Seeing Is Solving: Video Evidence Collection and Shooting Clearance Rates in New York City

Chin-Huai Shih
University of Pennsylvania

Charles E. Loeffler
University of Pennsylvania

Abstract

Over the last decade, video evidence has become a crucial part of many criminal investigations, supplementing human witnesses and forensic evidence. Using ground-truth measures of video camera density and video collection, this study shows that this growth in video usage was driven primarily by changes in investigative practices rather than by increased CCTV availability. We then estimate the causal effect of video evidence collection on clearance using a kernel k-nearest neighbors algorithm to identify spatiotemporally proximate cases with and without video evidence. Results indicate that the collection of video evidence increased clearance rates for shooting cases by between 33-44 percentage points (about 40 on average) and by 11.9 percentage points for each doubling of videos used. We also find that the technology-aided investigative practices play an important role in explaining the city's heightened clearance shooting clearance post 2012.

1. Introduction

Before the early 2010s, it was uncommon to see video evidence used in criminal investigations in the United States. Even the most serious criminal cases relied heavily on eyewitness testimony, and, to a much lesser extent, forensic evidence (Anderson et al., 2021; Baskin & Sommers, 2012; Horvath & Meesig, 1996; Peterson et al., 2013). Since then, criminal investigations increasingly rely on a range of video surveillance technologies, including closed-circuit television (CCTV), body-worn cameras (BWCs), license plate readers (LPRs), doorbell cameras, and dashcams (Prince et al., 2021). By supplying direct observation of people and places (Barao & Mastroianni, 2024), these publicly and privately installed devices help police and others reconstruct events from scattered snippets of streetscapes, building entrances, and hallways (Brookman & Jones, 2022; Dowling et al., 2019).

In several recent high-profile gun homicides, video evidence played a crucial role in identifying and locating suspects, often aided by eyewitness accounts following the release of CCTV footage by the police. For example, during the search for the suspect involved in the Brown University shooting in 2025, investigators canvassed the campus and nearby neighborhoods for surveillance cameras to gather images of the suspect (Arsenault et al., 2025). Beyond high-profile cases such as this, local police departments routinely use video evidence to solve cases ranging from theft to murder.

Given the recent nature of this change, it can be challenging to situate contemporary police camera footage practices within the existing scholarly literature on cameras and crime. Much of this large empirical literature, dating back decades, has focused on the deterrent effects of camera installations on reported crimes in their immediate vicinity with only a handful of recent studies exploring possible clearance effects of camera installations (Gerell, 2021; Jung & Wheeler, 2023; Koper & Lum, 2019b; Piza et al. 2014; Piza, Connealy, et al., 2025). This scholarly focus on the probability of crime within view of cameras, and, to a lesser extent, the probability of crime clearance for crimes still committed within view of cameras, likely reflects the legacy of a historically sparse deployment of cameras within many communities, which tended to produce little usable video evidence.

However, cameras are now much more commonplace, and the quality of the images they produce is substantially better than in the past. Local, state, and federal law enforcement agencies report exponential growth in the collection of video footage in criminal cases over the past decade (Blake, 2023). In cities with high camera density (Sheng et al., 2021), the question is no longer whether cameras exist, but whether more and better video improves investigations (Cook & Berglund, 2021). Investigators must also decide whether, and how extensively, to canvass for

video. These decisions determine whether a case is documented with footage or left to eyewitness statements and other forms of forensic and digital evidence.

This sea-change in availability and collection of video footage may explain why studies of the clearance effects of cameras have, to date, generally produced modest, estimated effects of camera coverage on crime clearance. Most studies estimate the intent-to-treat (ITT) effect of whether a camera is present, rather than whether video was collected and used. These two quantities diverge whenever investigators do not routinely canvass for footage. A camera that goes unchecked contributes nothing to a case, so its presence tells us little about video's true investigatory value. The modest effects in the literature potentially reflect low, or at least inconsistent collection rates, rather than a weak effect of video evidence itself.Besides the availability-usage gap, studies of public sector-deployed cameras often compare covered areas to areas without public sector cameras, while the latter may be extensively covered by private sector cameras, thereby blurring the treatment-control contrast. Many studies also measure camera contribution as a binary outcome, grouping cases with dozens of video clips with cases where little footage was retrieved. Both choices yield estimates that are potentially overconservative and harder to interpret.

Using a dataset of more than 18,000 geocoded outdoor shooting incidents drawn from one of the most surveyed American cities—New York (Sheng et al., 2021)—we estimate the effect of video evidence on case clearance with an instrumental variable method. The instruments are the share of nearby shootings, in space and time, that used any video, and the average number of videos they used. Both are built with a k-nearest neighbors (kNN) algorithm tuned by leave-one-out cross-validation. Since the video usage by the neighbors is not decided by the focal case's initial evidence condition, the instruments compare shootings that are otherwise similar in attributes but differ in the chance and extent of video being used, independent of investigators' discretion. We show that

the instruments are as-if random at the case level and are unlikely to affect clearance through spatiotemporal factors other than video use. We interpret this estimator as a local average treatment effect of video usage in a video-rich environment, which is much closer to the policy estimate of interest than the ITT.

Our findings indicate that video footage is more frequently and extensively collected in more difficult-to-solve cases (see also Koper & Lum, 2019b), leading to markedly larger estimated effects once case-based selection is excluded by the instruments. The presence of any video evidence increases the probability of clearance by arrest by approximately 40 percentage points. Moreover, every doubling of the number of cited videos increases the probability of clearance by about 11.9 percentage points, revealing an exponential decay of marginal return. Furthermore, increased video use throughout NYC has contributed to sustained improvement in shooting case clearance rates since 2013.

2. Prior Literature on CCTV and Crime Clearance

The literature on video surveillance and crime falls into two streams: studies of deterrence and studies of investigation. The larger and more established stream examines CCTV's deterrent effect on crime occurrence. It assumes that cameras' presence heightens potential offenders' perception of an arrest's certainty and celerity with the identification information the footage captures (Nagin, 2013; Piza et al., 2014). Installing more cameras, therefore, has long been proposed as a situational crime prevention strategy to reduce the attractiveness of offending in its surrounding physical environment (Clarke, 1995). These studies have consistently reported that cameras have a modest deterrent effect on crimes in public spaces, particularly property offenses (Welsh & Farrington, 2008, 2009), with larger effects when cameras are actively monitored (Piza et al., 2019). Although

scholars have long noted CCTV's potential as a case-solving tool (Brown, 1995; Ditton & Short, 1999),[1] only recently has there been a sustained focus on the relationship between CCTV and crime clearance (see also Ashby, 2017; Jung & Wheeler, 2023; Piza, Connealy, et al., 2025; Robin et al., 2021; Thomas, 2025). This smaller stream on camera and clearance has nonetheless produced mixed results.

In a study of newly installed public-sector security cameras in Dallas, Texas, Jung and Wheeler (2023) used varying distances to camera installations to form a pre- and post-comparison. They found a modest 2.45 percentage-point increase in clearance rates within a 500-foot buffer, an effect largely driven by theft and concentrated near the installation sites. Piza et al. (2025) found that CCTV did not improve clearance rates for fatal or non-fatal shootings within a 500-foot radius in Chicago, using entropy matching to compare treated and untreated cases. Finally, Thomas (2025) observed a 61% increase in crime clearance within the 283.5-foot viewshed of 76 newly installed cameras in Fayetteville, North Carolina, also through entropy matching. This effect was primarily due to a 160% increase in the clearance of disorder crimes.

A conservative summary of this nascent literature would be that CCTV cameras appear to improve the clearance of less serious crimes, such as public disorder (Robin et al., 2021; Thomas, 2025) and theft (Jung & Wheeler, 2023), rather than more serious index crimes, and within an extremely narrow area around the installed cameras. However, a commonly acknowledged limitation shared by these place-based evaluations of CCTV's clearance effects is the lack of data

[1] For instance, Brown (1995) attributed sustained crime reduction in Newcastle to a marked increase in arrest rates for burglary, criminal damage, vehicle theft, and drunken offenses following CCTV installation. Similarly, Ditton and Short (1999) report that Glasgow Police used CCTV in the retrospective investigation of serious offenses, such as murder, rather than crime prevention, and posited that enhancing investigative capabilities may eventually be the "real use" of open-street CCTV in major cities' centers (p. 217).

on video usage (Jung & Wheeler, 2023; Piza, Connealy, et al., 2025; Robin et al., 2021; Thomas, 2025). If police rarely search the newly installed cameras, then the null effects derived from data generated by these installations are better understood as estimates of non-utilization (MacQueen & Bradford, 2017).[2]

A second set of recent studies, using administrative case-level data rather than spatial proximity information, has reported substantial differences in clearance rates depending on the presence of video evidence. In an analysis of 253 variables across 9,655 police reports, Olphin et al. (2021) identified video evidence as the fourth of 40 significant predictors of case clearance. They found that cases with evidential video footage had a 73.5% increase in clearance rate, whereas the absence of such footage was associated with a 44.7% decrease.[3] Morgan and Dowling (2019) utilize nearest-neighbor matching to show that when video footage was requested by and provided to the police in the New South Wales rail network, clearance rates rose by 4.5 percentage points.

A few studies go beyond this by exploring how changes in organizational processes determine whether the collection of video evidence is institutionalized as part of the investigation. The quasi-random assignment of cases to *Polizia* and *Carabinieri* in Milan, Italy, illustrates that the routine collection of surveillance camera footage by Polizia's predictive policing team to generate suspect descriptions and link related incidents results in an 8- to 13-percentage-point higher clearance rate.

[2] Such an availability-usage gap is quite visible in parts of the research literature, such as work exploring the impact of large expansions in municipal camera networks (Piza et al., 2014). For a similar argument, see also Koper et al. (2014).

[3] "CCTV presence" and "no CCTV available" were coded as two different variables in the study. According to the variable inclusion criteria set by the authors, their correlation is lower than 0.9. A possible explanation for the absence of perfect collinearity is that "CCTV presence" combines coding from the Police system and free-text coding of investigation reports, while "no CCTV available" was only extracted from free-text coding.

Similarly, Pizarro et al. (2020) evaluate the tactical change that involved a heightened reliance on video canvassing and advanced review equipment within the Rochester Police Department's (NY) homicide unit, finding that it led to a 335% increase in clearance rates.[4] Overall, these incident-based studies produce strong positive associations between video evidence and case clearance across burglary, robbery, aggravated assaults, and homicides, all of which are serious crimes that are shown by place-based studies to benefit the least from the newly installed public surveillance cameras.

Still, the majority of the findings in these studies are statistical associations rather than causally identified relationships (for exceptions, Mastrobuoni, 2020; Morgan & Dowling, 2019), largely due to police investigator discretion in deciding how much and which types of evidence to collect in any case (Coupe, 2016; Fallik, 2018). The volume of resources deployed in a case directly determines the quantity and quality of evidence collected, which, in turn, influences the probability of a successful clearance (see Braga et al., 2019; Braga & Dusseault, 2018). And the investigative value of CCTV requires a series of intensive and resource-demanding steps to realize its potential (Brookman et al., 2019; Brookman & Jones, 2022), including canvassing cameras positioned at or near the scene (Pizarro et al., 2020), distinguishing pertinent footage from the irrelevant among continuous video feeds, and timely retrieval from often non-interoperable digital devices (Cook & Berglund, 2021).

Compounding this problem, it is not entirely clear whether investigative effort is negatively or positively associated with case solvability. Typically, serious (Roberts & Roberts, 2016) or high-

[4] Similarly, Barao and Mastroianni (2024) find that the most valuable Real-Time Crime Center activity for the Hartford Police Department (CT) in solving homicides and shootings is their use of video evidence. Having videos capturing the incident increases the clearance rate by 618%, surpassing other resources such as DNA analysis, latent print identification, shell-casing examination, and gunshot-detection alerts.

profile (Aziani & Persurich, 2023; Davies, 2007; Mastrorocco & Ornaghi, 2025) crimes command greater attention and police resource input and produce higher clearance rates.[5] On the other hand, Eck's (1983) "triage thesis" suggests that the relationship between investigative input and clearance is not strictly linear (see also Fallik, 2018).[6] Investigators may prioritize their limited resources and time to cases that are less likely to be solved without their additional input, and devote a minimum amount of time to cases that they consider to be "self-solving" (Coupe, 2016). Ashby (2017) finds that crimes solved within three days were significantly less likely to involve CCTV footage, even though, when the footage was available, it was 312% more likely to be considered useful by the investigator than in cases solved over a longer time period. As a result, a case's intrinsic solvability may confound the estimated association between CCTV usage and clearance and the investigator's judgment, if video evidence is primarily sought and used for inherently difficult cases (see a similar account for LPR in Koper & Lum, 2019b).

A final challenge in estimation arises from the changing landscape of public and private surveillance cameras in many communities. In a world with few cameras and even less footage, treating cameras as present or absent was a realistic representation of the surveillance landscape. However, with cameras now on many streets, they no longer must be pointed at the scene of the crime to be useful. Scholars often use the term "network" to describe the high density, clustering,

[5] Recent studies provide evidence that fatal shootings have higher clearance rates than nonfatal cases because the organization allocates more resources, such as a dedicated unit with a lower caseload and a higher priority for forensic evidence testing, to the former (Cook et al., 2019; Cook & Vila, 2023). Investigators gather significantly more evidence and consult more sources in fatal cases than in nonfatal cases, with video evidence being used 575% more often in Boston and 216% more often in Durham across these otherwise very similar case types (Cook et al., 2019; Cook & Vila, 2023). Therefore, the apparent association between using video evidence and case clearance may be *positively* confounded by unobserved fluctuations in available investigative resources due to discretionary allocation and investigators' caseloads (LoFaso, 2020; Roberts & Roberts, 2016; Vaughn, 2020).

[6] Fallik (2018) observed that in the Houston Police Department (TX), 28% of index crimes with known suspects received no detective attention, compared to 19.5% of cases with unknown suspects. This triaging practice is widely adopted by police agencies (Prince et al., 2021), particularly among those with lower clearance rates (Lum et al., 2018), where resource constraints may necessitate more rigorous resource management and case screening.

and potential connectivity of CCTV units across a city or region (Barao & Mastroianni, 2024; Ferrandino, 2021; Koper & Lum, 2019b; Mastrobuoni, 2020; Morgan & Dowling, 2019; Robin et al., 2021; Thomas, 2025), suggesting that such attributes offer synergistic advantages that isolated or sparsely distributed units cannot provide.

Despite the wide acceptance of this theoretical conceptualization, the current empirical literature predominantly relies on *binary* indicators to capture estimated effects or associations, assessing only whether at least one CCTV unit is present in the vicinity of a crime or at least some footage is recovered. As surveillance scholars note (Brayne, 2020b; Norris & Armstrong, 2020), the power of surveillance technologies in policing does not lie in a single device or data point, but in multiple digital sensors that generate observations from different times, angles, and locations that enable investigators to construct a coherent narrative about a suspect's activities, motivations, and intentions before and after the crime incident (Adams & Ferryman, 2015; Brookman & Jones, 2022). However, if the CCTV density around the incident is low, such intensive tracing becomes not only costly but also impractical.

Together, these three estimation challenges suggest that estimates in the existing literature may provide a lower bound on the investigative effects of CCTV rather than capturing the full range of CCTV's investigative effects. In order to recover these additional downstream investigative effects of video evidence, rather than camera or footage presence, it will be necessary to model video availability, canvassing and collection decisions, as well as the quantity of footage used in any given case. The next section describes how we address these challenges.

3. Research Design

3.1 Data Source and Restrictions

The data for this study were extracted from the New York Police Department's (NYPD) case management system. They contain details on all shootings investigated in New York City (NYC) from 2006 to April 2023, along with their case status as of April 2023 (N = 21,651). We restricted our scope to this population of cases for two reasons. First, these offenses exhibit a sizable clearance gap, with most shootings going unsolved (Roberts & Roberts, 2016; Vaughn, 2020). Second, by studying only one type of crime, we ensure that incidents under analysis receive comparable levels of organizational priority. Two additional restrictions were applied to filter the original dataset. Cases that were closed without an arrest were excluded from the analysis (N=238).[7] Additionally, indoor shootings were excluded from the analysis (N=3,047), given that video evidence for indoor cases plays a somewhat different role in shooting investigations from those occurring outside.[8] These restrictions resulted in a sample of 18,397 outdoor shootings.

Table 1 presents descriptive statistics for the 18,397 outdoor shooting incidents in the analytic sample. The outcome variable, $Arrest_i$, indicates whether an arrest was made in connection with the case at the time of data retrieval; 41.4% of cases resulted in an arrest. The two primary treatment variables measure video evidence use in the investigation: $AnyVideo_i$ is a binary

[7] According to the Federal Investigation Bureau's definition, these are cases cleared by "exceptional means" when suspects were identified but cannot be arrested due to factors beyond the agency's control, such as the suspect's death or extradition denial. However, a recent investigative report suggests significant uncertainty and discretionary leeway regarding when and in which cases police cited exceptional clearance to close a case (see Yeung et al., 2018).

