-0.005	0.052
Pre avg	-0.012	0.020	-0.58	0.563	-0.051	0.028
Post avg	0.023	0.015	1.55	0.121	-0.006	0.052
			Robustness: High bomb = 0 if Close			
ATT	0.020	0.008	2.63	0.009	0.005	0.035
Pre avg	-0.028	0.011	-2.47	0.014	-0.050	-0.006
Post avg	0.020	0.008	2.51	0.012	0.004	0.036

Notes: Estimates are obtained with the staggered DiD estimator of Callaway and Sant'Anna (2021). Treatment is defined by the opening of the nearest landfill; `close` indicates residence within 6 km at birth, and `high_bomb` equals 1 for municipalities with above-median wartime bombing intensity (1999). The negative-control panel fixes close=0 (farther than 6 km). Pre_avg and Post_avg report averages of event-time coefficients in pre- and post-treatment periods, respectively. Standard errors are clustered at the municipality level. Two-sided p-values reported.

B.6 DiD Results: Municipality and Year FE

Table B.4: Difference-in-Difference Results for Municipality and Opening Year Fixed Effects without covariates

Panel A: D×D×D			
Infant Mortality	(1)	(2)	(3)
Postdump	-0.018	-0.21***	-0.019***
	(0.008)	(0.007)	(0.006)
High Bomb	-0.032***	-0.026**	-0.028***
	(0.008)	(0.010)	(0.007)
Close	-0.012	-0.012	-0.008
	(0.008)	(0.009)	(0.007)
Postdump × High Bomb	–0.022	-0.016	-0.003
	(0.013)	(0.009)	(0.016)
Postdump × Close	-0.005	0.001	-0.007
	(0.008)	(0.008)	(0.006)
High Bomb × Close	-0.003	-0.016	0.028**
	(0.011)	(0.012)	(0.012)
Postdump × High Bomb × Close	0.037***	0.044***	0.002
	(0.012)	(0.010)	(0.019)
Covariates	X	X	X
Child Birth Year FE	X	X	X
N	5373	3249	5373
N_clust	26	17	26

Panel B: D×D			
Infant Mortality	(1)	(2)	(3)
Postdump	-0.022***	-0.021***	-0.026***
	(0.004)	(0.006)	(0.003)
High bomb	-0.025***	-0.033***	-
	(0.005)	(0.008)	)
Postdump × high bomb	0.017	0.033***	0.001
	(0.011)	(0.006)	(0.008)
Municipality FE	X	X	X
Child Birth Year FE	X	X	X
N	2945	1800	1383
N clust	15	8	11

Notes: The table displays coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, robust standard errors clustered at the municipality level are reported in parentheses. *Panel A* represents the triple difference-in-difference, while the simple difference-in-difference is reported in *Panel B*. The sample in *Panel B* is obtained by subsampling children living within 6 km of a waste landfill. It also includes weights to account for children born near more than one landfill within 6 km. The table shows results for infant mortality considering (1) all the waste landfills; (2) the externally funded landfills ; (3) the waste landfill non-constructed with the aid money. All withoutcovariates, standard errors are clustered at the municipality area.

B.7 Quadruple Interaction

B.7.1 Quadruple interaction

$$
\begin{aligned}
\text{Infant mortality}_{it} = \beta_0 &+ \beta_1 \cdot \text{Postdump}_{it} + \beta_2 \cdot \text{High bomb}_i + \beta_3 \cdot \text{Close}_i + \beta_4 \cdot \text{EAR\_d}_{it} \\
&+ \beta_5 \cdot (\text{Postdump}_{it} \times \text{High bomb}_i) + \beta_6 \cdot (\text{Postdump}_{it} \times \text{Close}_i) + \\
&+ \beta_7 \cdot (\text{Postdump}_{it} \times \text{EAR\_d}_{it}) + \beta_8 \cdot (\text{High bomb}_i \times \text{Close}_i) + \\
&+ \beta_9 \cdot (\text{High bomb}_i \times \text{EAR\_d}_{it}) + \beta_{10} \cdot (\text{close}_i \times \text{EAR\_d}_{it}) + \\
&+ \beta_{11} \cdot (\text{Postdump}_{it} \times \text{High bomb}_i \times \text{Close}_i) + \\
&+ \beta_{12} \cdot (\text{Postdump}_{it} \times \text{High bomb}_i \times \text{EAR\_d}_{it}) + \\
&+ \beta_{13} \cdot (\text{Postdump}_{it} \times \text{Close}_i \times \text{EAR\_d}_{it}) + \\
&+ \beta_{14} \cdot (\text{High bomb}_i \times \text{Close}_i \times \text{EAR\_d}_{it}) + \\
&+ \beta_{15} \cdot (\text{Postdump}_{it} \times \text{High bomb}_i \times \text{Close}_i \times \text{EAR\_d}_{it}) + \\
&+ \beta_{16} X_{it} + \nu_{it} + \epsilon_{it}
\end{aligned} \tag{B.1}
$$

Table B.5: Quadruple Difference-in-Difference

Panel A: D×D×D×D						
infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.030**	-0.035*	-0.037*	-0.029	-0.033	-0.037*
	(0.015)	(0.019)	(0.018)	(0.021)	(0.023)	(0.018)
High bomb	0.022	0.008	-0.006	0.001	-0.031**	
	(0.017)	(0.010)	(0.016)	(0.012)	(0.012)	
Postdump × High bomb	-0.018	-0.005	-0.003	-0.002	-0.007	-0.003
	(0.018)	(0.020)	(0.019)	(0.022)	(0.023)	(0.019)
close	-0.004	-0.008	-0.011	-0.007	-0.003	-0.011
	(0.013)	(0.008)	(0.008)	(0.008)	(0.011)	(0.008)
Postdump × Close	0.008	0.009	0.009	0.005	0.005	0.009
	(0.014)	(0.020)	(0.018)	(0.020)	(0.023)	(0.018)
High bomb × Close	-0.006	0.012	0.016	0.012	0.005	0.016
	(0.016)	(0.011)	(0.011)	(0.010)	(0.012)	(0.011)
Postdump × High bomb × Close	0.015	0.011	0.011	0.014	0.015	0.011
	(0.018)	(0.027)	(0.025)	(0.026)	(0.029)	(0.025)
EAR_d	-0.004	-0.009	-0.005	-0.016	-0.004	
	(0.016)	(0.010)	(0.014)	(0.013)	(0.012)	
Postdump × EAR_d	0.013	0.021	0.025	0.023	0.021	0.025
	(0.019)	(0.022)	(0.021)	(0.023)	(0.025)	(0.021)
High bomb × EAR_d	-0.010	0.001		0.004	0.031	
	(0.025)	(0.018)		(0.016)	(0.022)	
Postdump × High bomb × EAR_d	0.003	-0.012	-0.015	-0.010	-0.011	-0.015
	(0.026)	(0.027)	(0.025)	(0.026)	(0.029)	(0.025)
Close6 × EAR_d	0.009	0.017	0.012	0.016	0.005	0.012
	(0.017)	(0.010)	(0.013)	(0.010)	(0.017)	(0.013)
Postdump × Close × EAR_d	-0.012	-0.016	-0.017	-0.013	-0.014	-0.017
	(0.020)	(0.023)	(0.021)	(0.023)	(0.026)	(0.021)
High bomb × Close × EAR_d	-0.029	-0.047**	-0.041*	-0.043**	-0.032	-0.041*
	(0.023)	(0.019)	(0.021)	(0.017)	(0.023)	(0.021)
Postdump × High bomb × Close × EAR_d	0.032	0.038	0.039	0.031	0.035	0.039
	(0.026)	(0.031)	(0.030)	(0.030)	(0.033)	(0.030)
Covariates		X	X	X	X	X
Dump Opening year FE			X			
Child Birth Year FE				X		
Region FE					X	
Waste landfill FE						X
N	5373	4466	4466	4466	4466	4466
N clust	26	20	20	20	20	20