[8] Video footage on indoor shootings is naturally paired with information about who could usually access the area and therefore the offender-victim relation[8], which is helpful in investigations but often harder to ferret out for outdoor shootings (Jarvis et al., 2017). This additional layer of information suggests that it is not appropriate to group video usage into outdoor and indoor shootings when analyzing its benefits for crime investigations. Also, unlike outward-facing cameras, indoor-facing cameras provide almost no external values for other outdoor shootings, limiting their relevance for our estimation strategy.

indicator equal to 1 if any video evidence was collected and attached to the investigative file (37.5% of cases), while $VideoCount_i$ records the number of individual videos attached (mean = 2.63, ranging from 0 to 141 in the most intensively documented cases).

Among case characteristics, most shootings were motivated by disputes (35%) or fell in the "other" category (37.3%), with gang- and drug-motivated cases comprising 17.4% and 10.3% respectively. The vast majority of shootings were non-fatal (83.4%). Incidents occurred predominantly on streets (75.6%), with a notable share near public housing (17.8%). The kNN-based instrumental variables ($kNNVid_{\%i}$ and $kNNVid_{\#i}$) and the neighborhood clearance control ($ClusterArrest_{\%i}$) are described in Section 3.3.

Table 1: Descriptive Statistics.

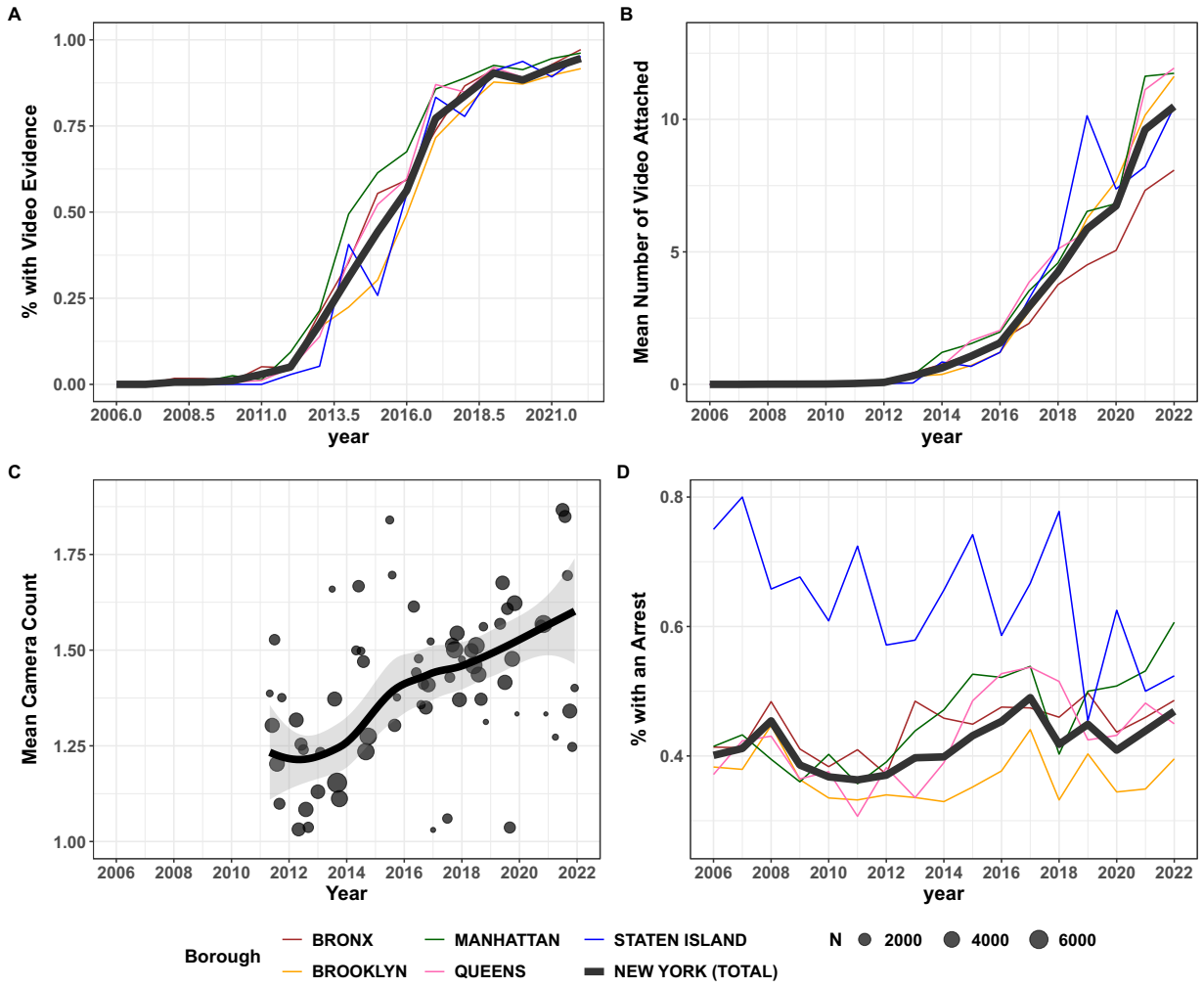
Variable	N	Mean	SD	Min	Pctl. 25	Pctl. 75	Max
$Arrest_i$	18397	0.414	0.493	0	0	1	1
$AnyVideo_i$	18397	0.375	0.484	0	0	1	1
$VideoCount_i$	18397	2.63	6.55	0	0	2	141
Primary Motive	18397						
... Dispute	6437	35%					
... Drugs	1890	10.30%					
... Gang	3205	17.40%					
... Other	6865	37.30%					
Fatal	18397						
... No	15346	83.40%					
... Yes	3051	16.60%					
Location	18397						
... Commercial	259	1.40%					
... Dwelling	268	1.50%					
... Housing	3271	17.80%					
... Other	98	0.50%					
... Parking Lot	60	0.30%					
... Playground	235	1.30%					

Variable	N	Mean	SD	Min	Pctl. 25	Pctl. 75	Max
... Street	13903	75.60%					
... Taxi/Livery	30	0.20%					
... Transit	32	0.20%					
... Vehicle	241	1.30%					
$kNNVid_{\%i}$	18397	0.373	0.398	0	0	0.845	1
$kNNVid_{\#i}$	18397	2.62	3.92	0	0	4.63	26
$ClusterArrest_{\%i}$	18397	0.412	0.125	0.0178	0.324	0.49	0.926
$IHS(VideoCount_i)$	18397	0.809	1.2	0	0	1.44	5.64
$IHS(kNNVid_{\%i})$	18397	0.341	0.357	0	0	0.768	0.881
$IHS(kNNVid_{\#i})$	18397	1.01	1.19	0	0	2.24	3.95

Figure 1 plots four different trends in NYC from 2006 to 2022: video evidence usage, video counts, CCTV units on the street, and the shooting arrest rate. Over this period, the video usage rate in shooting investigations rose dramatically—from zero to nearly full adoption (Panel A of Figure 1). The growth trajectory follows the S-curve commonly observed in technology diffusion (Geroski, 2000). Before 2012, the yearly usage rate grew slowly, from 0% to 5%. Between 2012 and 2019, it surged from 5% to over 90%, a compound annual growth rate of 43.79%. The growth then tapered off, reaching 94.4% by 2022. The yearly average video count in each case follows a similar trajectory for the first 14 years. Both experienced a growth boost around 2012 and 2013, but diverged as the video count continued to grow after 2019 (Panel B of Figure 1). During the same time, citywide shooting arrest rates in NYC rose around 2012 and have remained above 40% since 2016 (Panel D of Figure 1).[9]

[9] The growth is noteworthy and potentially underrepresented by the trend in the figure. Because more newer cases are subject to right-censoring and that shorter investigation time, the arreat rates are mechanically lower.

Figure 1: Trends of Video Evidence Usage, CCTV, and Arrest Rate



Using randomly sampled Google Street View images from areas with criminal shootings, Panel C of Figure 1 plots the mean monthly camera detection and the smoothed trend, weighted by the number of locations with available images. The trend shows that camera density growth is steady but much milder than the geometric growth observed in the NYPD video usage rate and count.[10] The acceleration in density around 2014 also does not align with the time when the boost

[10] Although Street View images are available before 2011, we restrict the graph to post-2011 data for two reasons. First, prior to 2011, coverage of Brooklyn was sparse, and Brooklyn constitutes a large share of the gunshot dataset. This restriction ensures that the data is comparable with other trends based on the gunshot dataset. Second, the resolution of 2007 Street View images (the first large collection) is too low for reliable camera detection, even by

in video evidence usage began. This mismatch implies that the change in the number of CCTV units is unlikely to be the sole reason for the rapid adoption of video evidence in shooting investigations.

3.2 Baseline Models

Our study targets two estimands: the treatment effect of using any video evidence versus none (*binary treatment*), and the dose-response effect of the number of videos used (*continuous treatment*). The baseline ordinary least squares (OLS) specifications are:

$$Arrest_i = \delta_1 AnyVideo_i + X_i'\beta + \lambda_{ym} + \gamma_{wh} + \pi_p + \epsilon_{i.} \quad (3\text{-}1)$$

$$Arrest_i = \delta_2 VideoCount_i + X_i'\beta + \lambda_{ym} + \gamma_{wh} + \pi_p + \epsilon_{i.} \quad (3\text{-}2)$$

Without any confounders, δ_1 captures the treatment effect of whether *any* video evidence is used, and δ_2 the dosage effect of adding *additional* video evidence on the probability of shooting clearance. $\boldsymbol{X_i}'$ is a row vector of case attributes, all categorical. $Fatal$ indicates whether the shooting was fatal. $Primary\ Motive$ of the shooting includes disputes, gangs, drugs, and others. $Location$ classifies the location of the outdoor shooting as commercial, dwelling, housing, parking lot, playground, street, taxi/livery, transit, vehicle, or other.

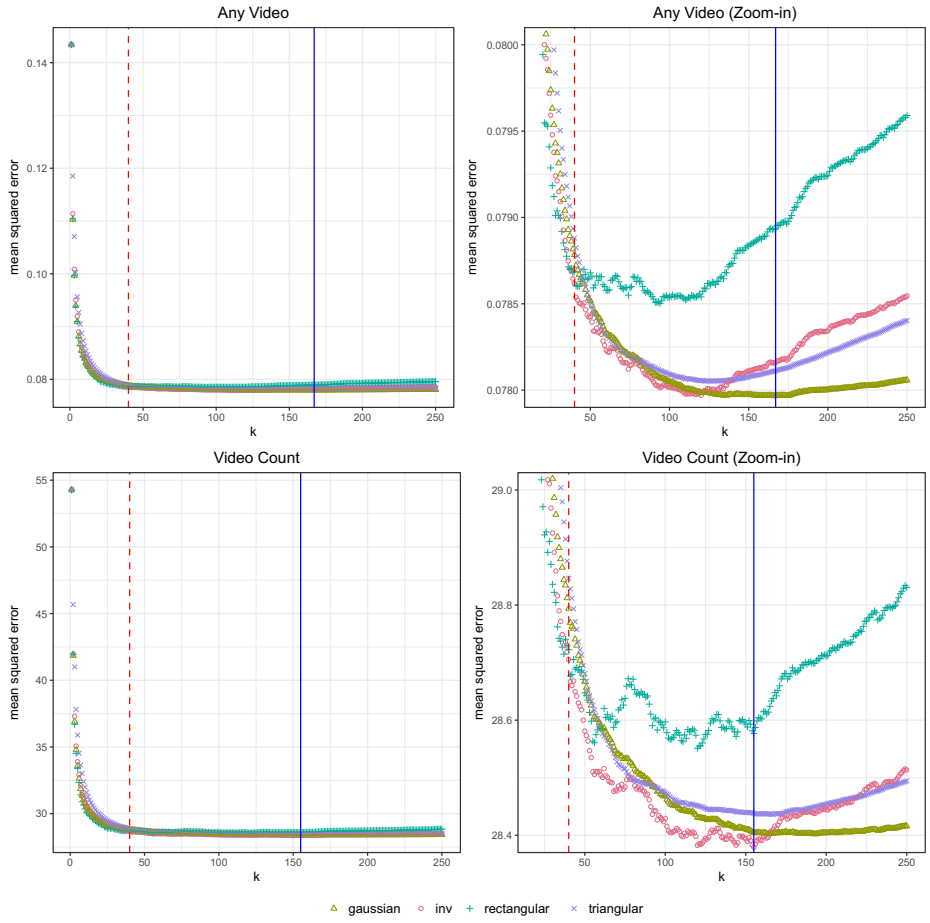
We include three sets of fixed effects to absorb unobservable temporal and neighborhood effects on shooting clearance. Year-by-month fixed effects (λ_{ym}, 208 parameters) account for variation in clearance risk over time or citywide social/policy changes. Weekday-by-hour fixed

human inspection. Including 2007 would produce a sharp jump in the average number of detected cameras per location, from 0.25 in 2007 (likely underestimated) to 1.25 in 2011. We do not attribute this to an actual increase in CCTV density; and even if it were real, the timing would not align with the sudden acceleration in video evidence usage observed only after 2012. Therefore, data points before 2011 are redacted to avoid misleading the readers. The non-redacted version is in Appendix.

effects (γ_{wh}, 168 parameters) account for the potential shift arrangement effect (see Vaughn, 2020) and the weekly pattern of criminal activities by hour in NYC. Precinct fixed effects (π_p, 76 parameters) control for time-invariant characteristics of the local neighborhood and police station in charge.

However, even with the covariate adjustments, δ_1 and δ_2 are unlikely to be unbiased. As our literature review shows, the decision to collect video in crime investigation is highly discretionary, often influenced by a department's priority, individual investigators' judgement on the cost-benefit of canvassing and analyzing videos, and the other evidence already at hand (e.g., footprints, DNA, or fingerprints), all with the potential to affect case clearance. A simple comparison of cases with and without collected video therefore mixes the effect of video with the selection process that prompted video collection in the first place. An alternative modeling approach is therefore needed.

Figure 2: K-Search for Video Usage Prediction.



Note: The search was up to k=250. The two graphs in the left panels and those on the right are from the same *k* search but different zoom-in levels. In particular, the y-axes show very different scales. The legend shows different kernel schemes.

3.3 Identification Strategy: IV Synthesis with Kernel-Weighted k-Nearest Neighbor Algorithm

To address these selection concerns, we utilize an instrumental variables (IV) approach that harnesses variation that only affects case clearance through its effect on video use thereby eliminating individual officer's decision on canvassing and collection of video evidence from the

estimand. The instrument we exploit is how commonly and extensively videos were collected in the investigations of *other* shootings that are close in space and time.

This spatiotemporal instrument is conceptually similar to the approach used by Weisburd et al. (2016), who estimated the effect of Stop-Question-Frisk on crime. Because police make more stops where crime is already high, stops on a street segment reflect that segment's existing crime risk. They instead used the level of stops in other precincts of the same borough, which captures stop practices without being driven by the segment's own risk. We apply the same idea to investigations: video use in neighboring cases reflects local camera coverage and canvassing practice, not the existing evidence in the focal shooting.

A simple example shows how the instrument sets aside detectives' case-specific choices. Suppose a solved shooting has 9 videos collected, while nearby shootings average 5. The extra 4 videos may signal a weak case, as police often canvass harder when witnesses or physical evidence are lacking, and the same deficiencies reduce the likelihood of an arrest. Now suppose the shooting instead had no video at all. This may mean the case was already strong, with DNA evidence or an arrest at the scene, so detectives saw no need to look for footage. In both versions, the raw video count reflects the strength of other evidence that we cannot fully observe, and a simple comparison would understate the value of video, because video is used most where cases are hardest to solve. The instrument ignores these departures. It uses only the level of video use expected from where and when the shooting occurred, about 5 videos here, which does not depend on the evidence in that particular case.

A valid instrument requires three conditions, which we examine in Section 3.4 and below. First, nearby video use must predict video use in the focal case. We expect this for two reasons. For one, the same cameras that supplied footage for nearby shootings can also capture the focal shooting;

and for two, where police canvass for video routinely, they do so in most shootings, not only the difficult or salient ones. Second, nearby video use must be unrelated to the focal case's solvability. This holds if the exact time and place of a shooting within a precinct is *as good as random* with respect to how solvable it is. For investigators, this is plausible because the time and location of a shooting are given before it is known and cannot be manipulated *ex post*. Third, nearby video use should affect the focal case's clearance only through the focal case's own video use. This might be where our instruments are the most vulnerable, as the locations where video evidence is collected more often may be the communities that are more cooperative with police investigation or have more competent detective squads, which are directly associated with higher clearance rates regardless of the video use. We address this by controlling for the investigative environment and by drawing on a large enough set of neighboring cases that no single related shooting drives the measure.

Our instrument is constructed at a higher resolution than precincts using a k-nearest neighbors (kNN) algorithm that identifies the k shootings spatially and temporally closest to case i and computes a distance-weighted average of video usage among those neighbors.[11] In essence, the instruments are weighted counts of nearby cases. $kNNVid_{\%i}$ is the weighted proportion of the k nearest neighboring shootings in which any video evidence was used, and $kNNVid_{\#i}$ is the weighted average number of videos used in those neighboring cases. Weights may decay with distance according to kernel functions, giving closer cases more influence on the predicted value for case i.

[11] Combining machine learning with instrumental variable methods to extract identifying variation from individually weak exogenous predictors (i.e., one neighbor) and synthesize them into a strong instrument has gained growing traction among applied researchers (see Wu et al., 2025).

To select the optimal number of neighbors k and weights, we search over values from 1 to 250 using leave-one-out cross-validation with four different kernel functions (inverse, Gaussian, rectangular, triangular), measuring prediction error for each candidate (Figure 2). The neighbors are identified via Euclidean distance (d) across the x-coordinate, y-coordinate, and incident time. Prediction error declines sharply with small values of k and bottoms at around $k = 167$ for $AnyVideo_i$ and $k = 155$ for $VideoCount_i$. We adopt $k^* = 40$ for both instruments as a conservative choice. At this value, the inverse-distance kernel ($\frac{1}{d}$) achieves the best predictive accuracy among the kernel options tested, and the instrument draws on neighbors that are meaningfully local rather than spanning large distances.[12] Accordingly, the IVs for $AnyVideo_i$ and $VideoCount_i$ can be synthesized as:

$$kNNVid_{\%i} = \frac{\sum_{j \in \mathcal{N}_{k^*i}} w_{ji}^* \cdot AnyVideo_j}{\sum_{j \in \mathcal{N}_{k^*i}} w_{ji}^*},$$

$$kNNVid_{\#i} = \frac{\sum_{j \in \mathcal{N}_{k^*i}} w_{ji}^* \cdot VideoCount_j}{\sum_{j \in \mathcal{N}_{k^*i}} w_{ji}^*},$$

where $\mathcal{N}_{k^*i}$ is a set of 40 neighbors (j) for case i; w_{ji}^* is the inverse-distance weights.

With the instruments, we rely on two-stage least squares (2SLS) to estimate the binary and continuous treatment effects of video use. In the first stage, each case's actual video evidence usage ($AnyVideo_i$ or $VideoCount_i$) is regressed on the kNN instrument and controls. This yields a predicted value, $\widehat{AnyVideo_i}$ or $\widehat{VideoCount_i}$. The predicted values replace the original endogenous treatment variables in the second stage to obtain the 2SLS coefficients (θ_1, θ_2). They

[12] The median weighed geographical distance to the 40 nearest neighbors has already been 2275.49 feet (and with the median time distance of 138 days). These are more three times more distant than the previous research conclusion that a surveillance camera only increases clearance within a 500-feet buffer (Jung & Wheeler, 2023).

are interpreted as local average treatment effects (LATE): the effect of video evidence use among cases whose video collection was driven by local norms of video collection in nearby cases, purged of the case-specific characteristics that drive selective video use. The directions of $\delta_1 - \theta_1$ and $\delta_2 - \theta_2$, therefore, indicate the overall direction of selection biases.

Granted, omitted variable bias could arise if spatiotemporal clustering of video evidence use is linked to other factors that directly affect clearance, violating the exclusion restriction assumption for a valid instrument (Angrist et al., 1996). For example, precincts with highly effective detective teams may both collect more video and solve more cases, regardless of the video itself. We mitigate this concern by including the average clearance rate of the 40 neighbors ($ClusterArrest_{\%i}$) as a control that summarizes the shared investigative environment that could otherwise confound the instrument. In practice, it is constructed as

$$\frac{\sum_{j \in \mathcal{N}_{k^*i}} w_{ji}^* \cdot Arrest_j}{\sum_{j \in \mathcal{N}_{k^*i}} w_{ji}^*}.$$

In Section 4.3, we address this concern even more directly through an alternative fixed-effects specification that nullifies the effect of this term.

Accordingly, the 2SLS regressions of our study can be written as

$$\begin{cases} AnyVideo_i = \phi_1 kNNVid_{\%i} + \eta_1\, ClusterArrest_{\%i} + X_i'\beta + \tau_i + \epsilon_i \ \text{(1st stage)} \\ Arrest_i = \theta_1 \widehat{AnyVideo}_\iota + \eta_2\, ClusterArrest_{\%i} + X_i'\beta + \tau_i + \epsilon_i \text{(2nd stage)} \end{cases}; \quad (3\text{-}3)$$

$$\begin{cases} VideoCount_i = \phi_2 kNNVid_{\#i} + \eta_1\, ClusterArrest_{\%i} + X_i'\beta + \tau_i + \epsilon_i \ \text{(1st stage)} \\ Arrest_i = \theta_2 \widehat{VideoCount}_\iota + \eta_2\, ClusterArrest_{\%i} + X_i'\beta + \tau_i + \epsilon_i \text{(2nd stage)} \end{cases}. \quad (3\text{-}4)$$

τ_i is the shorthand for the fixed effects (λ_{ym}, γ_{hw}, and π_p) in Equations 3-1 and 3-2. Since the models will be fitted with 2SLS, for binary outcome variables like $Arrest_i$ and $AnyVideo_i$, the

regressions are linear probability models (LPM). The interpretation of the coefficients is then the change in percentage points (pp) from moving the regressors by one unit.

3.4 Examining the Synthesized IVs

Here, we examine the properties of the instruments synthesized by the algorithm, including their spatiotemporal distribution, relevance to the treatment variables, and their balance across the case attributes in our data. Figure 3 plots the time trend of $kNNVid_{\%i}$ and $kNNVid_{\#i}$ against $AnyVideo_i$ and $VideoCount_i$. $kNNVid_{\%i}$ and $kNNVid_{\#i}$ represent smoother, less noisy copies that closely trace the growth of $AnyVideo_i$ and $VideoCount_i$. Figure 4 and Figure 5 plot $kNNVid_{\%i}$ and $kNNVid_{\#i}$ on the map by year, showing that they are not merely smoothed citywide time series of the original variables. They also capture the micro-level spatial clustering of video usage that is unevenly distributed across NYC.

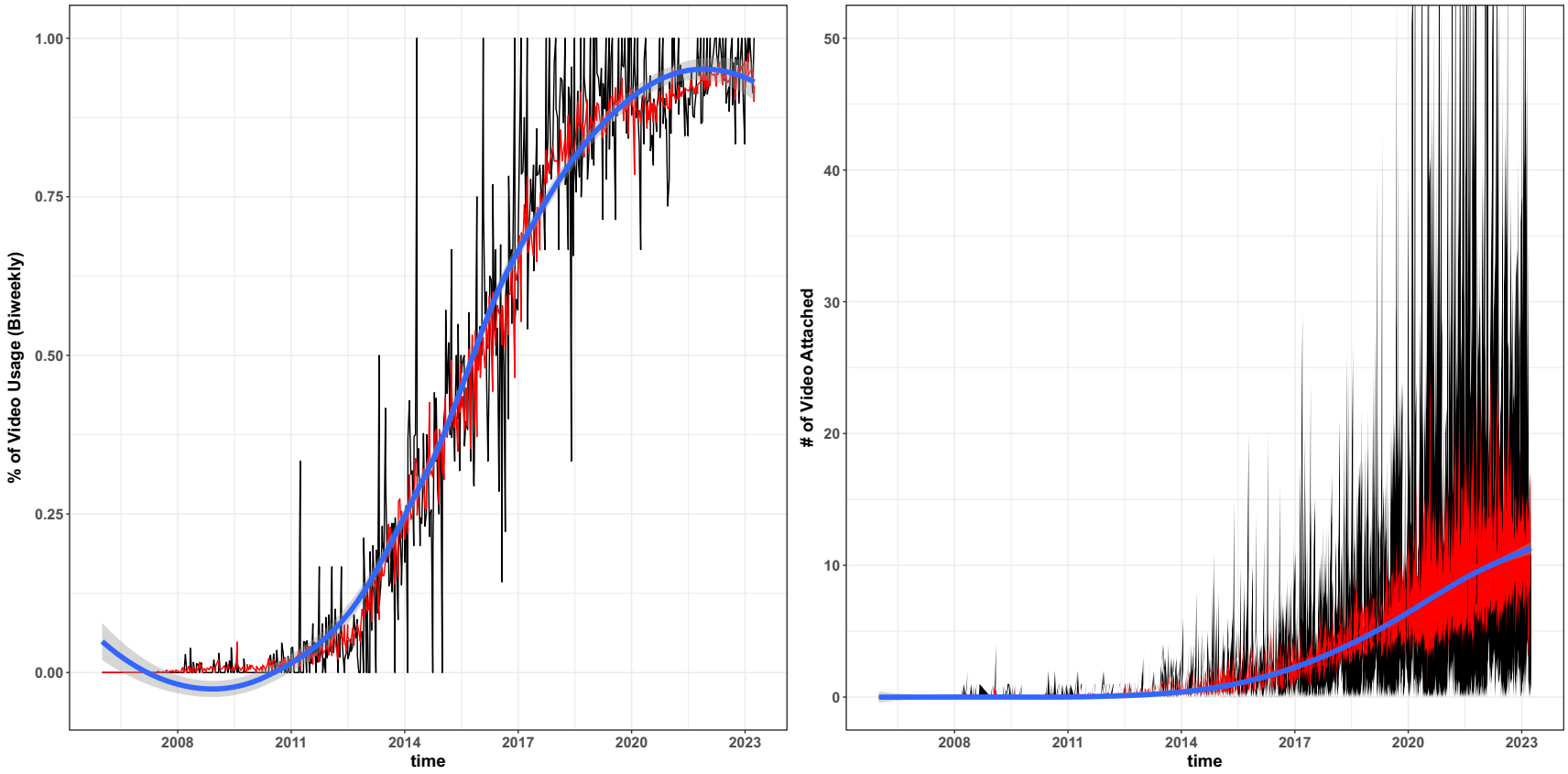
Figure 6**Error! Reference source not found.** plots the time trends in prediction errors for $AnyVideo_i$ and $VideoCount_i$ using only $kNNVid_{\%i}$ and $kNNVid_{\#i}$, respectively. The prediction errors for $VideoCount_i$ are essentially indistinguishable from zero across the entire observation period and across all five NYC boroughs (right panel). For $AnyVideo_i$, the prediction errors show a positive trend between 2014 and 2021. We do not view this as a direct challenge to the instrument's validity, since the sudden surge in video collection might be the very endogeneity we aim to rule out, such as a change in investigative priorities. Also, $kNNVid_{\%i}$ has excellent predictive performance by many standards (Accuracy = 0.892, ROC–AUC = 0.954, Brier Score = 0.079).

Lastly, we examine whether the instruments succeed in decoupling video evidence usage from the case characteristics and the solvability they entail that drive selective collection. Table 2

regresses the treatment variables ($AnyVideo_i$ and $VideoCount_i$) and the kNN instruments on case attributes, using $Primary\ Motive$, $Fatal$, $Location$, along with fixed effects. Model 1 and Model 4 confirm the endogeneity concern. They show that fatal shootings, gang- or drug-motivated cases, and incidents near public housing are all significantly more likely to involve video evidence, consistent with investigators devoting greater resources to harder-to-solve or higher-priority cases. These patterns illustrate why a naïve OLS estimate of video evidence's effect on clearance would be confounded.

Model 2 and Model 5 replace the endogenous treatment with the kNN instruments. The coefficients on all case characteristics collapse to near zero and become statistically indistinguishable from zero with high precision (lower SEs). What the change suggests is that, by conditioning on neighboring cases' video usage, the instruments strip away the individual case features that drive selective collection. What remains is thus the variation in video usage driven by local camera availability and investigative norms rather than case difficulty. We consider this as the basis for favoring the 2SLS over OLS estimates in the following analysis.

Figure 3: Real Video Usage Trends versus kNN-Predicted Trends.



Note: The black lines are the original trends from $\boldsymbol{AnyVideo_i}$. and $\boldsymbol{VideoCount_i}$. The blue lines are their locally estimated scatterplot smoothing (LOESS) curves. The red lines are the trends for $\boldsymbol{kNNVid_{\#i}}$ and $\boldsymbol{kNNVid_{\%i}}$

Figure 4: KNN-Predicted Probability of Any Video Usage in Shooting Investigation.