Panel B: D×D						
infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.022***	-0.028***	-0.031***	-0.013	-0.031***	-0.031***
	(0.004)	(0.004)	(0.004)	(0.014)	(0.004)	(0.004)
High bomb	0.016**	0.022***	-0.024***	-0.004	-0.011	
	(0.007)	(0.005)	(0.006)	(0.010)	(0.007)	
Postdump × High bomb	-0.003	0.003	0.007	0.025	0.006	0.007
	(0.008)	(0.012)	(0.013)	(0.017)	(0.013)	(0.013)
EAR_d	0.005	0.001	-0.003	-0.020**	-0.002	
	(0.004)	(0.007)	(0.004)	(0.007)	(0.004)	
Postdump × EAR_d	0.001	0.006	0.010	0.024***	0.010	0.010
	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)
High bomb× EAR_d	-0.039***	-0.045***		-0.026**		
	(0.008)	(0.007)		(0.008)		
Postdump × High bomb × EAR_d	0.035***	0.030*	0.026*	0.013	0.026*	0.026*
	(0.010)	(0.013)	(0.014)	(0.019)	(0.014)	(0.014)
Covariates		X	X	X	X	X
Dump Opening year FE			X			
Child Birth Year FE				X		
Region FE					X	
Waste landfill FE						X
N	2945	2366	2366	2366	2366	2366
N clust	15	10	10	10	10	10

Notes: The table displays coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, followed by standard errors in parentheses. *Panel A* represents the triple difference in difference while the simple difference in difference is reported in *Panel B*. *Panel B* sample is obtained by subsampling the children living within 6 km to a waste landfill. Furthermore, it also includes weights to account for children born in places with more than one dump within 6 km. The table shows the results for infant mortality where (1) represents the triple difference in difference without covariates or fixed effects; (2) is the same as the previous one with covariates; (3) contains waste landfill opening years fixed effects; (4) contains birth year fixed effects; and (5) has regional fixed effects.

Table B.6: Predictive Margins for Infant Mortality by Bombing Intensity and Landfill Proximity

	Margin	**Std. Err.**	**P>|t|**
Panel A: High bomb × Close			
0 0	0.0286***	0.0052	0.000
0 1	0.0297**	0.0085	0.000
1 0	0.0152***	0.0070	0.000
1 1	0.0442***	0.0066	0.000
Joint contrast	df = 1, F = 5.44		p = 0.035
Panel B: Postdump × High bomb × Close			
0 0 0	0.0296***	0.0071	0.000
0 0 1	0.0374***	0.0052	0.000
0 1 0	0.0386***	0.0086	0.000
0 1 1	0.0262***	0.0070	0.000
1 0 0	0.0198**	0.0074	0.007
1 0 1	0.0223***	0.0067	0.001
1 1 0	0.0157*	0.0063	0.012
1 1 1	0.0354***	0.0061	0.000
Joint contrast	df = 1, F = 7.43		p = 0.012

Notes: Table reports predictive margins for infant mortality rates estimated from the triple-difference specification. Panel A shows margins by bombing intensity and landfill proximity; Panel B adds post-dump exposure. Reported margins correspond to subgroup means (e.g., 0 0 1 indicates postdump = 0, high bomb = 0, close = 1). Joint F-tests evaluate the statistical significance of the interaction terms. Standard errors are clustered at the municipality level. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B.8 Summary Statistics

Table B.7: Summary Statistics

	All		close		High Bomb		close & High Bomb		Not close& Low Bomb	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Infant mortality	0.029	0.167	0.028	0.166	0.032	0.177	0.030	0.171	0.024	0.154
High bomb	0.484	0.500	0.494	0.500	1.000	0.000	1.000	0.000	0.000	0.000
Postdump	0.441	0.497	0.415	0.493	0.444	0.497	0.419	0.493	0.467	0.499
Close	0.548	0.498	1.000	0.000	0.559	0.497	1.000	0.000	0.000	0.000
Rural	0.520	0.500	0.396	0.489	0.598	0.490	0.477	0.500	0.598	0.490
Albanian	0.914	0.281	0.904	0.295	0.896	0.305	0.892	0.310	0.947	0.224
Serb	0.022	0.147	0.026	0.161	0.004	0.065	0.000	0.000	0.023	0.149
Other ethnicities	0.064	0.245	0.070	0.255	0.100	0.299	0.108	0.310	0.030	0.172
No educ	0.026	0.160	0.028	0.165	0.025	0.155	0.027	0.162	0.027	0.161
Primary educ	0.051	0.221	0.046	0.208	0.057	0.232	0.054	0.225	0.055	0.229
Second educ	0.546	0.498	0.523	0.500	0.612	0.487	0.595	0.491	0.519	0.500
Up Second educ	0.263	0.440	0.280	0.449	0.226	0.418	0.241	0.428	0.273	0.446
Poor	0.216	0.411	0.207	0.406	0.244	0.430	0.238	0.426	0.202	0.402
middle class	0.195	0.397	0.179	0.383	0.218	0.413	0.191	0.393	0.183	0.386
Upper middle class	0.194	0.396	0.189	0.391	0.205	0.404	0.210	0.407	0.202	0.402
Rich	0.193	0.395	0.195	0.396	0.168	0.374	0.184	0.387	0.230	0.421
Observations	5,373		2,945		2,601		1,455		1,282	

The table provides summary statistics for the variables used in the analysis, categorized by bombing intensity and geographical location. The data is divided into four groups: all observations, clusters in high-bombing municipalities, clusters in low-bombing municipalities, and clusters within and beyond 6 km of waste landfills in high-bombing municipalities.

The variables presented are all dummy variables taking value 1 if the charachteristics is present and zero otherwise. It include infant mortality, which measures the proportion of infant deaths, and high bombing, a binary variable indicating whether the municipality experienced bombing intensity above the median. Rural indicates whether the cluster is in a rural area. Other ethnicities reflect the proportion of ethnic minorities such as ignificant minorities include Bosniaks, Turks, Askhali, Egyptian, Gorani and Roma in the municipality. No ownership indicates the proportion of households without property ownership, poor represents the proportion of households in the lower-income class, and separated ethnic groups is a binary variable indicating whether the municipality has distinct residential areas for different ethnic groups. Upper middle class captures the proportion of households in the upper-middle-income class. Lastly, observations show the total number of data points for each category.