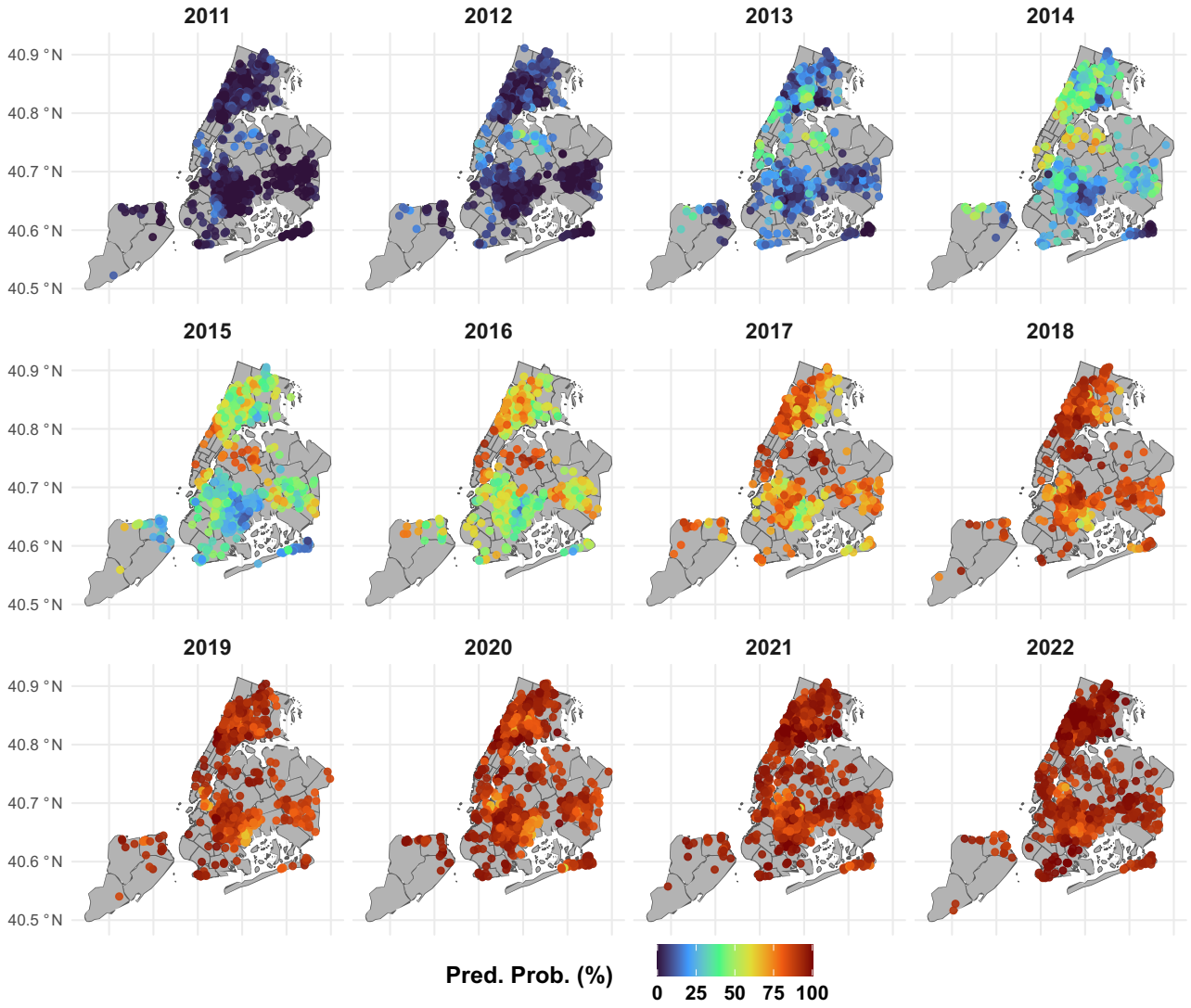
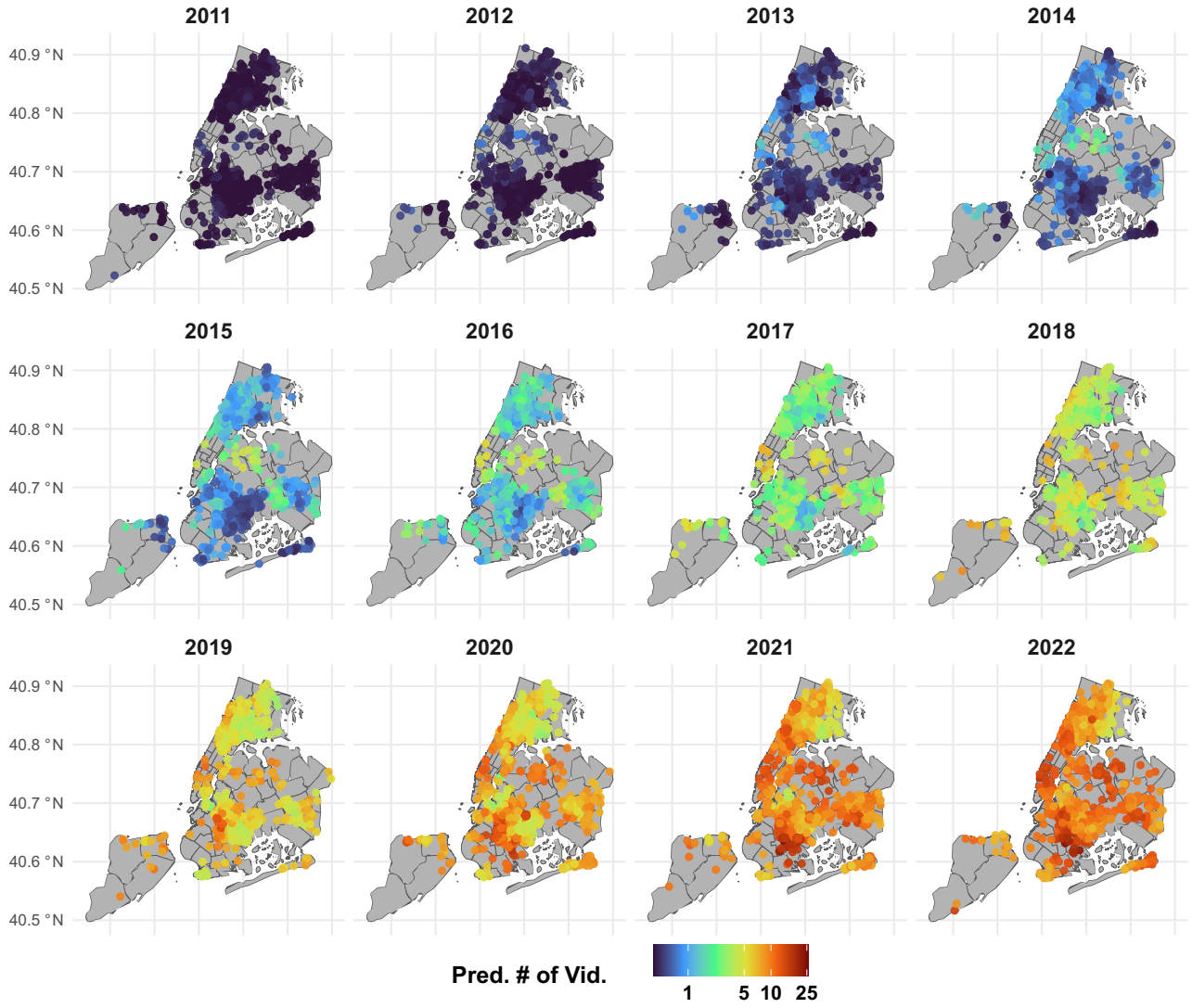
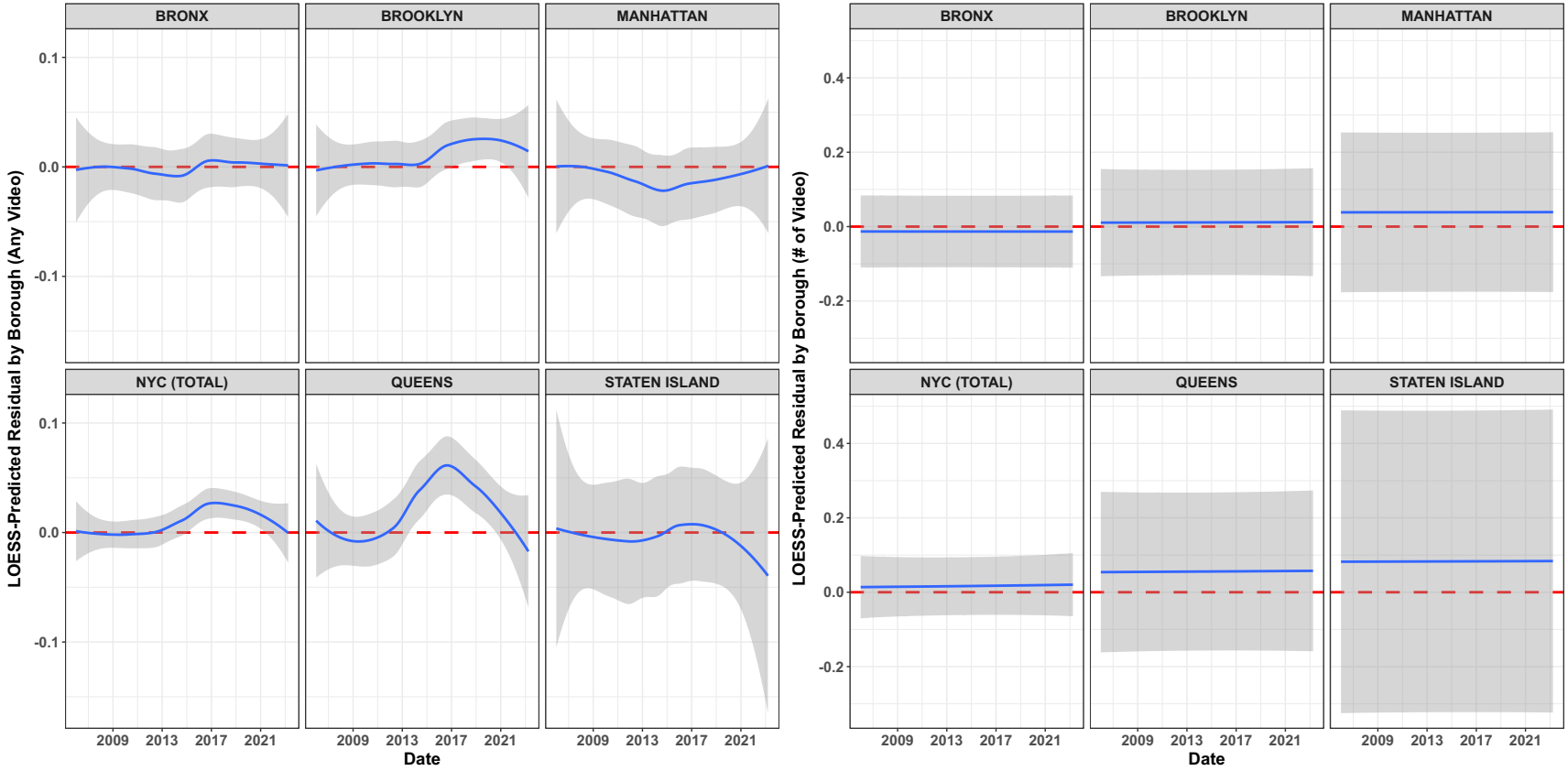


Figure 5: KNN-Predicted Number of Video Usage in Shooting Investigation.



Note: The predicted numbers are transformed by the inverse hyperbolic sine function (IHS) to increase color contrast.

Figure 6: LOESS-Fitted kNN Prediction Residuals by Time and Borough.



Note: The blue lines are LOESS curves for the residuals by time. The shaded areas are 95% confidence intervals.

Table 2: Predicting Different Treatment Definitions with Case Characteristics.

	Binary Treatment						Continuous Treatment					
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Dep. Var.:	$AnyVideo_i$		$kNNVid_{\%i}$		$kNNVid_{\%i}^{(2)}$		$VideoCount_i$		$kNNVid_{\#i}$		$kNNVid_{\#i}^{(2)}$	
Characteristic	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]
Primary Motive												
Dispute	—		—		—		—		—		—	
Drugs	0.011	0.008	0.004	0.002	0.003	0.002	0.426***	0.123	0.037	0.034	0.021	0.031
Gang	0.026**	0.008	-0.004	0.003	-0.004	0.003	1.882***	0.211	-0.001	0.048	0.015	0.044
Other	-0.006	0.005	0.000	0.001	0.000	0.001	0.157	0.084	0.027	0.029	0.028	0.025
Fatal												
No	—		—		—		—		—		—	
Yes	0.052***	0.005	-0.002	0.001	-0.001	0.001	2.751***	0.183	-0.026	0.026	-0.022	0.021
Location												
Commercial	—		—		—		—		—		—	
Dwelling	-0.030	0.019	0.005	0.004	0.004	0.004	-0.669	0.549	0.173	0.090	0.119	0.070
Housing	-0.048**	0.015	-0.006	0.004	-0.002	0.003	-0.876	0.452	-0.061	0.088	-0.008	0.063
Other	-0.042	0.029	-0.004	0.008	-0.006	0.006	-1.507	0.829	0.116	0.202	-0.037	0.175
Parking Lot	-0.031	0.041	0.005	0.011	0.012	0.008	0.056	0.964	0.038	0.178	0.117	0.146
Playground	-0.087***	0.025	0.008	0.006	0.010	0.005	-0.280	0.535	0.011	0.107	0.020	0.081
Street	0.000	0.013	0.000	0.004	0.001	0.003	0.002	0.439	0.090	0.072	0.073	0.058
Taxi/Livery	0.003	0.060	-0.006	0.012	-0.009	0.008	0.774	0.763	0.146	0.180	0.151	0.149
Transit	0.009	0.044	-0.009	0.012	-0.007	0.009	-0.495	0.912	0.873	0.611	0.236	0.175
Vehicle	0.002	0.020	0.003	0.006	0.004	0.004	0.816	0.578	0.035	0.130	0.031	0.098
Adjusted R^2	0.668		0.968		0.982		0.368		0.880		0.923	
N	18,396		18,396		18,396		18,396		18,396		18,396	
Wald Test F	21.298***		1.672		1.802		26.699***		1.924		1.425	

[1]*p<0.05; **p<0.01; ***p<0.001. [2]SE = Standard error clustered by precinct. $kNNVid_{\%i}^{(2)}$ and $kNNVid_{\#i}^{(2)}$ in Model 3 and Model 6 are constructed using a second-order feed-forward approach: the weighted sum of the 40 nearest neighbors' weighted sums of their 40 nearest neighbors' video usage (binary) and their number (count), with incident i removed if it is a neighbor's neighbor.

4. Results

4.1 First-Stage Regressions.

This section aims to establish the connection between the instruments and the actual video usage pattern in practice. Across all four models, the video evidence usage in the other spatiotemporally proximate cases strongly predicts whether video evidence will be collected in the focal case and to what extent. All t-statistics are larger than 8, far exceeding the commonly used threshold of 3 (Angrist, 2006), and remain robust to non-linear transformations.

Table 3 reports the first-stage regressions for Equations 3-3 and 3-43-4, each estimated under two different variable specifications. In Model 7, when $kNNVid_{\%i}$ predicts the probability of video usage in case i to be 1, the actual probability that video evidence is used in that case increases by 54.5 pp ($t = 8.963$). Model 8 replaces $kNNVid_{\%i}$ with its inverse hyperbolic sine (IHS) transformation,[13] $IHS(kNNVid_{\%i})$. This specification yields a slightly better fit ($t = 9.175$) and implies a 55 pp increase in the probability of video usage when the kNN-predicted probability equals 1.[14]

For the continuous treatment, a one-unit increase in the predicted number of attached videos, $kNNVid_{\#i}$, is associated with a 0.575 increase in $VideoCount_i$ ($t = 10.012$) (Model 9). Applying IHS transformation to both $kNNVid_{\#i}$ and VideoCount_i, a 1% increase in $kNNVid_{\#i}$ leads to an

[13] The IHS transformation, $\log(x + \sqrt{x^2 + 1})$ is chosen over the more common natural logarithm because $kNNVid_{\%i}$, $kNNVid_{\#i}$, and $VideoCount_i$ can take the value zero, which would yield $-\infty$ under the log transformation. IHS is preferred to $log(x + \epsilon)$ because it preserves continuity around zero and reduces researcher discretion in choosing ϵ. For count variables, the interpretation of coefficients with IHS-transformed outcomes is close to the standard percentage-change interpretation of log-transformed variables (not sensitive to unit, see Norton, 2022).

[14] $(IHS(1) = 0.881) \times 0.624 \times 100$.

approximate 0.453% increase in the number of videos attached to case i ($t = 10.452$) (Model 10). Notably, the adjusted R^2 indicates a substantially better fit than in Model 9, which uses the original count-level data.

Regarding the other covariates, $ClusterArrest_{\%}$ is significantly associated with a higher likelihood of video usage and with the number of attached videos when the latter is IHS-transformed. Consistent with Table 2, shootings that are gang-motivated, fatal, or occur near commercial properties are more likely to involve video evidence. In addition, drug-motivated shootings have significantly more videos collected.

4.2 2SLS Estimates

Table 4 and Table 5 show the estimated effects of using video evidence in shooting investigation clearance under different model specifications. Across all models, we find that the estimated benefits of video evidence are substantially larger than the OLS estimates. This implies video evidence is collected strategically, concentrated in cases that would otherwise be the hardest to solve.

For the binary treatment, the 2SLS shows that using any video evidence substantially increases the probability of case clearance by arrest by approximately 38.6 pp (Model 12). This effect is nearly nine times larger than the naïve OLS estimate of 4.2 pp (Model 11). The magnitude of this gap reveals strong negative selection: investigators disproportionately seek video evidence in cases that are harder to solve by other means, so OLS estimates severely understate video evidence's contribution. The logistic (2SRI) specification in Model 14 confirms the direction and magnitude.

Using video evidence increases the odds of making an arrest by 28.9%, consistent with the large positive effect found under 2SLS.[15]

With continuous treatment, each additional video attached to a case file increases clearance probability by 1.4 pp under 2SLS (Model 17), far larger than the OLS estimate of 0.2 pp (Model 15). The preferred IHS-transformed specification (Model 18) suggests that each doubling of the number of videos increases clearance probability by approximately 11.9 pp, reflecting the diminishing returns visible in the data. The first few videos matter most, and the marginal benefit of additional footage declines as the case file grows (see Section 4.4). Translating this to the original scale, to reach the 38.6 pp increase in arrest probability as in Model 13, it requires about 4–5 videos attached to the case report ($sinh\left(\frac{0.386}{0.172}\right) = 4.663$). This seems to be consistent with the binary treatment, in that the median number of videos attached to a case with any videos is 4, and the mean is 7. The logistic (2SRI) specification with the IHS-transformed video counts and IVs indicates that attaching one video results in a 25.1% increase in the odds of making an arrest. Adding 5 videos increases the odds by 79.9%.

For the other covariates, $ClusterArrest_{\%}$ shows a strong positive association with the arrest probability in all 8 models in Table 4 and Table 5. It suggests that shooting clearance is, to some

[15] Some may feel uncomfortable with the fact that LPM can result in fitted values that defy the basic definition of probability (between 1 and 0), even though LPM is commonly used in policy studies and applied econometrics due to its simplicity and intuitiveness. Serving as a robustness check, we will also present θ_1 and θ_2 estimated through logistic regression with the two-stage residual inclusion (2SRI) (Terza, 2017).

$$Arrest_i = \theta_1 AnyVideo_i + \gamma\zeta_i + \eta_2\, ClusterArrest_{\%i} + \boldsymbol{X_i}'\boldsymbol{\beta} + \tau_i + \epsilon_{i.}$$

$$Arrest_i = \theta_2 VideoCount_i + \gamma\zeta_i + \eta_2\, ClusterArrest_{\%i} + \boldsymbol{X_i}'\boldsymbol{\beta} + \tau_i + \epsilon_{i.}$$

The method uses the residuals from the first stage as an explicit control variable, ζ_i, for endogeneity in the second stage and replaces $\widehat{AnyVideo_i}$ and $\widehat{VideoCount_i}$ in Equations 3-1 and 3-2 with $AnyVideo_i$ and $VideoCount_i$. The uncertainty of 2SRI estimates will be quantified through clustered (by precinct) bias-corrected and accelerated (BCa) bootstrap confidence intervals.

extent, spatiotemporally clustered. Compared to shootings motivated by disputes, cases that are motivated by drugs, gangs, or other unknown reasons are 10–20 pp less likely to result in an arrest. Fatal shootings are 10–15 pp higher than non-fatal ones. Compared to shootings around commercial buildings, those that happened in a playground, on the street, or in a vehicle are less likely to result in an arrest, while those in a taxi increase the likelihood. No individual case attribute (fatality, motivation, or location) has as large a coefficient as the instrumented video evidence effect in Table 4. Investigative technology, when actually deployed, is the single most decisive factor in our analysis.

Table 3: First-Stage Regressions.

	Any Video				Video Count		Video Count (IHS)	
Characteristic	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]
	Model 7		**Model 8**		**Model 9**		**Model 10**	
$kNNVid_{\%i}$	0.545***	0.061						
$IHS(kNNVid_{\%i})$			0.624***	0.068				
$kNNVid_{\#i}$					0.561***	0.056		
$IHS(kNNVid_{\#i})$							0.453***	0.043
$ClusterArrest_{\%}$	0.047*	0.023	0.046*	0.023	0.500	0.424	0.212***	0.054
Primary Motive								
Dispute		—		—		—		—
Drugs	0.009	0.007	0.009	0.007	0.405**	0.122	0.061***	0.017
Gang	0.028***	0.007	0.028***	0.007	1.885***	0.202	0.216***	0.024
Other	-0.006	0.005	-0.006	0.005	0.143	0.080	0.001	0.014
Fatal								
No		—		—		—		—
Yes	0.053***	0.005	0.053***	0.005	2.767***	0.181	0.340***	0.018
Location								
Commercial		—		—		—		—
Dwelling	-0.032	0.019	-0.032	0.019	-0.764	0.545	-0.139**	0.052
Housing	-0.045**	0.014	-0.045**	0.014	-0.842	0.436	-0.167***	0.043
Other	-0.040	0.028	-0.039	0.028	-1.575	0.826	-0.264**	0.085
Parking Lot	-0.034	0.039	-0.034	0.038	0.029	0.928	-0.029	0.102
Playground	-0.092***	0.024	-0.092***	0.024	-0.287	0.522	-0.147*	0.056
Street	0.000	0.013	0.000	0.013	-0.048	0.433	0.001	0.040
Taxi/Livery	0.006	0.062	0.006	0.062	0.681	0.791	0.150	0.126
Transit	0.014	0.047	0.015	0.047	-0.983	0.985	-0.053	0.109
Vehicle	0.000	0.020	0.001	0.020	0.794	0.566	0.130*	0.060
Adjusted R^2	0.674		0.675		0.382		0.686	
N	18,396		18,396		18,396		18,396	

[1]*p<0.05; **p<0.01; ***p<0.001. [2]Clustered by precinct. Abbreviations: CI = Confidence Interval, SE = Standard Error

Table 4: Regressions of Clearance with Binary Treatment.