B.9 ATT alternative exposure groups

Appendix Table B.8 reports results for alternative exposure groups that should not be jointly affected by bombing and landfill exposure. In low-bombing municipalities, children living near landfills show a positive ATT, but the pre-treatment coefficients are also significantly different from zero, suggesting violations of parallel trends. This indicates that the low-bomb group is not a valid counterfactual for the main treated units and reinforces the importance of conditioning treatment on both landfill proximity and bombing intensity.

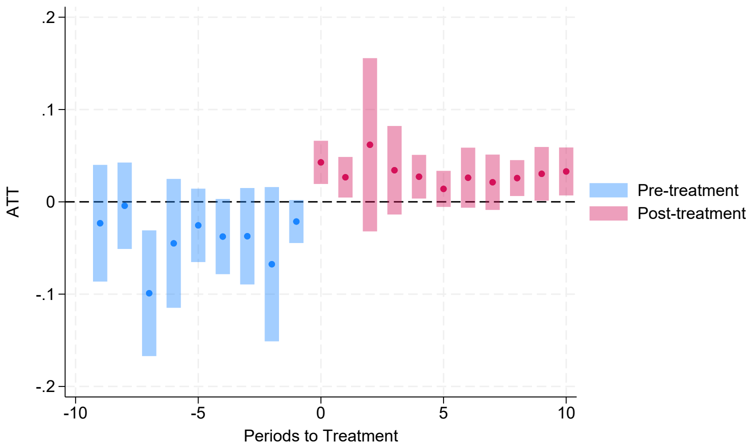
Table B.8: Average Treatment Effects on Infant Mortality: Alternative Exposure Groups

ATT	**Pre-treatment avg.**	**Post-treatment avg.**
Panel A Within 6 km, Low bombing		
0.032**	-0.040***	0.031**
(0.013)	(0.013)	(0.013)
Panel B >6 km, High bombing		
0.002	-0.024	0.002
(0.016)	(0.026)	(0.016)
Panel C >6 km, Low bombing		
-0.037	0.061**	-0.036
(0.026)	(0.027)	(0.026)

Notes: These panels report results for exposure groups that are not jointly exposed to both bombing and landfill proximity. Panel A shows effects for children near landfills in low-bombing municipalities; significant pre-trends raise concerns about validity in this group. Panels B and C report falsification checks for children residing farther than 6 km from landfills. Standard errors clustered at the municipality level are in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B.10 DD and DDD of all the different waste landfills categories

Figure B.6: Event-study estimates: Children in non-bombed municipalities, within 6 km of a landfill



Note: Estimates obtained using the Callaway and Sant'Anna (2021) staggered DiD estimator with doubly robust inverse probability weighting. Treatment is defined as landfill openings. The sample is restricted to children residing within 6 km of a landfill in municipalities with below-median wartime bombing intensity. Significant pre-treatment differences are observed, indicating violations of the parallel trends assumption. Standard errors clustered at the municipality level.

Table B.9: Difference in Difference results for the Non–European Agency for Reconstruction waste landfills

Panel A: D×D×D						
Infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.023	-0.022	-0.022	-0.029	-0.024	-0.022
	(0.015)	(0.026)	(0.027)	(0.022)	(0.026)	(0.027)
High bomb	0.024	0.013	-0.020	-0.009	-0.021	
	(0.016)	(0.014)	(0.012)	(0.020)	(0.012)	
Close_o	-0.002	0.005	0.004	-0.002	0.009	0.004
	(0.013)	(0.012)	(0.010)	(0.011)	(0.014)	(0.010)
Postdump × High bomb	-0.025	-0.016	-0.016	0.004	-0.014	-0.016
	(0.018)	(0.026)	(0.027)	(0.033)	(0.026)	(0.027)
Postdump × Close_o	0.001	-0.007	-0.004	-0.008	-0.004	-0.004
	(0.014)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
High bomb × Close_o	-0.007	0.001	0.002	0.009	-0.004	0.002
	(0.015)	(0.016)	(0.015)	(0.016)	(0.016)	(0.015)
Postdump × High bomb × Close_o	0.022	0.024	0.020	0.024	0.020	0.020
	(0.018)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
Covariates		X	X	X	X	X
Dump Opening year FE			X			
Child Birth Year FE				X		
Region FE					X	
Waste landfill FE						X
N	2564	1657	1657	1657	1657	1657
N clust	18	11	11	11	11	11
Panel B: D×D						
infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.022***	-0.030***	-0.029**	-0.033	-0.029**	-0.029**
	(0.004)	(0.007)	(0.008)	(0.021)	(0.008)	(0.008)
High bomb	0.017**	0.016**		-0.000		
	(0.007)	(0.005)		(0.020)		
Postdump × High bomb	-0.003	0.004	0.004	0.031	0.004	0.004
	(0.008)	(0.009)	(0.011)	(0.019)	(0.011)	(0.011)
Covariates		X	X	X	X	X
Dump Opening year FE			X			
Child Birth Year FE				X		
Region FE					X	
Waste landfill FE						X
N	1383	830	804	804	804	804
N clust	11	7	6	6	6	6

Notes: Coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; robust SE in parentheses. *Panel A* reports the triple-difference; *Panel B* the standard DiD on births within 6 km of a landfill (weighted for places with > 1 dump within 6 km). Columns sequentially add: covariates, landfill-opening year FE, birth-year FE, region FE, and landfill FE.

Table B.10: Difference in Difference results for the Municipal waste landfills

Panel A: D×D×D						
Infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.021	-0.006	-0.006	-0.018	-0.007	-0.006
	(0.014)	(0.016)	(0.017)	(0.015)	(0.017)	(0.017)
High bomb	0.000	-0.004	-0.036	-0.009	-0.035	
	(0.023)	(0.023)	(0.020)	(0.025)	(0.021)	
Postdump × High bomb	-0.017	-0.028	-0.029	-0.028	-0.028	-0.029
	(0.021)	(0.024)	(0.025)	(0.022)	(0.025)	(0.025)
Close_m	-0.009	-0.001	-0.002	0.001	-0.002	-0.002
	(0.010)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)
Postdump × Close_m	0.002	-0.013	-0.014	-0.017	-0.015	-0.014
	(0.014)	(0.016)	(0.017)	(0.016)	(0.017)	(0.017)
High bomb × Close_m	0.002	-0.006	-0.012	-0.004	-0.012	-0.012
	(0.023)	(0.024)	(0.018)	(0.023)	(0.018)	(0.018)
Postdump × High bomb × Close_m	-0.001	0.010	0.012	0.012	0.012	0.012
	(0.023)	(0.026)	(0.026)	(0.025)	(0.026)	(0.026)
Covariates		X	X	X	X	X
Dump Opening year FE			X			
Child Birth Year FE				X		
Region FE					X	
Waste landfill FE						X
N	2061	1930	1930	1930	1930	1930
N clust	16	14	14	14	14	14
Panel B: D×D						
Infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.019***	-0.023***	-0.023***	-0.001	-0.023***	
	(0.005)	(0.004)	(0.004)	(0.018)	(0.004)	
High bomb	0.003	-0.026		-0.009		
	(0.003)	(0.016)		(0.008)		
Postdump × High bomb	-0.018***	0.008	-0.014**	-0.017	-0.014***	
	(0.005)	(0.015)	(0.004)	(0.014)	(0.004)	
Covariates		X	X	X	X	X
Dump Opening year FE			X			
Child Birth Year FE				X		
Region FE					X	
Waste landfill FE						X
N	1119	1204	1099	1098	1099	
N clust	8	6	6	6	6	

Notes: Coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; robust SE in parentheses. *Panel A* reports the triple-difference; *Panel B* the standard DiD on births within 6 km of a landfill (weighted for places with > 1 dump within 6 km). Columns sequentially add: covariates, landfill-opening year FE, birth-year FE, region FE, and landfill FE.