	OLS		2SLS		2SLS (IHS)[3]		Logit (2SRI)	
	Model 11		Model 12		Model 13		Model 14	
Characteristic	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**OR**[14]	**95% CI**[5]
Any Video	0.042***	0.012	0.386***	0.105	0.380***	0.102	1.289***	1.134, 1.452
First-Stage Residual							0.940*	0.902, 0.987
$ClusterArrest_{\%}$	0.215***	0.047	0.178**	0.052	0.179***	0.052	2.477***	1.694, 4.080
Primary Motive								
Dispute		—		—		—		—
Drugs	-0.146***	0.015	-0.150***	0.015	-0.150***	0.015	0.502***	0.446, 0.564
Gang	-0.115***	0.012	-0.124***	0.012	-0.124***	0.012	0.531***	0.467, 0.604
Other	-0.215***	0.011	-0.213***	0.011	-0.213***	0.011	0.366***	0.338, 0.395
Fatal								
No		—		—		—		—
Yes	0.165***	0.016	0.147***	0.018	0.147***	0.018	1.798***	1.553, 2.081
Location								
Commercial		—		—		—		—
Dwelling	-0.054	0.047	-0.044	0.045	-0.044	0.045	0.815	0.557, 1.193
Housing	-0.096**	0.036	-0.079*	0.037	-0.080*	0.037	0.682**	0.511, 0.909
Other	-0.130*	0.057	-0.115	0.059	-0.116	0.059	0.593	0.346, 1.017
Parking Lot	-0.059	0.075	-0.048	0.077	-0.048	0.077	0.777	0.417, 1.449
Playground	-0.165**	0.053	-0.134*	0.054	-0.135*	0.054	0.478***	0.321, 0.711
Street	-0.078*	0.035	-0.078*	0.034	-0.078*	0.034	0.701*	0.532, 0.924
Taxi/Livery	0.201*	0.078	0.200*	0.077	0.200*	0.077	2.551*	1.061, 6.130
Transit	0.100	0.090	0.097	0.087	0.097	0.087	1.757	0.757, 4.080
Vehicle	-0.134**	0.045	-0.135**	0.046	-0.135**	0.046	0.507***	0.341, 0.755
Adjusted R^2	0.102		0.059		0.060		—	
Pseudo R^2	—		—		—		0.104	
N	18,396		18,396		18,396		18,396	
First-Stage F	—		342.888		354.994		—	

[1]*p<0.05; **p<0.01; ***p<0.001. [2]Clustered by precinct. [3]The IV was transformed by the inverse hyperbolic sine function in the first stage. [4]The p-values were estimated through inverting the confidence intervals. [5]Bias-correct and accelerated bootstrap confidence interval (clustered by precinct, B=1,500). Abbreviations: CI = Confidence Interval, OR = Odds Ratio, SE = Standard Error.

Table 5: Regression of Clearance with Continuous Treatment.

	OLS		OLS (IHS)		2SLS		2SLS (IHS)[3]		Logit (2SRI)	
	Model 15		Model 16		Model 17		Model 18		Model 19	
Characteristic	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**Beta**[1]	**SE**[2]	**OR**[12]	**95% CI**[5]
# of Video	0.002*	0.001			0.014**	0.005				
# of Video$_{IHS}$[3]			0.021***	0.005			0.172***	0.035	1.158**	1.050, 1.275
First-Stage Residual									0.973	0.928, 1.029
ClusterArrest$_{\%}$	0.217***	0.046	0.211***	0.046	0.199***	0.045	0.153**	0.049	2.620***	1.732, 4.279
Primary Motive										
Dispute		—		—		—		—		—
Drugs	-0.147***	0.015	-0.147***	0.015	-0.152***	0.015	-0.157***	0.016	0.509***	0.453, 0.572
Gang	-0.118***	0.012	-0.119***	0.012	-0.140***	0.013	-0.151***	0.014	0.573***	0.515, 0.638
Other	-0.216***	0.011	-0.215***	0.011	-0.217***	0.011	-0.216***	0.011	0.368***	0.340, 0.398
Fatal										
No		—		—		—		—		—
Yes	0.162***	0.016	0.160***	0.016	0.129***	0.019	0.109***	0.020	2.005***	1.803, 2.228
Location										
Commercial		—		—		—		—		—
Dwelling	-0.054	0.047	-0.053	0.047	-0.046	0.050	-0.034	0.048	0.795	0.544, 1.161
Housing	-0.096**	0.036	-0.094*	0.036	-0.086*	0.038	-0.068	0.039	0.661**	0.496, 0.880
Other	-0.129*	0.057	-0.126*	0.057	-0.111	0.059	-0.086	0.061	0.560*	0.327, 0.957
Parking Lot	-0.060	0.075	-0.059	0.075	-0.061	0.077	-0.055	0.078	0.777	0.417, 1.446
Playground	-0.168**	0.054	-0.165**	0.053	-0.164**	0.053	-0.143**	0.053	0.472***	0.318, 0.703
Street	-0.078*	0.035	-0.078*	0.035	-0.078*	0.036	-0.079*	0.035	0.702*	0.533, 0.924
Taxi/Livery	0.199*	0.078	0.197*	0.078	0.190*	0.079	0.173*	0.078	2.597*	1.081, 6.241
Transit	0.101	0.091	0.101	0.090	0.107	0.094	0.105	0.091	1.718	0.742, 3.978
Vehicle	-0.136**	0.046	-0.137**	0.046	-0.145**	0.047	-0.157**	0.046	0.523**	0.352, 0.777
Adjusted R^2	0.102		0.102		0.102		0.103			
Pseudo R^2	—				—		—		0.105	
N	18,396		18,396		18,396		18,396		18,396	
First-Stage F	—				388.535		351.378			

[1]*p<0.05; **p<0.01; ***p<0.001. [2]Clustered by precinct. [3]The IV was also transformed by the inverse hyperbolic sine function in the first stage. [4]The p-values were estimated through inverting the confidence intervals. [5]Bias-correct and accelerated bootstrap confidence interval (clustered by precinct, B=1,500). Abbreviations: CI = Confidence Interval, OR = Odds Ratio, SE = Standard Error.

4.3 Robustness Checks

The preferred model estimates depend on choices about covariates, sample composition, instrument construction, and variable operationalization. Table 6 varies each of these choices to test whether the core findings are sensitive to any single decision. Nine variants are tested for the three treatment variables (any video, video count in levels, and IHS-transformed video count), yielding a 9×3 grid of robustness estimates. The baseline preferred estimates from Section 4.2 appear in Row (1) for comparison. All models exhibit strong first-stage F-statistics for the instruments, ranging from 136.05 to 471.48.

Rows (2) and (3) re-estimate the preferred models without any case-level covariates and with $ClusterArrest_{\%i}$ as the sole covariate, respectively. The estimates are slightly higher when all covariates for motive, fatality, location, and surrounding-area average clearance are dropped. When the models include only $ClusterArrest_{\%i}$ and omit other case attributes, however, the estimates become slightly smaller than in the preferred models.

Row (4) restricts the samples to the 16,575 outdoor criminal shootings without an arrest within two days. According to the solvability thesis, some shootings are inherently easier to solve, regardless of the subsequent resources/effort devoted to the investigation (Cook & Mancik, 2024), especially those with an arrest within 2 days (Ashby, 2017; Cook et al., 2019). Including these "self-solving" cases, such as on-scene arrests, may result in an underestimation of video evidence's contribution, given that detectives might not even consider collecting it in the first place. However, the estimates from the restricted samples show only mild increases across the three models with reduced precision.

Table 6: Summary of Robustness Checks.

		Treatment Variable		
	Specification	**Any Video**	**# of Video (Level)**	**# of Video (IHS)**
(1)	Preferred	0.386***	0.014**	0.172***
		(0.105)	(0.005)	(0.035)
(2)	No Case Covariates	0.441***	0.016*	0.191***
		(0.106)	(0.006)	(0.039)
(3)	$ClusterArrest_{\%}$ only	0.368***	0.013*	0.162***
		(0.105)	(0.005)	(0.037)
(4)	Exclude within 2-day arrests (N=16,575)	0.417**	0.014**	0.177***
		(0.123)	(0.005)	(0.037)
(5)	Best-k IV	0.405***	0.014**	0.185***
		(0.107)	(0.005)	(0.040)
(6)	Best-k for IV & $ClusterArrest_{\%i}$	0.333**	0.012*	0.161***
		(0.110)	(0.005)	(0.042)
(7)	k-of-k neighbors IV	0.364***	0.013*	0.164***
		(0.106)	(0.005)	(0.037)
(8)	Case Closure as the Dep. Var.	0.383***	0.014**	0.174***
		(0.103)	(0.004)	(0.032)
(9)	Year-month-borough Fixed Effects (N=18,320)	0.398*	0.016*	0.187***
		(0.169)	(0.006)	(0.050)

Note: All models include precinct, year-month, and weekday-by-hour fixed effects. Standard errors are clustered on the precinct level.

Rows (5) and (6) examine alternative tuning parameter k for the IVs and for $ClusterArrest_{\%i}$. Row (5) switches the k from 40 to 167 and 155, respectively, which deliver the best out-of-sample prediction of case i's video usage (see Section 3.3). Under these alternative instruments, the estimates become slightly larger than in the preferred models. Row (6) further adjusts the k for $ClusterArrest_{\%i}$ to the value that best predicts Arrest_i. The kernel kNN does not find an optimum within $k \leq 250$, so we extend the search to 600. The best prediction occurs at k = 360, yet accuracy improves only from 59.4% (k = 40) to 59.8%. When all kNN-based variables in the preferred model are replaced with their best-k versions, the shrinkage in the estimated effect sizes

becomes more pronounced. However, the estimates still do not differ significantly from the preferred ones.[16]

To further address concerns about potential residual endogeneity of the proposed IVs, Row (7) draws on the literature that uses peers-of-peers in a leave-one-out network as valid instruments for peer effects (Bramoullé et al., 2009; Jochmans, 2023). We apply this idea to neighbors-of-neighbors: for each case i, we take the inverse-distance-weighted mean, across its 40 nearest neighbors, of each neighbor's own inverse-distance-weighted mean over its 40 nearest neighbors, always excluding the case i itself. Model 3 and Model 6 in Table 2 show that these second-order, k-of-k neighbor IVs correlate even more weakly with case-level covariates. Using these new instruments, the estimates barely change relative to the preferred models.

Using arrest as a proxy for clearance may raise the concern that the video-driven increase in arrests is inflated. The presence of video evidence may make police overconfident in their suspect identification when other important evidence (such as DNA, ballistic analysis, or witnesses) is missing, leading to more wrongful arrests that are subsequently challenged by the district attorney or in court. If this is common, video evidence may not substantially improve case clearance or deliver justice to victims, but instead render investigations less efficient and more costly.

To provide an alternative measure of case clearance, Row (8) replaces the binary arrest outcome with a stricter standard: whether the department considers the case "closed". The latter is a subset of the former, as only 84.48% of cases with an arrest are also tagged as closed. Even under this stricter definition, the estimated effects of video evidence usage on clearance barely change,

[16] The control of $ClusterArrest_{\%}$ with 360 neighbors appears concerning not just because of the low accuracy, but also the extensive scope of search that seems implausible and noisy. The median absolute difference to the neighbors under this scope is 2 years in time and 0.957 miles in distance. This is why this specification only serves as a robustness check.

providing further confidence that the improved clearance does not reflect more haphazard arrests prompted by video footage.[17]

Lastly, because our instruments rely on spatiotemporal patterns of video use, they may also pick up unobserved local factors, such as investigative capacity or community cooperation, that directly affect arrest probability. The significant positive coefficients on $ClusterArrest_{\%}$ point to the existence of such factors, showing that shooting clearances cluster in space and time even net of clustering in video use.

Row (9) addresses this by adding finer spatiotemporal fixed effects. Including year–month–borough fixed effects (dropping 75 singletons) halves the coefficient on $ClusterArrest_{\%}$ and makes it insignificant, while the video evidence estimates remain stable. This suggests that the spatiotemporal clustering of clearance is driven mainly by video use rather than by other unobserved local factors under this specification.

To summarize, the estimated benefits of video evidence usage identified by our proposed IVs remain stable across a range of conditions. We therefore conclude that the preferred models in Row (1) are robust and adopt them as the primary reference for the subsequent analyses and policy discussion.

[17] We acknowledge that this remains an imperfect measure of clearance, since we have no data on whether arrestees were ultimately charged or convicted.

4.4 The Non-Linear Reward of Video Canvassing

As Section 4.2 showed, the dose-response curve between the number of videos attached to a case and the clearance rate is distinctly non-linear, tracing the shape of an inverse hyperbolic sine function (Model 18). The marginal return to having one additional video is therefore not constant but depends on the amount of video evidence already collected. Assuming that video evidence is collected sequentially, as detectives gradually broaden the scope of their search (Barao & Mastroianni, 2024; Brookman & Jones, 2022), rather than all at once, and that there is a certain point where no more additional relevant videos can be easily recovered, then the distinction between constant and non-linear returns carries an important policy trade-off: sometimes the search for video should be expanded and other times it should be halted to prioritize other forms of evidence.

Based on Model 18, Figure 7 maps the marginal effects of attaching one more video to the case file when 0 to 29 videos have already been attached, approximating this sequential decision process. The effects are estimated using G-computation, which assigns the same treatment value to every shooting in the dataset, computes the expected outcome under the model, and compares two assignment scenarios to recover the average treatment effect; the corresponding variance is obtained via the Delta method. All computations are carried out using the `marginaleffects` package in R (Arel-Bundock et al., 2024).

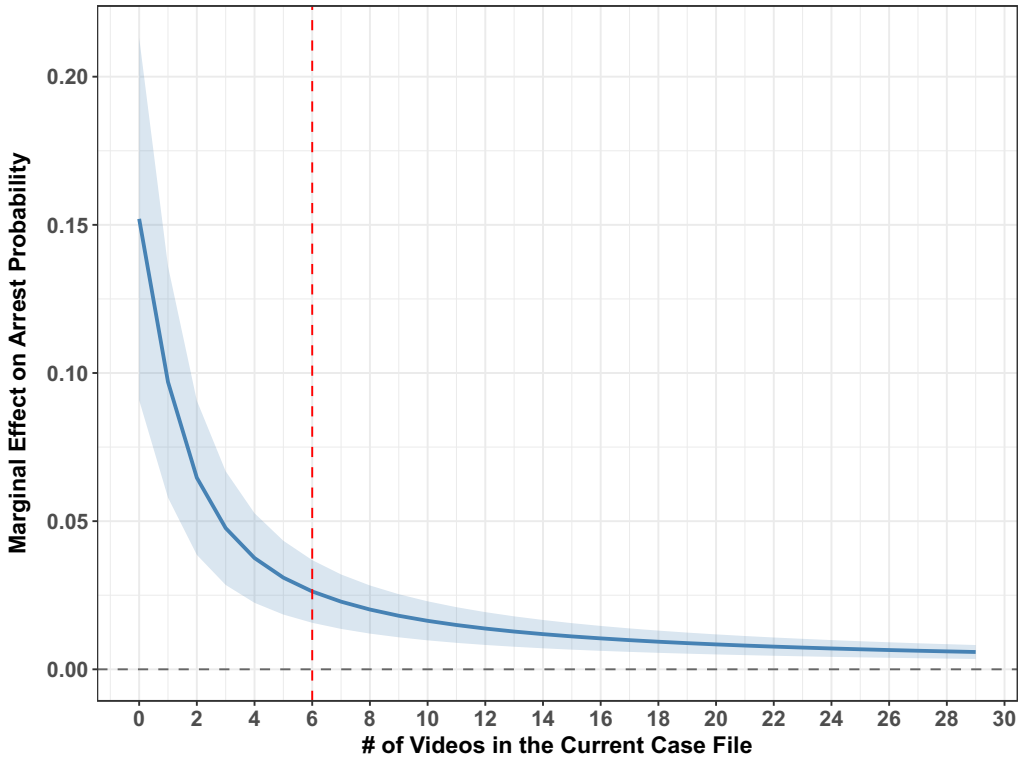
The marginal effects resemble an exponential decay, strongest at the outset and flattening in the tail. The largest effect occurs when a shooting moves from having no video to a single video, raising the probability of an arrest by 15.2 pp. Moving from one to two videos, the additional gain (9.95 pp) is less than two-thirds of the first. The decline in marginal benefit then slows and stabilizes after six videos, at around 2.63 pp, and the cumulative effect of having seven videos

reaches 45.61 pp. Notably, the linear effect originally estimated in Model 17, 1.4 pp per video, corresponds to the marginal effect after twelve videos have already been attached, showing how different the conclusions would have been if the dose-response relationship had been mis-specified.