Table B.11: Difference in Difference results for the Transfer waste landfills

Panel A: D×D×D						
Infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.043	-0.058	-0.058	-0.023	-0.051	-0.058
	(0.061)	(0.066)	(0.075)	(0.071)	(0.071)	(0.075)
High bomb	-0.025	-0.044		-0.068		
	(0.042)	(0.038)		(0.036)		
Postdump × High bomb	0.006	0.024	0.023	0.037	0.016	0.023
	(0.062)	(0.065)	(0.075)	(0.068)	(0.071)	(0.075)
Close_t	-0.042	-0.063	-0.075*	-0.071	-0.054	-0.075*
	(0.040)	(0.036)	(0.037)	(0.038)	(0.047)	(0.037)
Postdump × Close_t	0.022	0.041	0.041	0.046	0.033	0.041
	(0.061)	(0.066)	(0.078)	(0.068)	(0.073)	(0.078)
High bomb × Close_t	0.042	0.068	0.090*	0.089*	0.068	0.090*
	(0.041)	(0.039)	(0.040)	(0.043)	(0.052)	(0.040)
Postdump × High bomb × Close_t	-0.010	-0.029	-0.028	-0.040	-0.020	-0.028
	(0.063)	(0.067)	(0.078)	(0.070)	(0.074)	(0.078)
Covariates		X	X	X	X	X
Dump Opening year FE			X			
Child Birth Year FE				X		
Region FE					X	
Waste landfill FE						X
N	1548	1548	1548	1548	1548	1548
N clust	7	7	7	7	7	7
Panel B: D×D						
infant mortality	(1)	(2)	(3)	(4)	(5)	(6)
Postdump	-0.021	-0.047**	-0.030	0.027**	-0.031	-0.030
	(.)	(0.011)	(0.012)	(0.007)	(0.012)	(0.012)
High bomb	0.017*	0.015		0.016		
	(0.007)	(0.004)		(0.011)		
Postdump × High bomb	-0.004	0.016	-0.002	-0.001	-0.001	-0.002
	(0.008)	(0.011)	(0.013)	(0.005)	(0.012)	(0.013)
Covariates		X	X	X	X	X
Dump Opening year FE			X			
Child Birth Year FE				X		
Region FE					X	
Waste landfill FE						X
N	823	432	414	414	414	414
N clust	4	3	3	3	3	3

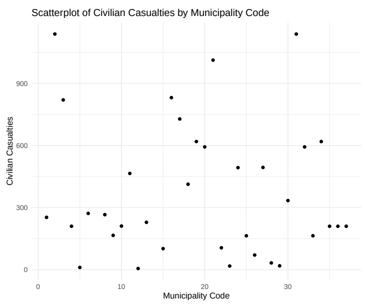
Notes: Coefficients with $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; robust SE in parentheses. *Panel A* reports the triple-difference; *Panel B* the standard DiD on births within 6 km of a landfill, with weights for places having > 1 dump within 6 km. Columns add (in order): covariates, landfill-opening year FE, birth-year FE, region FE, and landfill FE.

Appendix C

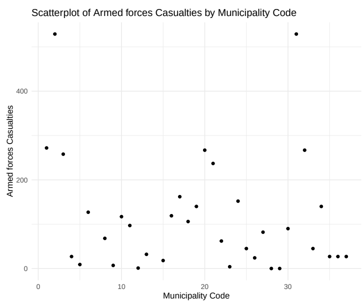
Appendix to Chapter 3

APPENDIX

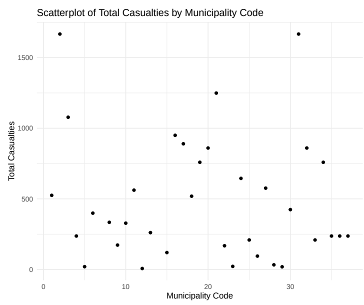
C.1 Casualties Scatterplot



(a) Civilian casualties



(b) Armed forces casualties



(c) Total killings

Figure C.1: Bin scatterplots of different types of casualties by municipality bin.

C.2 Coverage and Ranges of Key Variables

Table C.1: Coverage and Ranges of Key Variables

Variable	Observations = NA	Observations ≠ NA	Unique Values	Min	Max
Corruption for					
Customs	4,251	18,106	2	0	1
Court	4,040	18,317	2	0	1
Education	3,460	18,897	2	0	1
Local Gov	3,631	18,726	2	0	1
Central Gov	4,448	17,909	2	0	1
Political Parties	3,919	18,438	2	0	1
Eulex	4,045	18,312	2	0	1
Police	2,880	19,477	2	0	1
Hospital	12,683	9,674	2	0	1
International Org	5,595	16,762	2	0	1
Pak	1,692	20,665	2	0	1
Demographic Variables					
Casualties High	1,692	20,665	2	0	1
Civilian High	1,692	20,665	2	0	1
Armed Force High	1,692	20,665	2	0	1
Female	100	22,257	78	18	100
Age	122	22,235	7	1	99
Marital Status	1,022	21,335	220	1	15,000
Income	431	21,926	2	0	1
House Ownership	153	22,204	2	0	1
Urban (1=Yes, 0=No)	153	22,204	2	0	1
Unemployed	153	22,204	2	0	1
Public Sector Employment	153	22,204	2	0	1
Private Sector Employment	153	22,204	2	0	1
Student Employment	153	22,204	2	0	1
Municipality code	546	21,811	37	1	37

The table presents summary statistics for key variables used in the analysis, including perceptions of corruption across multiple sectors (e.g., customs, courts, education, police), socio-economic characteristics (e.g., gender, age, income, employment status), and violence exposure indicators (e.g., casualties, civilian casualties, armed force presence). The summary statistics cover the total sample and subsets of individuals categorized by their exposure to violence (e.g., high casualties, high civilian casualties, high armed force presence). The "Obs" columns represent the number of observations with and without missing data for each variable.

C.3 Collinearity Diagnostics for Exposure Variables

Civilian and Armed force casualties are positively correlated across municipalities. The simple municipality-level correlation is $\rho = 0.76$ and, within the estimation samples used in the joint regressions, $\rho = 0.75$. To assess whether this correlation undermines inference when both exposures enter together, we compute standard collinearity diagnostics.

Variance inflation factors (VIFs) for the two exposure variables in the joint models are approximately 2.25 for both civilian and Armed force casualties, well below common thresholds of concern (5 or 10). The condition number of the regressor matrix is $\kappa \approx 4.24$, far below levels typically associated with problematic collinearity (above 30). These diagnostics indicate that while the two measures are substantively correlated, they do not generate harmful multicollinearity.

We conclude that the differential signs in the joint specifications—positive for civilian casualties and negative or null for Armed force casualties—are not artifacts of collinearity but reflect substantively meaningful differences in the long-term legacy of violence.

Table C.2: Collinearity Diagnostics (Joint Specifications)

	Corr(civilian, armed)	VIF (civilian / armed)	Condition number κ
Local outcome sample	0.745	2.249 / 2.249	4.24
Central outcome sample	0.745	2.246 / 2.246	4.24
Municipality-level (simple)	0.755	—	—

Notes: Corr(civilian, armed) reports the Pearson correlation between civilian- and Armed force casualties per 1990 municipal population. VIFs are from OLS joint specifications with the listed outcome (no fixed effects; VIF pertains to the regressor matrix). κ is the condition number of the centered regressor matrix excluding the intercept. Values rounded to three decimals for correlations and to two decimals for κ.

C.4 Corruption Correlation matrix

Table C.3: Correlation Matrix of Corruption Perception Variables

	Customs	Court	Edu	LocalGov	CentralGov	Parties	Police	Hospital	CivSoc	PTK	PAK	Banks	KEDS	TAK
Customs	1.000													
Court	0.608	1.000												
Education	0.347	0.318	1.000											
LocalGov	0.375	0.399	0.446	1.000										
CentralGov	0.470	0.465	0.411	0.511	1.000									
Parties	0.467	0.481	0.343	0.413	0.464	1.000								
Police	0.336	0.326	0.401	0.401	0.375	0.293	1.000							
Hospital	0.363	0.379	0.426	0.406	0.388	0.380	0.364	1.000						
CivSoc	0.288	0.287	0.367	0.357	0.328	0.297	0.387	0.316	1.000					
PTK	0.406	0.400	0.359	0.352	0.405	0.376	0.350	0.336	0.393	1.000				
PAK	0.444	0.441	0.328	0.366	0.436	0.405	0.335	0.389	0.374	0.499	1.000			
Banks	0.338	0.323	0.378	0.353	0.365	0.318	0.366	0.329	0.440	0.449	0.406	1.000		
KEDS	0.415	0.399	0.357	0.380	0.418	0.364	0.365	0.370	0.364	0.493	0.493	0.489	1.000	
TAK	0.399	0.400	0.355	0.378	0.399	0.352	0.369	0.365	0.381	0.477	0.498	0.441	0.526	1.000

This table reports pairwise Pearson correlation coefficients among variables capturing perceptions of corruption across various public institutions, based on a sample of 15,361 respondents. All values range from 0 (no correlation) to 1 (perfect positive correlation). The highest correlations are generally observed between perceptions of central political institutions, such as central government, political parties, and customs, and those of closely associated public services like the tax authority (TAK) and telecommunications providers (PTK, PAK). Perceptions of corruption in education and civil society exhibit lower correlations with the rest, indicating some differentiation in how citizens assess integrity across institutional domains.

Table C.4: Loadings from PCA and Factor Analysis (First Component/Factor)

Variable	**PCA Loadings**			**Factor Analysis Loadings**		
	Local	**Central**	**Overall**	**Local**	**Central**	**Overall**
Corruption: Education	0.683		0.595	0.610		0.550
Corruption: Local Government	0.676		0.645	0.602		0.605
Corruption: Police	0.627		0.566	0.545		0.522
Corruption: Hospital	0.633		0.592	0.552		0.549
Corruption: Civil Society	0.666		0.614	0.593		0.572
Corruption: Banks	0.694		0.631	0.631		0.593
Corruption: KEDS	0.674		0.681	0.604		0.652
Corruption: Customs		0.734	0.635		0.690	0.600
Corruption: Court		0.733	0.634		0.691	0.599
Corruption: Central Government		0.717	0.679		0.651	0.646
Corruption: Political Parties		0.670	0.614		0.599	0.576
Corruption: International Org.		0.591	0.651		0.501	0.610
Corruption: PAK		0.721	0.677		0.652	0.648
Corruption: TAK		0.678	0.676		0.600	0.646

Note: This table reports loadings for the first principal component (PCA) and the first factor (FA) from separate analyses on corruption perception variables grouped into local, central, and overall sets. Blank cells indicate variables not included in the respective subset. PCA and FA capture related but conceptually distinct dimensions of latent corruption perception. The PCA loadings are from unrotated components, and FA loadings are from maximum likelihood estimation without rotation.

C.5 Conditional placebos at the municipality level

Table C.5: Conditional Placebo Tests — Municipality-Level Exposure (Continuous Shares)

Outcome Variable	**Permuted Exposure (other held fixed)**	**True Coefficient**	p**-value**
Civilian share permuted (armed force share fixed)			
Corruption (Local Avg)	Civilian casualties	1.832	0.499
Corruption (Central Avg)	Civilian casualties	4.385	0.223
Corruption (All Avg)	Civilian casualties	3.339	0.254
Armed force share permuted (civilian share fixed)			
Corruption (Local Avg)	Armed force casualties	-7.842	0.444
Corruption (Central Avg)	Armed force casualties	-5.650	0.583
Corruption (All Avg)	Armed force casualties	-7.857	0.418

Notes: Exposures are measured as casualties per 1990 municipal population. Coefficients are interpreted per one–percentage–point increase in the exposure (i.e., a 0.01 change in the regressor). Each row reports the coefficient from the original joint regression with both continuous exposures and individual covariates and the permutation p-value from 10,000 replications that randomly reassign the *permuted* exposure while keeping the *other* exposure fixed. Standard errors in the original model are clustered by municipality; municipality or year fixed effects are not included in these cross-sectional placebo designs.

C.6 Ethnicity effect

Table C.6: Heterogeneity by Ethnicity: Interaction of Conflict Exposure and Albanian Ethnicity

	Local Corruption			**Central Corruption**		
	(1)	(2)	(3)	(4)	(5)	(6)
Civilian Casualties						
Civilian Casualties	2.441**	4.405***	4.081***	9.377***	10.480***	8.685***
	(0.932)	(0.994)	(0.992)	(0.844)	(0.902)	(0.894)
Albanian	0.034***	0.034***	0.036***	0.081***	0.088***	0.084***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Civilian Casualties × Albanian	-3.823***	-6.104***	-6.467***	-9.992***	-11.100***	-9.800***
	(1.158)	(1.215)	(1.216)	(1.046)	(1.100)	(1.093)
Armed Force Casualties						
Armed Casualties	3.343	7.903**	5.157·	20.946***	22.120***	14.320***
	(2.526)	(2.691)	(2.693)	(2.280)	(2.437)	(2.420)
Albanian	0.031***	0.028***	0.025***	0.072***	0.076***	0.068***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Armed Casualties × Albanian	-10.668***	-15.120***	-13.460***	-26.099***	-27.340***	-20.650***
	(3.076)	(3.238)	(3.230)	(2.774)	(2.929)	(2.901)
Observations	25,802	24,289	24,289	25,514	24,007	24,007
Covariates	No	Yes	Yes	No	Yes	Yes
Year FE	No	No	Yes	No	No	Yes
Human Rights Violations during and after war						
HRV During	-0.153***	-0.140***	-0.142***	-0.153***	-0.147***	-0.135***
	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
HRV After	0.169***	0.158***	0.104***	0.137***	0.134***	0.087***
	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)
Albanian	-0.048***	-0.057***	-0.065***	-0.003	-0.010	-0.011
	(0.013)	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)
HRV During × Albanian	0.099***	0.095***	0.081***	0.130***	0.131***	0.103***
	(0.018)	(0.018)	(0.018)	(0.017)	(0.017)	(0.017)
HRV After × Albanian	-0.125***	-0.121***	-0.091***	-0.160***	-0.161***	-0.121***
	(0.024)	(0.024)	(0.023)	(0.022)	(0.023)	(0.022)
Observations	8,504	5,801	5,749	8,515	5,814	5,762
Covariates	No	Yes	Yes	No	Yes	Yes
Year FE	No	No	Yes	No	No	Yes
Municipality FE	No	No	Yes	No	No	Yes