One might consider the dose-response relation implausibly smooth and that a more flexible functional form would better capture it. We re-estimate Model 18 with the IHS-transformed video count entering the same regression through a natural cubic spline using 2SRI (Figure 8).[18] The two marginal effect curves track closely, with the estimates with the spline model showing a slower decay from six to eleven videos and a faster decay thereafter. There is no clear point at which additional videos cease to be helpful (a marginal effect of zero) or actively reduce the probability of clearance (a negative marginal effect).

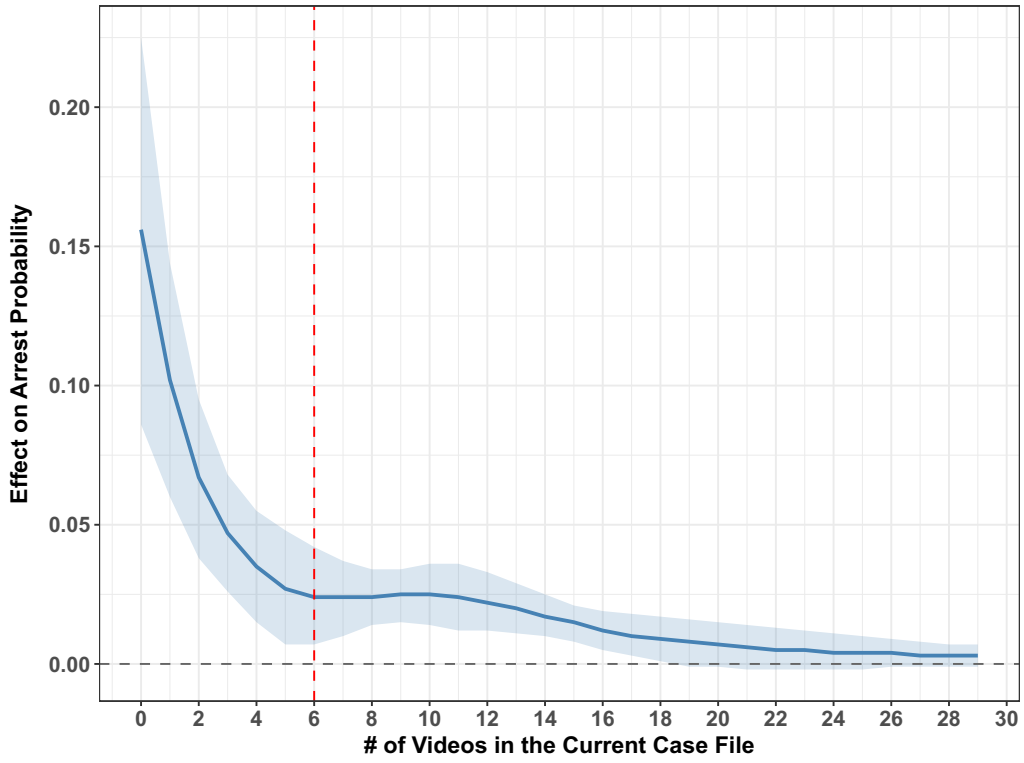
[18] 2SRI was used because the spline terms make the treatment variable enter the second stage nonlinearly through multiple basis functions, which do not fit within the traditional 2SLS framework. 2SRI, on the other hand, preserves the comparability between the original specification and the regressions with splines. We also compared the model with the raw video count entering the spline function, but the fit is not as good. For more details, such as model specifications and goodness-of-fit for the three models, please see the Appendix.

Figure 7: Marginal Effects of One Additional Video Attached (Model 18).



Note: The thick blue line represents the marginal effect on arrest probability of adding one more video to the case management file, based on Model 18. The shaded area is the 95% confidence interval.

Figure 8: Marginal Effects of One Additional Video Attached (Spline).



Note: The thick blue line represents the marginal effects on arrest probability by adding one more video to the case management file, estimated through a 2SRI model with a natural cubic spline with five interior knots at $IHS(1)$, $IHS(5)$, $IHS(10)$, $IHS(15)$, and $IHS(25)$. The shaded area is the 95% confidence interval estimated from 1,500 clustered bootstrap samples at the precinct level.

4.5 Video Evidence and Citywide Clearance Trend.

As briefly discussed in Section 3.1, Figure 1 reveals a peculiar upward inflection in the shooting clearance rate since 2012. The open question is whether the rise in video evidence usage in NYC shooting investigations helps explain this shift.

To probe this question, we examine time fixed effects across different model specifications. The logic is as follows: if our model adequately controls for time-varying factors that contribute to shifts in citywide shooting clearance, the estimated time fixed effects on clearance probability should trace a monotonically decreasing trend that reflects only the effect of time pressure/abundance on case clearance. That is, the shorter the time from the shooting to the date, the less likely an arrest has been made. When the time fixed effects deviate from this expected trajectory, it suggests they are absorbing unobservable and thus unexplained temporal trends that contribute to the improving shooting clearance.[19]

Figure 9 presents year-quadrimester (4 months) time fixed effects from four model specifications: two adapted from the OLS estimators in Equations 3-1 (Panel A) and 3-2 (Panel C), and two from the 2SLS estimators in Equations 3-3 (Panel B) and 3-4 (Panel D).[20] All four models include $ClusterArrest_{\%}$ as a covariate. The last quadrimester of 2012 is set as the reference period—the time when the Domain Awareness System (DAS) started expanding from

[19] The converse does not hold. A monotonically decreasing trend does not imply that the model captures the entire temporal trend through the covariates. Omitted time-covarying variables may still influence the speed (acceleration) of the downward trend.

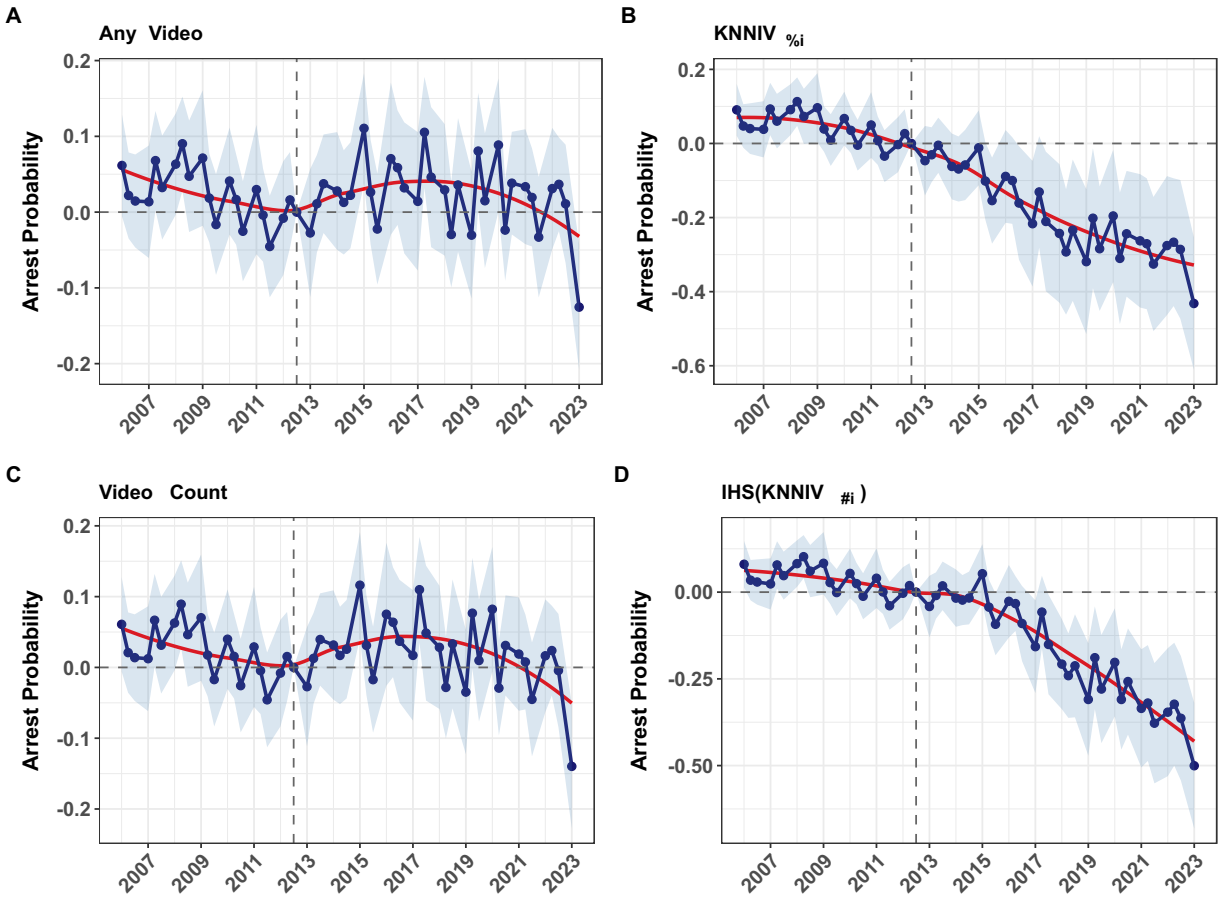
[20] The condensation of the year-month fixed effects into year-quadrimester fixed effects is to improve the readability of the plots and reduce noise from having numerous fixed-effect estimates.

lower Manhattan across the city (Ungerleider, 2012), and the video evidence used in shooting investigations experienced a sudden acceleration.[21]

The key finding from the analysis is in the contrast between the OLS and IV panels. Under OLS (Panels A and C), the time fixed effects still show an unexplained upward inflection beginning at the end of 2012, suggesting that time-varying confounders are still being absorbed into the coefficients even after the covariate adjustments, including video use, in the regressions. Once video evidence is properly instrumented through 2SLS (Panels B and D), the trend becomes monotonically decreasing throughout the observation period, as we expect: later cases are mechanically harder to clear due to time pressure from the administrative censoring date. The "correction" of the trajectory provides strong evidence that the increased clearance rates since 2013 can be explained by the growth in video evidence collection across the city. An important nuance, though, is that this growth is better captured by how video usage practice evolves across different microlocations of the city at the time than by selective video collection at the individual case level.

[21] We also conduct an analysis using the public announcement date, August 8, 2012, of the system's expansion as a cutoff to understand if the system has an immediate effect on video collection practice in shooting investigations. Using a regression kink design, we demonstrate an acceleration in the growth trend of video evidence collection around the cutoff date. However, this analysis has limitations. It focuses only on a very short period, and the robust z-statistics are insufficient for a second-stage analysis of its impact on clearance rates. For more details, see the Appendix.

Figure 9: Year-Quadrimester Time Fixed Effects.



Note: The blue dots are point estimates from the quadrimester time fixed effects under different model specifications. The shaded areas are 95% confidence intervals. The red line is the smoothed trend for these estimates. The title of each panel represents the treatment variable in the models with the time fixed effects. All the models estimated include precinct and weekday-by-hour fixed effects.

5. Discussion and Conclusions

Drawing on data from New York City, this study provides evidence that the use of video evidence in criminal investigations substantially increases the probability of case clearance by arrest. On average, a shooting with video evidence attached to its case file is 33.3 to 44.1 pp more likely to result in an arrest than one without. This result is notable for how it appears to depart from much of the prior U.S. evidence. As discussed in the literature review, we believe the discrepancy stems mainly from our focus on the actual use of video evidence in investigations, rather than the mere

proximity of an incident to known public CCTV cameras (Jung & Wheeler, 2023; Piza, Connealy, et al., 2025; Robin et al., 2021; Thomas, 2025). This shift closes the latent gap between researchers' intended measure of treatment and actual investigative practice. It also avoids the measurement error introduced by the unknown distribution of private, non-police-owned cameras, which often constitute a non-negligible share of the video evidence accessible to police (Cook & Berglund, 2021).

The selective collection of video evidence further complicates the treatment-on-treated and intent-to-treat gap. By instrumenting the use of video evidence in cases with video usage patterns from their spatiotemporally close neighbors, our analysis reveals strong *negative selection* in the use of video evidence: Police investigators seek video evidence more often in cases where suspects are intrinsically harder to identify or apprehend (see also Ashby, 2017). This pattern is more consistent with the triage thesis, in which limited resources, including time and effort for video retrieval and review, are concentrated on the harder cases that would otherwise go unsolved (Coupe, 2016; Eck, 1983; Prince et al., 2021). The true strength of this triage tendency may be even greater than our estimates suggest, because other unobserved investigative input (e.g., witness interviews, DNA collection, ballistic analysis) is likely deployed more intensively alongside video evidence as a result of triaging, inducing an offsetting positive selection that pushes the naïve estimate back toward overestimation.

Beyond establishing that video evidence matters, the study maps a dose-response relationship between the amount of video collected and the probability of case clearance. The results are broadly consistent with the surveillance network hypothesis: the more video is used, the more likely a case will be cleared. The nuance, however, is that the marginal contribution of each additional video is uneven. The strongest effect appears in the move from no video to a single

video, which raises the clearance rate by 15.2 pp; the additional return then decays smoothly along a soft L-shaped curve as more videos accumulate.

Based on a sequential search-and-addition assumption (Section 4.4), the pattern we observed aligns with previous research indicating that the first few videos recovered from the immediate scene are the most critical. They are more likely to reveal key details about the suspects, including their appearance, demeanor, gait, vehicles, and weapons (Chen, 2016). These traits, in turn, allow investigators to trace suspects backward and forward in time across a broader geographic area to generate circumstantial information (Brookman & Jones, 2022; Chen, 2016), such as preceding events (drug transactions, fights, or quarrels), contacts, or movement after the incident. According to Field, Barao, and Mastroianni (2024), this type of information, although less helpful when video directly links identification to criminal actions, can still significantly increase arrest rates. Moreover, because the search for identifying footage must begin from the location of the incident and expand outward, the diminishing returns also mirror Jung and Wheeler's (2023) finding that the clearance benefit of the nearest public CCTV camera decays rapidly with distance, assuming that subsequent videos come from cameras farther away from the incident location.[22]

By contrast, the "mosaic theory" frequently invoked in legal scholarship on technological surveillance (Brayne, 2020a; Kerr, 2012) does not fit in our L-shaped dosage effect estimates. The theory holds that key identifying information *emerges* when many data points are assembled, with any single point (a video) or a single tile of the mosaic carrying little meaning until it is embedded

[22] Still, taking the elbow method as a rule of thumb, a case that already has around seven videos attached yet remains unsolved may mark the point at which investigators should begin weighing other forms of evidence. Beyond this point, the gain in clearance probability is small, while the cost of canvassing for additional videos, which we do not observe in our data, likely continues to rise. We discuss the policy implications and limitations of this analysis in greater detail in the last section.

within the larger picture formed by the rest (many videos). If this logic governed the value of video evidence in crime investigation, the marginal-effect curve would instead have traced a reverse-L shape, with small-to-zero marginal effects for the first few videos until a suspect's identity abruptly emerges once a sufficient amount of the "big picture" is collected.[23]

At the macro level, our study contributes to the ongoing discussion of trends in crime clearance across major U.S. cities (Cook & Mancik, 2024; Ferrandino, 2021) and the role of technology in shaping them. Consistent with much of the literature on police technology, our findings suggest that the presence of a technology alone is insufficient to realize its expected benefits. In NYC, the relatively slow growth in street-level CCTV cameras contrasts with the S-shaped trajectory of video evidence use in shooting investigations, suggesting that technological availability alone cannot explain changes in investigative practice. Instead, the organizational processes through which police routinize technology use appear central to realizing its benefits (Koper et al., 2014; Koper & Lum, 2019a; Lum et al., 2019; Piza et al., 2014). Our IV analysis, together with our examination of the unobserved influences captured by the time fixed effects in the model, provides strongly suggestive evidence that NYC's overall improvement in shooting clearance is better explained by the variation in time and place that video evidence collection and its extensiveness became a shared norm across different cases than by individual investigators' discretionary use of video when investigations stall. These results again challenge "technological determinism" in policing (Norris & Armstrong, 2020; Piza, Arietti, et al., 2025; Piza et al., 2014) and highlight the

[23] Our findings are not a conclusive falsification of the "mosaic theory" as a whole. The theory may still hold true for various other types of digital evidence that tend to be more circumstantial than what video footage offers, including IP addresses, GPS locations, social media posts, or personal communications.

limitations of research designs that estimate the effect of technology installation without observing its use.

5.1 Limitations and Future Research Directions

The findings of this study should be interpreted in light of NYC's distinctive surveillance environment and the limited availability of more granular data from the NYPD's case management system. NYC has a relatively high density of CCTV cameras compared with many other major U.S. cities. As a point of reference, Sheng et al. (2021) estimate that, among the cities included in their survey, NYC's CCTV density in 2021 was 0.62 cameras per square kilometer, second only to Boston at 0.63. This figure is higher than London (0.45), a city commonly associated with extensive public surveillance. Our results do not identify the minimum camera density around a crime incident required for police organizations to use video evidence effectively in investigations. It therefore remains unclear whether the estimated dose-response relationship in NYC would hold in jurisdictions with substantially lower camera density, different spatial layouts, or less developed video-evidence infrastructure. Future research should therefore replicate this study design in other locations.

The second major limitation is that our data do not allow us to examine whether video substitutes for or complements other types of evidence, such as witness testimony, DNA, ballistics, or latent prints. Part of the estimated effect of video may come not from the footage itself but from additional evidence it helps investigators obtain, such as identifying witnesses or locating surfaces for print collection. Conversely, the presence of video may lead investigators to collect less traditional evidence that could still aid the investigation. Future research should examine whether video collection supports the collection of other forms of corroborating evidence or instead crowds it out.

References

Adams, A. A., & Ferryman, J. M. (2015). The future of video analytics for surveillance and its ethical implications. *Security Journal*, *28*(3), 272–289. https://doi.org/10.1057/sj.2012.48

Anderson, J. M., Matthies, C. F., Greathouse, S. M., & Chari, A. V. (2021). The Unrealized Promise of Forensic Science: A Study of Its Production and Use. *Berkeley Journal of Criminal Law*, *26*. https://www.rand.org/pubs/external_publications/EP68695.html

Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: What, why and how. *Journal of Experimental Criminology*, *2*(1), 23–44. https://doi.org/10.1007/s11292-005-5126-x

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to Interpret Statistical Models Using marginaleffects for R and Python. *Journal of Statistical Software*, *111*, 1–32. https://doi.org/10.18637/jss.v111.i09

Arsenault, M., Bogel-Burroughs, N., & Gibbons-Neff, T. (2025, December 19). How a Reddit Post Helped the Police Identify a Suspect in the Brown University Shooting. *The New York Times*. https://www.nytimes.com/2025/12/19/us/brown-mit-shooting-investigation-reddit.html?campaign_id=190&emc=edit_ufn_20251219&instance_id=168208&nl=from-the-times&regi_id=258229798&segment_id=212509&user_id=1e604ce3b38801e725e127fbca1920ea

Ashby, M. P. J. (2017). The Value of CCTV Surveillance Cameras as an Investigative Tool: An Empirical Analysis. *European Journal on Criminal Policy and Research*, *23*(3), 441–459. https://doi.org/10.1007/s10610-017-9341-6

Aziani, A., & Persurich, C. (2023). Homicide clearance: Discretionary and non-discretionary factors. *European Journal of Criminology*, *20*(3), 947–972. https://doi.org/10.1177/14773708221136049

Barao, L., & Mastroianni, C. (2024). Creating Solvability With Real-Time Crime Centers (RTCCs): Impacts on Homicide and Shooting Investigations. *Police Quarterly*, 10986111241290143. https://doi.org/10.1177/10986111241290143

Baskin, D., & Sommers, I. (2012). The Influence of Forensic Evidence on the Case Outcomes of Assault and Robbery Incidents. *Criminal Justice Policy Review*, *23*(2), 186–210. https://doi.org/10.1177/0887403410395576

Blake, S. (2023, August 16). How Understanding Video Evidence Is Vital for Your Organization. *Police Chief Magazine*. https://www.policechiefmagazine.org/amped-how-understanding-video-evidence-is-vital-for-your-organization/?ref=d44f031b778ddc40e9cf7efeabcda78a

Braga, A. A., & Dusseault, D. (2018). Can Homicide Detectives Improve Homicide Clearance Rates? *Crime & Delinquency*, *64*(3), 283–315. https://doi.org/10.1177/0011128716679164

Braga, A. A., Turchan, B., & Barao, L. (2019). The Influence of Investigative Resources on Homicide Clearances. *Journal of Quantitative Criminology*, *35*(2), 337–364. https://doi.org/10.1007/s10940-018-9386-9

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, *150*(1), 41–55. https://doi.org/10.1016/j.jeconom.2008.12.021

Brayne, S. (2020a). Algorithmic Suspicion and Big Data Searches: The Inadequacy of Law in the Digital Age. In S. Brayne (Ed.), *Predict and Surveil: Data, Discretion, and the Future of Policing* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780190684099.003.0007

Brayne, S. (2020b). Dragnet Surveillance: Our Incriminating Lives. In S. Brayne (Ed.), *Predict and Surveil: Data, Discretion, and the Future of Policing* (pp. 37–55). Oxford University Press. https://doi.org/10.1093/oso/9780190684099.003.0003

Brookman, F., & Jones, H. (2022). Capturing killers: The construction of CCTV evidence during homicide investigations. *Policing and Society*, *32*(2), 125–144. https://doi.org/10.1080/10439463.2021.1879075

Brookman, F., Maguire, E. R., & Maguire, M. (2019). What Factors Influence Whether Homicide Cases Are Solved? Insights From Qualitative Research With Detectives in Great Britain and the United States. *Homicide Studies*, *23*(2), 145–174. https://doi.org/10.1177/1088767918793678

Brown, B. (1995). *CCTV in town centres: Three case studies*. Home Office, Police Department London, UK.

Chen, S.-H. (2016). *Application of CCTV in promoting the efficiency of criminal investigation (A Case Study in Taipei County)* [Master, Central Police University]. https://hdl.handle.net/11296/x359b6

Clarke, R. V. (1995). Situational Crime Prevention. *Crime and Justice*, *19*, 91–150. https://doi.org/10.1086/449230

Cook, P. J., & Berglund, A. (2021). More and better video evidence for police investigations of shootings: Chicago's area technology centers. *Policing: An International Journal*, *44*(4), 655–668. https://doi.org/10.1108/PIJPSM-12-2020-0186

Cook, P. J., Braga, A. A., Turchan, B. S., & Barao, L. M. (2019). Why do gun murders have a higher clearance rate than gunshot assaults? *Criminology & Public Policy*, *18*(3), 525–551. https://doi.org/10.1111/1745-9133.12451

Cook, P. J., & Mancik, A. (2024). The Sixty-Year Trajectory of Homicide Clearance Rates: Toward a Better Understanding of the Great Decline. *Annual Review of Criminology*, *7*(1), 59–83. https://doi.org/10.1146/annurev-criminol-022422-122744

Cook, P. J., & Vila, A. (2023). *Gun Violence in Durham, NC, 2017-2021: Investigation and Court Processing of Fatal and Nonfatal Shootings*.

Coupe, T. R. (2016). Evaluating the effects of resources and solvability on burglary detection. *Policing and Society*, *26*(5), 563–587. https://doi.org/10.1080/10439463.2014.989155

Davies, H. J. (2007). Understanding Variations in Murder Clearance Rates: The Influence of the Political Environment. *Homicide Studies*, *11*(2), 133–150. https://doi.org/10.1177/1088767907300751

Ditton, J., & Short, E. (1999). Yes, it works, no it doesn't: Comparing the effects of open-street CCTV in two adjacent Scottish town centres. In K. Painter & N. Tilley (Eds.), *Surveillance of Public Space: Cctv, Street Lighting and Crime Prevention* (Vol. 10, pp. 201–223). Lynne Rienner Publishers.

Dowling, C., Morgan, A., Gannoni, A., & Jorna, P. (2019). How do police use CCTV footage in criminal investigations? *Trends and Issues in Crime and Criminal Justice*, (575), 1–15. https://doi.org/10.3316/informit.336889825164842

Eck, J. E. (1983). *Solving Crimes: The Investigation of Burglary and Robbery*. Police Executive Research Forum.

Fallik, S. W. (2018). Detective effort: What contributes to arrests during retrospective criminal investigations? *Policing and Society*, *28*(9), 1084–1104. https://doi.org/10.1080/10439463.2016.1275625

Ferrandino, J. (2021). Getting Away With Murder: Homicide Clearance by Arrest in Chicago and Its Community Areas. *Homicide Studies*, *25*(2), 164–188. https://doi.org/10.1177/1088767920941563

Gerell, M. (2021). CCTV in deprived neighbourhoods – a short-time follow-up of effects on crime and crime clearance. *Nordic Journal of Criminology*, *22*(2), 221–239. https://doi.org/10.1080/2578983X.2020.1816023

Horvath, F., & Meesig, R. (1996). The Criminal Investigation Process and the Role of Forensic Evidence: A Review of Empirical Findings. *Journal of Forensic Sciences 1972-2005*, *41*(6), 963–969. https://doi.org/10.1520/JFS14032J

Jarvis, J. P., Mancik, A., & Regoeczi, W. C. (2017). Police Responses to Violent Crime: Reconsidering the Mobilization of Law. *Criminal Justice Review*, *42*(1), 5–25. https://doi.org/10.1177/0734016816684198

Jochmans, K. (2023). Peer effects and endogenous social interactions. *Journal of Econometrics*, *235*(2), 1203–1214. https://doi.org/10.1016/j.jeconom.2022.10.004

Jung, Y., & Wheeler, A. P. (2023). The effect of public surveillance cameras on crime clearance rates. *Journal of Experimental Criminology*, *19*(1), 143–164. https://doi.org/10.1007/s11292-021-09477-8

Kerr, O. (2012). The Mosaic Theory of the Fourth Amendment. *Michigan Law Review*, *111*(3), 311–354.

Koper, C. S., & Lum, C. (2019a). Critic: The Limits of Police Technology. In A. A. Braga & D. Weisburd (Eds.), *Police Innovation: Contrasting Perspectives* (2nd ed., pp. 517–543). Cambridge University Press. https://doi.org/10.1017/9781108278423.023

Koper, C. S., & Lum, C. (2019b). The Impacts of Large-Scale License Plate Reader Deployment on Criminal Investigations. *Police Quarterly*, *22*(3), 305–329. https://doi.org/10.1177/1098611119828039

Koper, C. S., Lum, C., & Willis, J. J. (2014). Optimizing the Use of Technology in Policing: Results and Implications from a Multi-Site Study of the Social, Organizational, and Behavioural Aspects of Implementing Police Technologies. *Policing: A Journal of Policy and Practice*, *8*(2), 212–221. https://doi.org/10.1093/police/pau015

Levine, E. S., Tisch, J., Tasso, A., & Joy, M. (2017). The New York City Police Department's Domain Awareness System. *Interfaces*, *47*(1), 70–84. https://doi.org/10.1287/inte.2016.0860

LoFaso, C. A. (2020). Solving Homicides: The Influence of Neighborhood Characteristics and Investigator Caseload. *Criminal Justice Review*, *45*(1), 84–103. https://doi.org/10.1177/0734016819874395

Lum, C., Stoltz, M., Koper, C. S., & Scherer, J. A. (2019). Research on body-worn cameras. *Criminology & Public Policy*, *18*(1), 93–118. https://doi.org/10.1111/1745-9133.12412

Lum, C., Wellford, C., Scott, T. L., Vovak, H., & Scherer, J. A. (2018). *Identifying Effective Investigative Practices: A National Study Using Trajectory Analysis, Case Studies, and Investigative Data* (Final Report to the Laura and John Arnold Foundation). George Mason University.

MacQueen, S., & Bradford, B. (2017). Where did it all go wrong? Implementation failure—and more—in a field experiment of procedural justice policing. *Journal of Experimental Criminology*, *13*(3), 321–345. https://doi.org/10.1007/s11292-016-9278-7

Mastrobuoni, G. (2020). Crime is Terribly Revealing: Information Technology and Police Productivity. *The Review of Economic Studies*, *87*(6), 2727–2753. https://doi.org/10.1093/restud/rdaa009

Mastrorocco, N., & Ornaghi, A. (2025). Who Watches the Watchmen? Local News and Police Behavior in the United States. *American Economic Journal: Economic Policy*, *17*(2), 285–318. https://doi.org/10.1257/pol.20230356

Morgan, A., & Dowling, C. (2019). *Does CCTV help police solve crime?* Australian Institute of Criminology. https://doi.org/10.52922/ti09166

Nagin, D. S. (2013). Deterrence in the Twenty-First Century. *Crime and Justice*, *42*(1), 199–263. https://doi.org/10.1086/670398

Norris, C., & Armstrong, G. (2020). *The Maximum Surveillance Society: The Rise of CCTV*. Routledge. https://doi.org/10.4324/9781003136439

Norton, E. C. (2022). The inverse hyperbolic sine transformation and retransformed marginal effects. *The Stata Journal: Promoting Communications on Statistics and Stata*, *22*(3), 702–712. https://doi.org/10.1177/1536867X221124553

Olphin, T. W. L., Bandyopadhyay, S., & Banerjee, A. (2021). Solving Burglary Offences: Building a Model to Predict Clearance of Burglary Following Initial Investigation. *Policing: A Journal of Policy and Practice*, *15*(2), 832–847. https://doi.org/10.1093/police/paz002

Peterson, J. L., Hickman, M. J., Strom, K. J., & Johnson, D. J. (2013). Effect of Forensic Evidence on Criminal Justice Case Processing. *Journal of Forensic Sciences*, *58*(s1). https://doi.org/10.1111/1556-4029.12020

Piza, E. L., Arietti, R. A., Carter, J. G., & Mohler, G. O. (2025). The effect of gunshot detection technology on evidence collection and case clearance in Kansas City, Missouri. *Journal of Experimental Criminology*, *21*(1), 127–152. https://doi.org/10.1007/s11292-023-09594-6

Piza, E. L., Caplan, J. M., & Kennedy, L. W. (2014). Is the Punishment More Certain? An Analysis of CCTV Detections and Enforcement. *Justice Quarterly*, *31*(6), 1015–1043. https://doi.org/10.1080/07418825.2012.723034

Piza, E. L., Connealy, N. T., Arietti, R. A., Mohler, G. O., & Carter, J. G. (2025). The relative and joint effects of gunshot detection technology and video surveillance cameras on case clearance in Chicago. *Criminology & Public Policy*, 1745-9133.70001. https://doi.org/10.1111/1745-9133.70001

Piza, E. L., Welsh, B. C., Farrington, D. P., & Thomas, A. L. (2019). CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis. *Criminology & Public Policy*, *18*(1), 135–159. https://doi.org/10.1111/1745-9133.12419

Pizarro, J. M., Terrill, W., & LoFaso, C. A. (2020). The Impact of Investigation Strategies and Tactics on Homicide Clearance. *Homicide Studies*, *24*(1), 3–24. https://doi.org/10.1177/1088767918816741

Prince, H., Lum, C., & Koper, C. S. (2021). Effective police investigative practices: An evidence-assessment of the research. *Policing: An International Journal*, *44*(4), 683–707. https://doi.org/10.1108/PIJPSM-04-2021-0054

Roberts, A., & Roberts, J. M. (2016). Crime Clearance and Temporal Variation in Police Investigative Workload: Evidence from National Incident-Based Reporting System (NIBRS) Data. *Journal of Quantitative Criminology*, *32*(4), 651–674. https://doi.org/10.1007/s10940-015-9270-9

Robin, L., Peterson, B. E., & Lawrence, D. S. (2021). How do close-circuit television cameras impact crimes and clearances? An evaluation of the Milwaukee police department's public surveillance system. *Police Practice and Research*, *22*(2), 1171–1190. https://doi.org/10.1080/15614263.2020.1772783

Sheng, H., Yao, K., & Goel, S. (2021). *Surveilling Surveillance: Estimating the Prevalence of Surveillance Cameras with Street View Data* (arXiv:2105.01764). arXiv. https://doi.org/10.48550/arXiv.2105.01764

Terza, J. V. (2017). Two-stage Residual Inclusion Estimation: A Practitioners Guide to Stata Implementation. *The Stata Journal: Promoting Communications on Statistics and Stata*, *17*(4), 916–938. https://doi.org/10.1177/1536867X1801700409

Thomas, A. L. (2025). Evaluating the effect of CCTV on case clearances in Fayetteville, NC: A case for entropy balancing. *Journal of Experimental Criminology*. https://doi.org/10.1007/s11292-025-09703-7

Ungerleider, N. (2012, August 8). *NYPD, Microsoft Launch All-Seeing "Domain Awareness System" With Real-Time CCTV, License Plate Monitoring [Updated]*. Fast Company. https://www.fastcompany.com/3000272/nypd-microsoft-launch-all-seeing-domain-awareness-system-real-time-cctv-license-plate-monito

Vaughn, P. E. (2020). The effects of devaluation and solvability on crime clearance. *Journal of Criminal Justice*, *68*, 101657. https://doi.org/10.1016/j.jcrimjus.2020.101657

Weisburd, D., Wooditch, A., Weisburd, S., & Yang, S.-M. (2016). Do Stop, Question, and Frisk Practices Deter Crime? *Criminology & Public Policy*, *15*(1), 31–56. https://doi.org/10.1111/1745-9133.12172

Welsh, B. C., & Farrington, D. P. (2008). Effects of Closed Circuit Television Surveillance on Crime. *Campbell Systematic Reviews*, *4*(1), 1–73. https://doi.org/10.4073/csr.2008.17

Welsh, B. C., & Farrington, D. P. (2009). Public Area CCTV and Crime Prevention: An Updated Systematic Review and Meta-Analysis. *Justice Quarterly*, *26*(4), 716–745. https://doi.org/10.1080/07418820802506206

Wu, A., Kuang, K., Xiong, R., & Wu, F. (2025). Instrumental Variables in Causal Inference and Machine Learning: A Survey. *ACM Computing Surveys*, *57*(11), 1–36. https://doi.org/10.1145/3735969

Yeung, B., ProPublica, Greenblatt, M., Fahey, M., Newsy, Harris, E., & Reporting, R. from T. C. for I. (2018, November 15). When It Comes to Rape, Just Because a Case Is Cleared Doesn't Mean It's Solved. *ProPublica*. https://www.propublica.org/article/when-it-comes-to-rape-just-because-a-case-is-cleared-does-not-mean-solved

6. Appendix: Regression Kink Design (RKD) with Domain Awareness System Expansion Announcement Date

One explanation for the NYPD's uptick in video evidence collection in gunshot investigations is the better integration of existing CCTV via the Domain Awareness System (DAS) at the end of 2012. The system was originally developed in 2008 to integrate intelligence and footage from public and private CCTV units for counterterrorism purposes in lower Manhattan (Levine et al., 2017). After recognizing DAS's potential for general policing in late 2012, the system expanded to connect all five boroughs' CCTVs and became a centralized repository of camera footage for investigators (Levine et al., 2017; Parascandola & Moore, 2012). Conversations with a former NYPD manager suggest that DAS substantially reduced the cost of searching for video evidence and created institutional pressure on detectives to retrieve available footage, as supervisors could check whether video was available in the area of an incident by using past cases in the system.