Notes: Each column reports coefficients from separate OLS regressions. The dependent variable is perceived corruption for local institutions (cols. 1–3) and central institutions (cols. 4–6). Columns 1 and 4 include no covariates and no fixed effects. Columns 2 and 5 add individual covariates (age, female, urban residence, unemployment, pensioner status, years of schooling, homeownership, and income). Columns 3 and 6 further include year fixed effects in addition to the covariates. Regressions interact conflict exposure (civilian or armed casualties divided by the 1990 municipal population) and human rights violations (HRV) during and after the conflict with an indicator for Albanian ethnicity, where applicable. Standard errors are robust.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

C.7 Local, Central and All institutions

Table C.7: Local, Central and All institutions regression results

	Simple Average		PCA		Factor Analysis	
	(1)	(4)	(2)	(5)	(3)	(6)
Panel A: Local Institutions						
Civilian casualties	2.416^{***}	1.770^{**}	2.341^{**}	1.766^{*}	2.254^{**}	1.676
	(0.813)	(0.826)	(1.016)	(1.031)	(1.014)	(1.029)
Armed force casualties	-8.836^{***}	-7.707^{***}	-5.652^{**}	-5.080^{*}	-5.619^{**}	-5.037^{*}
	(2.211)	(2.241)	(2.779)	(2.816)	(2.774)	(2.811)
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	27,491	27,055	18,264	18,034	18,264	18,034
Panel B: Central Institutions						
Civilian casualties	4.342^{***}	4.361^{***}	4.393^{***}	4.533^{***}	4.425^{***}	4.574^{***}
	(0.770)	(0.782)	(1.014)	(1.025)	(1.014)	(1.024)
Armed force casualties	-5.676^{***}	-5.599^{***}	-3.519	-3.864	-3.593	-3.935
	(2.101)	(2.142)	(2.774)	(2.823)	(2.773)	(2.822)
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	27,472	27,034	17,154	16,953	17,154	16,953
Panel C: All Institutions						
Civilian casualties	3.589^{***}	3.289^{***}	4.172^{***}	4.079^{***}	4.330^{***}	4.254^{***}
	(0.724)	(0.736)	(1.025)	(1.037)	(1.026)	(1.037)
Armed force casualties	-8.335^{***}	-7.750^{***}	-4.526	-4.641	-4.434	-4.579
	(2.003)	(2.038)	(2.856)	(2.894)	(2.855)	(2.893)
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	27,762	27,318	15,413	15,248	15,413	15,248

Notes: OLS coefficients with municipality–clustered standard errors in parentheses. SIgnificance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Columns (1),(2),(3) exclude individual covariates; columns (4),(5),(6) include age, gender, urban residence, years of schooling, employment status (unemployed, public, private, student), and homeownership. No municipality or year fixed effects are included because the exposure variables are time–invariant at the municipality level. Violence exposures enter jointly in all columns and are defined as casualties per 1990 municipal population. PCA and FA indices are linearly rescaled to match the simple average's mean and standard deviation and re-oriented so higher values denote more perceived corruption. Sample sizes vary due to missing covariates and index coverage.

Table C.8: Regression Results – Impact of Post-Conflict Human Rights Violations on Corruption Measures

	Simple Average		PCA		Factor Analysis	
	(1)	(4)	(2)	(5)	(3)	(6)
Panel A: Local Institutions						
HRV during conflict	-0.096^{***}	-0.090^{***}	-0.108^{***}	-0.104^{***}	0.108^{***}	-0.104^{***}
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
HRV after conflict	0.042^{***}	0.039^{***}	0.037^{***}	0.035^{***}	0.036^{***}	0.034^{***}
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,488	8,464	7,190	7,170	7,190	7,170
Panel B: Central Institutions						
HRV during conflict	-0.058^{***}	-0.055^{***}	-0.063^{***}	-0.059^{***}	-0.063^{***}	-0.059^{***}
	(0.008)	(0.009)	(0.009)	(0.030)	(0.009)	(0.009)
HRV after conflict	-0.005	-0.005	-0.013	-0.014	-0.012	-0.013
	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,514	8,489	7,089	7,070	7,089	7,070
Panel C: All Institutions						
HRV during conflict	-0.075^{***}	-0.071^{***}	-0.078^{***}	-0.074^{***}	-0.076^{***}	-0.072^{***}
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)
HRV after conflict	0.017^{*}	0.016	0.004	0.003	0.004	0.003
	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,531	8,506	6,721	6,703	6,721	6,703

Note: Standard errors are reported in parentheses below the coefficient estimates. Statistical significance is denoted as follows: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Sample sizes vary across models due to missing data in covariates and differ by institutional domain.
This table reports the estimated coefficients from linear regressions assessing the impact of post-conflict human rights violations (HRV) on various measures of perceived corruption. The dependent variables are constructed from multiple corruption perception indicators using three alternative methods: PCA and FA indices have been linearly rescaled to match the Simple Average's mean and standard deviation and re-oriented so higher values indicate more perceived corruption.
Panel A (Local Institutions): HRV after the conflict is positively and significantly associated with higher corruption perception in local institutions across all methods and specifications. The magnitude and statistical significance are robust to inclusion of covariates.
Panel B (Central Institutions): The relationship is weaker and less consistent. Only PCA and FA without covariates show marginally significant positive effects.
Panel C (All Institutions): The overall relationship remains positive and significant in PCA and FA models, suggesting that post-conflict HRV has contributed to persistent institutional distrust, particularly in local-level services.
Covariates include age, gender, urban residence, education, employment status, and home ownership.

Table C.9: Regression Results - Impact of Wartime Human Rights Violations on Corruption Measures with Municipality FE and Interaction

	Simple Average		PCA		Factor Analysis	
Panel A: Local Institutions						
HRV during conflict	-0.106***	-0.100***	-0.121***	-0.155***	-0.120***	-0.115***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
HRV after conflict	-0.024	-0.023	-0.041*	-0.040	-0.041*	-0.040
	(0.022)	(0.022)	(0.025)	(0.025)	(0.025)	(0.026)
HRV during × after	0.086***	0.081***	0.100***	0.097***	0.099***	0.095***
	(0.025)	(0.025)	(0.028)	(0.028)	(0.028)	(0.028)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,488	8,464	7,190	7,170	7,190	7,170
Panel B: Central Institutions						
HRV during conflict	-0.063***	-0.060***	-0.070***	-0.066***	-0.070***	-0.066***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
HRV after conflict	-0.038*	-0.034	-0.061***	-0.059**	-0.061***	-0.059**
	(0.021)	(0.021)	(0.023)	(0.023)	(0.023)	(0.023)
HRV during × after	0.044*	0.038	0.062**	0.058**	0.063**	0.059**
	(0.024)	(0.024)	(0.026)	(0.026)	(0.026)	(0.026)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,514	8,489	7,089	7,070	7,089	7,070
Panel C: All Institutions						
HRV during conflict	-0.083***	-0.078***	-0.087***	-0.082***	-0.085***	-0.081***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
HRV after conflict	-0.033*	-0.031	-0.056**	-0.054**	-0.055**	-0.053**
	(0.020)	(0.020)	(0.023)	(0.023)	(0.023)	(0.023)
HRV during × after	0.065***	0.060***	0.077***	0.073***	0.075***	0.071***
	(0.022)	(0.023)	(0.026)	(0.026)	(0.026)	(0.026)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes	No	Yes
Num. Obs.	8,531	8,506	6,721	6,703	6,721	6,703

Note: Standard errors in parentheses. Statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.PCA and FA indices have been linearly rescaled to match the Simple Average's mean and standard deviation and re-oriented so higher values indicate more perceived corruption (see Footnote 8). Regressions estimate the association between wartime HRV exposure, postwar HRV exposure, and their interaction on perceptions of corruption in local (Panel A), central (Panel B), and all institutions (Panel C). Corruption indices are constructed using Simple Average, PCA, and Factor Analysis. All models include municipality and year fixed effects. Covariates in columns (2), (4), and (6) include age, gender, education, employment status, urban residence, and homeownership. Sample sizes vary due to item nonresponse and index-specific coverage.

C.8 Age interaction

Table C.10: Regression Results - Impact of Civilian Casualties on Corruption Measures with Age Interactions

	Local Avg	Local Avg	Local PCA	Local PCA	Local FA	Local FA
Panel A: Local Institutions						
Civilian Casualties	-1.187	-1.465	-1.426	-1.653	-1.551	-1.783
	(0.993)	(1.003)	(1.183)	(1.193)	(1.183)	(1.193)
Age 30–44	-0.005	0.002	-0.010	-0.001	-0.009	-0.001
	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)	(0.010)
Age 45–59	-0.039***	-0.026***	-0.032***	-0.019*	-0.031***	-0.018
	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)	(0.011)
Age 60+	-0.043***	-0.017	-0.055***	-0.028**	-0.054***	-0.027**
	(0.010)	(0.011)	(0.013)	(0.014)	(0.013)	(0.014)
Civilian Casualties × Age 30–44	-0.047	-0.146	1.588	1.339	1.682	1.436
	(1.388)	(1.401)	(1.686)	(1.702)	(1.686)	(1.702)
Civilian Casualties × Age 45–59	3.798**	3.574**	3.934**	3.638*	3.954**	3.674**
	(1.516)	(1.532)	(1.846)	(1.862)	(1.846)	(1.862)
Civilian Casualties × Age 60+	1.568	1.843	5.907**	5.946**	6.011***	6.060***
	(1.813)	(1.829)	(2.313)	(2.333)	(2.314)	(2.333)
Num.Obs.	27,491	27,077	18,264	18,043	18,264	18,043
Covariates	No	Yes	No	Yes	No	Yes
Panel B: Central Institutions						
Civilian Casualties	2.953***	3.118***	3.704***	3.722***	3.794***	3.822***
	(0.897)	(0.908)	(1.104)	(1.111)	(1.104)	(1.111)
Age 30–44	0.010	0.019**	0.006	0.015	0.007	0.016*
	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Age 45–59	-0.012	0.001	-0.011	0.001	-0.011	0.002
	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)	(0.010)
Age 60+	-0.020**	-0.004	-0.029**	-0.011	-0.028**	-0.009
	(0.009)	(0.010)	(0.012)	(0.013)	(0.012)	(0.013)
Civilian Casualties × Age 30–44	-1.925	-2.059	-2.215	-2.110	-2.330	-2.227
	(1.254)	(1.268)	(1.568)	(1.580)	(1.568)	(1.580)
Civilian Casualties × Age 45–59	0.337	-0.022	-0.174	-0.290	-0.294	-0.408
	(1.373)	(1.390)	(1.724)	(1.738)	(1.724)	(1.738)
Civilian Casualties × Age 60+	2.306	2.369	3.650*	4.023*	3.528	3.904*
	(1.637)	(1.655)	(2.185)	(2.203)	(2.185)	(2.202)
Num.Obs.	27,472	27,057	17,154	16,965	17,154	16,965
Covariates	No	Yes	No	Yes	No	
Panel C: All Institutions						
Civilian Casualties	0.808	0.817	2.051*	1.949*	2.311**	2.216**
	(0.869)	(0.880)	(1.118)	(1.125)	(1.118)	(1.125)
Age 30–44	0.001	0.011	-0.003	0.005	-0.002	0.005
	(0.007)	(0.007)	(0.009)	(0.010)	(0.009)	(0.010)
Age 45–59	-0.025***	-0.010	-0.023**	-0.009	-0.022**	-0.008
	(0.008)	(0.008)	(0.010)	(0.011)	(0.010)	(0.011)
Age 60+	-0.031***	-0.007	-0.044***	-0.021	-0.043***	-0.021
	(0.009)	(0.010)	(0.012)	(0.013)	(0.012)	(0.013)
Civilian Casualties × Age 30–44	-0.885	-1.083	-0.546	-0.441	-0.606	-0.500
	(1.214)	(1.228)	(1.597)	(1.610)	(1.597)	(1.610)
Civilian Casualties × Age 45–59	1.906	1.535	1.511	1.283	1.320	1.086
	(1.327)	(1.343)	(1.755)	(1.767)	(1.755)	(1.767)
Civilian Casualties × Age 60+	1.780	1.862	5.308**	5.574**	5.212**	5.483**
	(1.582)	(1.599)	(2.215)	(2.234)	(2.215)	(2.233)
Num.Obs.	27,762	27,341	15,413	15,255	15,413	15,255
Covariates	No	Yes	No	Yes	No	Yes

Note: Standard errors are in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Controls include gender, urban/rural, education, employment sector, student status, and home ownership.PCA and FA indices have been linearly rescaled to match the Simple Average's mean and standard deviation and re-oriented so higher values indicate more perceived corruption (see Footnote 8).

Table C.11: Regression Results – Impact of Armed Forces Casualties on Corruption Measures with Age Interactions

	Local Avg	Local Avg	Local PCA	Local PCA	Local FA	Local FA
Panel A: Local Institutions						
Armed Forces Casualties	-4.233	-4.393*	-2.556	-2.713	-2.699	-2.852
	(2.584)	(2.610)	(3.122)	(3.147)	(3.122)	(3.147)
Age 30–44	0.000	0.007	0.002	0.010	0.003	0.011
	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Age 45–59	-0.030***	-0.016**	-0.023**	-0.010	-0.022**	-0.009
	(0.008)	(0.008)	(0.009)	(0.010)	(0.009)	(0.010)
Age 60+	-0.038***	-0.011	-0.040***	-0.013	-0.039***	-0.012
	(0.009)	(0.010)	(0.011)	(0.012)	(0.011)	(0.012)
Armed Forces Casualties × Age 30–44	-4.138	-4.137	-3.702	-4.017	-3.664	-3.986
	(3.644)	(3.680)	(4.440)	(4.479)	(4.440)	(4.479)
Armed Forces Casualties × Age 45–59	5.222	4.673	5.797	5.240	5.761	5.201
	(3.949)	(3.982)	(4.841)	(4.877)	(4.841)	(4.877)
Armed Forces Casualties × Age 60+	1.182	2.040	8.416	8.156	8.454	8.206
	(4.581)	(4.627)	(5.704)	(5.770)	(5.704)	(5.770)
Num.Obs.	27,491	27,077	18,264	18,043	18,264	18,043
Covariates	No	Yes	No	Yes	No	Yes
Panel B: Central Institutions						
Armed Forces Casualties	3.649	3.916*	7.491***	7.769***	7.667***	7.942***
	(2.336)	(2.363)	(2.897)	(2.919)	(2.897)	(2.919)
Age 30–44	0.009	0.018***	0.008	0.017**	0.008	0.018**
	(0.006)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Age 45–59	-0.011	0.001	-0.007	0.005	-0.007	0.006
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)
Age 60+	-0.017**	-0.001	-0.020**	0.000	-0.019*	0.001
	(0.008)	(0.009)	(0.010)	(0.011)	(0.010)	(0.011)
Armed Forces Casualties × Age 30–44	-5.621*	-5.940*	-8.008*	-8.196**	-8.212**	-8.378**
	(3.295)	(3.336)	(4.133)	(4.164)	(4.133)	(4.164)
Armed Forces Casualties × Age 45–59	0.319	-0.025	-3.466	-4.003	-3.818	-4.316
	(3.568)	(3.606)	(4.496)	(4.529)	(4.496)	(4.529)
Armed Forces Casualties × Age 60+	5.274	5.733	5.368	5.383	5.070	5.106
	(4.136)	(4.186)	(5.397)	(5.456)	(5.397)	(5.456)
Num.Obs.	27,472	27,057	17,154	16,965	17,154	16,965
Covariates	No	Yes	No	Yes	No	Yes
Panel C: All Institutions						
Armed Forces Casualties	-0.737	-0.580	5.004*	5.157*	5.540*	5.700*
	(2.260)	(2.286)	(2.954)	(2.971)	(2.954)	(2.970)
Age 30–44	0.004	0.013**	0.005	0.013	0.005	0.013
	(0.006)	(0.007)	(0.008)	(0.009)	(0.008)	(0.009)
Age 45–59	-0.020***	-0.006	-0.015*	-0.001	-0.014	0.000
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)
Age 60+	-0.027***	-0.003	-0.030***	-0.006	-0.030***	-0.006
	(0.008)	(0.009)	(0.010)	(0.012)	(0.010)	(0.012)
Armed Forces Casualties × Age 30–44	-5.009	-5.343*	-7.399*	-7.549*	-7.433*	-7.581*
	(3.185)	(3.224)	(4.232)	(4.261)	(4.232)	(4.261)
Armed Forces Casualties × Age 45–59	2.412	1.908	-0.857	-1.700	-1.295	-2.154
	(3.447)	(3.483)	(4.587)	(4.612)	(4.587)	(4.611)
Armed Forces Casualties × Age 60+	2.461	2.986	7.306	6.908	7.140	6.720
	(3.999)	(4.047)	(5.478)	(5.537)	(5.478)	(5.537)
Num.Obs.	27,762	27,341	15,413	15,255	15,413	15,255
Covariates	No	Yes	No	Yes	No	Yes

Note: Standard errors in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Controls include gender, urban/rural, education (years), employment sector dummies (public/private), student status, and home ownership. PCA and FA indices are linearly rescaled to match the Simple Average's mean and standard deviation and re-oriented so higher values indicate more perceived corruption. Casualty variables are per 1,000 of the 1990 municipal population. Baseline age group is 18–29.

C.9 Sensitivity test: Omitted variable bias

Table C.12: Sensitivity to Omitted Variables: Oster (2019) Bounds

Outcome	**Exposure**	$\hat{\beta}_0$	$\hat{\beta}_1$	R_0	R_1	$\delta \to 0$	$\beta^*(\delta = 1)$
Panel A. Local corruption perceptions							
Local Avg	Armed-force casualties	-8.335	-7.750	0.000928	0.005000	0.2046	-7.9656
Local Avg	Civilian casualties	3.589	3.289	0.000928	0.005000	0.2472	3.3994
Local PCA	Armed-force casualties	-5.650	-5.080	0.000310	0.005880	0.3558	-5.264
Local PCA	Civilian casualties	2.340	1.770	0.000310	0.005880	1.0286	1.95
Local FA	Armed-force casualties	-5.620	-5.040	0.000294	0.005730	0.3651	-5.22
Local FA	Civilian casualties	2.250	1.680	0.000294	0.005730	1.0889	1.86
Panel B. Central corruption perceptions							
Central Avg	Armed-force casualties	-5.676	-5.599	0.001385	0.003565	0.02801	-5.637
Central Avg	Civilian casualties	4.342	4.361	0.001385	0.003565	-0.00875	4.351
Central PCA	Armed-force casualties	-3.520	-3.860	0.001768	0.006460	-0.2164	-3.72
Central PCA	Civilian casualties	4.390	4.530	0.001768	0.006460	-0.07491	4.48
Central FA	Armed-force casualties	-3.590	-3.930	0.001784	0.006680	-0.2125	-3.79
Central FA	Civilian casualties	4.420	4.570	0.001784	0.006680	-0.07981	4.51
Panel C. Overall corruption perceptions							
All Avg	Armed-force casualties	-8.335	-7.750	0.000928	0.005000	0.2046	-7.9656
All Avg	Civilian casualties	3.589	3.289	0.000928	0.005000	0.2472	3.3994
All PCA	Armed-force casualties	-4.530	-4.640	0.001477	0.006240	-0.06288	-4.60
All PCA	Civilian casualties	4.170	4.080	0.001477	0.006240	0.05762	4.12
All FA	Armed-force casualties	-4.430	-4.580	0.001639	0.006460	-0.07897	-4.52
All FA	Civilian casualties	4.330	4.250	0.001639	0.006460	0.04468	4.28

Notes: This table reports sensitivity statistics following Oster (2019). $\hat{\beta}_0$ and R_0 come from the restricted specification (baseline controls); $\hat{\beta}_1$ and R_1 come from the fully controlled specification. $\delta \to 0$ is the value of δ (relative selection on unobservables vs. observables) required for the coefficient to equal zero. $\beta^*(\delta = 1)$ is the implied bias-adjusted coefficient under equal selection on observables and unobservables. The maximum attainable R^2 is set to $R_{\max} = \min\{1, 1.3 \times R_1\}$.