To test whether the announcement of DAS expansion caused a change in the original growth trajectory of video evidence usage, we conduct a nonparametric regression kink design (RKD) using August 8, 2012, the date of the city's public announcement of DAS's expansion (see Parascandola & Moore, 2012; Ungerleider, 2012), as a cutoff. The pool regression of the RKD can be written as

$$Y_i = \alpha + \beta_1 \tilde{T}_i + \tau \cdot \tilde{T}_i \cdot D_i + \sum_{j=2}^{p} (\gamma_j \tilde{T}_i^{\,j} + \delta_j \tilde{T}_i^{\,j} D_i) + \varepsilon_i,$$

$$\tilde{T}_i = T_i - c.$$

Y_i is the video evidence usage in case i. $\tilde{T}_i$ is the day since the expansion for case i. τ is the immediate change in the trend in video evidence usage right after the expansion. D_i indicates that case i happened after August 8, 2012, denoted as c. $\sum_{j=2}^{p} (\gamma_j \tilde{T}_i^{\,j} + \delta_j \tilde{T}_i^{\,j} D_i)$ is the local polynomial to control for the nonlinear trend of video evidence usage before and after the expansion of the DAS system.

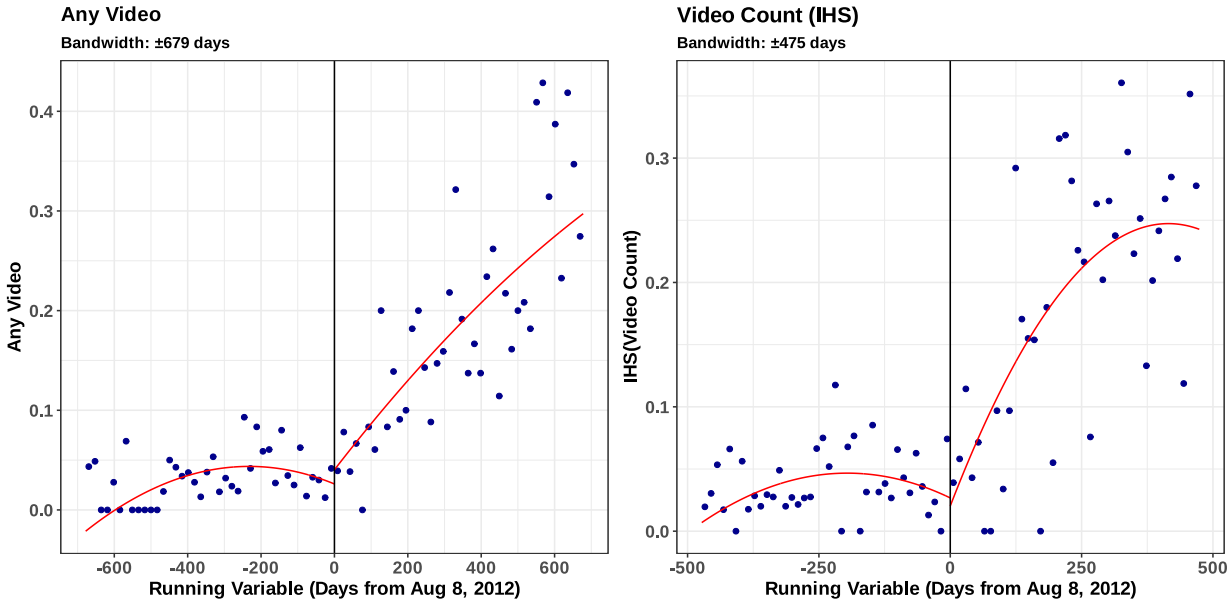
Estimation used triangular kernels, with optimal bandwidth and local polynomial order selected via the rdrobust package in R (Calonico et al., 2015). The results indicate that the expansion induced an abrupt increase in the slope for both the usage rate ($\beta = 0.0007, p = 0.009$) and the inverse hyperbolic sine (IHS) transformed number of videos attached per case ($\beta = 0.0016, p = 0.014$), while the level (intercept) remained unchanged (Figure A1). The absence of a direct level shift is consistent with a gradual rollout of DAS in the city rather than an immediate overhaul.

We present the estimate for video count from the IHS-transformed model in the text because it provides a better fit than the count-level model. For transparency, the count-level model indicates a marginally significant increase in the slope for video amount by day ($\beta = 0.0024, p = 0.057, \mathrm{z} = 1.905$). The selected bandwidth remains consistent with that of the IHS-transformed model.

As a first-order approximation, the RKD models suggest that the probability of using video evidence increased by 0.05 percentage points, and the number of videos used increased by approximately 0.13% per day, since the expansion was announced. This pattern suggests that the

change in evidence practices in NYPD criminal gunshot investigations was not driven solely by the number of technological hardware but was also catalyzed by department policy and system design to realize the technology's potential (Koper et al., 2014).

Figure A1: Regression Kink Design on Video Evidence Usage.



Note: The blue dots represent the 40-day binned average, while the red curves depict a quadratic local polynomial regression weighted by a triangular kernel around the cutoff. The polynomial order for these fitted lines was chosen using the rdrobust package.

Unfortunately, the robust z-statistics in the RKDs are only 2.62 for any video evidence and 2.47 for IHS-transformed video count, both lower than the common 3.15 threshold for a strong instrument (Keane & Neal, 2024). This is unsurprising given that RKD is inherently less statistically powered, even less than the regression discontinuity design, as it aims at detecting higher-order changes in trend. This limitation is exacerbated when first-stage uncertainty is propagated into the second-stage estimate within the instrumental variable framework. Moreover, because the RKD focuses only on a transient period of change, it limits our understanding of the extent to which video use in criminal gunshot investigations contributes to arrest rates in Figure 1.

These limitations motivated our alternative empirical strategy in the main text, which better exploits localized variation in non-selective video evidence practices: spatiotemporal instruments that capture the timing and location at which video evidence is more commonly or intensively used across cases, regardless of case attributes.

7. Appendix: Comparing Nonlinear Modelings of the Dosage Effect from Video Count

In Section 4.4 of the main text, we aim to compare the dosage model with the IHS-transformed video count using spline regressions that allow a more flexible functional form for video count. However, the typical two-stage least squares framework cannot accommodate an endogenous variable that enters multiple basis functions. To keep the comparison on a common footing, we reestimate Model 18 in the main text, together with the other two spline regressions, by the two-stage residual inclusion (2SRI; see the main text for more details about this method) procedure:

$$Arrest_i = \theta\, IHS(VideoCount_i) + \gamma\zeta_i + \eta_2\, ClusterArrest_{\%i} + \boldsymbol{X_i}'\boldsymbol{\beta} + \tau_i + \epsilon_{i.} \quad (1)$$

$$Arrest_i = \sum_{m=1}^{6} \rho_m\, B_m(VideoCount_i) + \gamma\zeta_i + \eta_2\, ClusterArrest_{\%i} + \boldsymbol{X_i}'\boldsymbol{\beta} + \tau_i + \epsilon_{i.} \quad (2)$$

$$Arrest_i = \sum_{m=1}^{6} \rho_m\, B_m\big(IHS(VideoCount_i)\big) + \gamma\zeta_i + \eta_2\, ClusterArrest_{\%i} + \boldsymbol{X_i}'\boldsymbol{\beta} + \tau_i + \epsilon_{i.} \quad (3)$$

$B_m(\cdot)$ denotes the natural cubic spline basis with interior knots at counts of 1, 5, 10, 15, and 25 videos, yielding six basis functions ($m = 1, \dots, 6$); in Equation 3 the knots are placed on the transformed scale, i.e., at $IHS(1), \dots, IHS(25)$. The ρ_m are the associated basis coefficients, which individually have no direct interpretation and are best understood through marginal-effects analysis. ζ_i is the first-stage residual from the 2SRI procedure, formed by regressing each endogenous treatment on its linear instrument: $VideoCount_i$ on $KNNIV_{\#i}$ in Equation 1, and $IHS(VideoCount_i)$ on $IHS(KNNIV_{\#i})$ in Equations 2 and 3. We compare the three models using the adjusted R^2, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). The locations of the interior knots and their numbers were also selected based on these three criteria.

Equation 1 provides the baseline, with adjusted $R^2 = 0.1032$, AIC = 24,604, and BIC = 28,248. Equation 2 using a spline of the raw count fits worst (adjusted $R^2 = 0.1028$, AIC = 24,616, BIC = 28,292). Equation 3 using a spline of the IHS-transformed count attains the highest adjusted R^2 (0.1034) but higher AIC (24,605) and BIC (28,280) than the baseline, suggesting a marginally better in-sample fit at the cost of added complexity and potentially weaker generalizability to other data. But given Equation 3's slightly higher adjusted R^2, we still plot the estimates in Figure 8 of the main text.

8. Appendix: CCTV Camera Detections with Google Street View

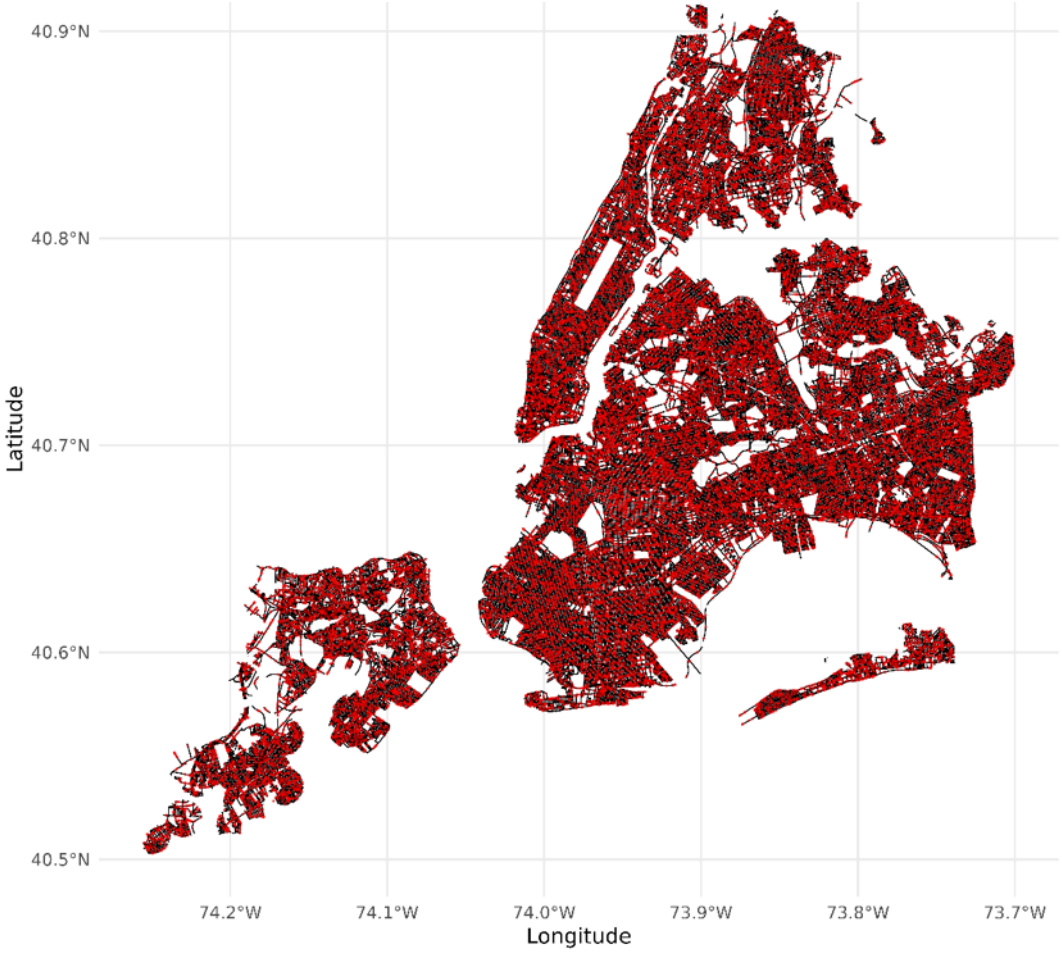
Google Street View, which provides access to both contemporary and historical imagery for the vast majority of streets and intersections in the U.S., represents an unprecedented repository of information on features of the built environment, including security cameras and other security-related fixtures (e.g., street lights, window bars, security gates, fences, etc.). However, processing large volumes of Street View imagery via traditional systematic social observation (SSO) methods is challenging to implement quickly and inexpensively. Automated extraction and feature detection to produce spatiotemporally granular estimates of feature prevalence represents an alternative methodology.

8.1 Street View Sampling, Collection, and Camera Detection Procedure

To reduce computational and API usage costs, we focus only on areas within 1 km of all shooting incidents in our dataset. The 1 km buffer areas were created for all shooting incidents and merged into a single area using the st_union function from the sf package. Sampling was conducted on street grids where Google's panorama is available. The street centerline data used for the location sampling is from "DCM_StreetCenterLine" on NYC OpenData. The street lines were "scooped" by the merged buffer and serve as the basis for our random location sampling. The scooped area is about 690.73 km^2.

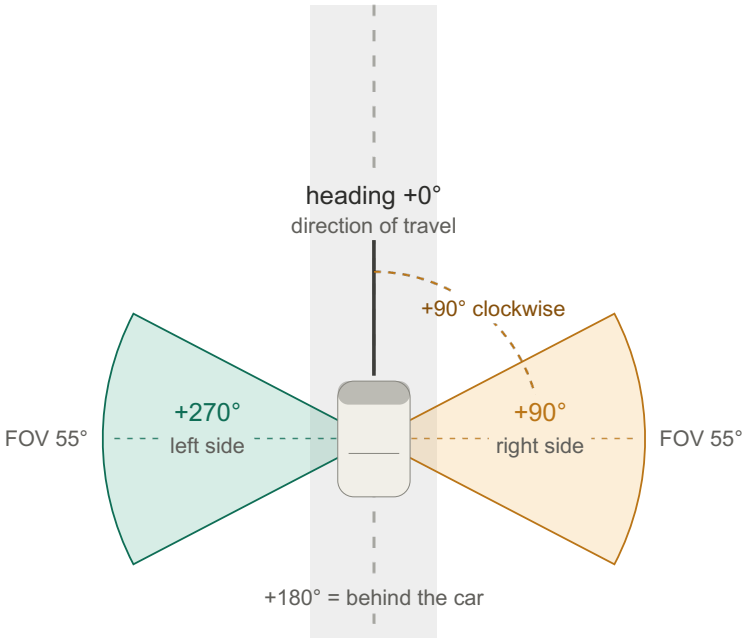
The sampling density is set to be approximately 128 points per 1 km buffer (3.14 $km^{2)}$. Therefore, the predefined sample size is $690.73 \div 7.376 \times 300 \approx 28{,}094$ points. The coordinates of these sampled points were then queried using the Street View package in Python to retrieve metadata on the coordinates, ID, recorded time (year and month), and heading (the moving direction of the Street View car) for the panorama. Note that the location of the panoramas obtained from the query will not be the same as the original queried coordinate (mathematically impossible). To ensure the panoramas were relevant to the original query, they were limited to those within 5 meters of the median distance from all responses to the query. Those that were not attached to a clear date were also excluded. These restrictions resulted in only 26,918 sampled locations with any panoramas obtained. From the 26,918 sampled locations, 189,805 panoramas were obtained, meaning that, on average, one location query yields 7.05 panoramas photographed at different times. The mean distance from the actual panorama location to the queried location was 4.7 meters. Figure A2 plots the street lines that we sampled from and the locations sampled.

Figure A2: Randomly Sampled Locations.



Note: The plot shows the street grids on which the samples were based. Each red dot is a sampled location. The missing streets are those without criminal gunshots within 1 kilometer from 2006 to March 2023.

Figure A3: The Static Street View API Parameters.



Note: The plot was made with Claude Opus 4.8.

The actual street view pictures were acquired through Google's Street View Static API (https://developers.google.com/maps/documentation/streetview). The API limits the response image to be 640 by 640 pixels, which is not, by definition, a "panorama". To save on API calls, we set the heading to 90 (right side) and 270 (left side) to obtain pictures of the roadside. The field of view (FOV) was set to 55 degrees to reduce overlap when sampling a nearby location. The smaller the FOV, the more zoomed-in the image is (see Figure A3 for an illustration). This is also a scope that we consider sufficient for the computer vision algorithm to perform meaningful detection on a 640 by 640 pixel image. In total, $189{,}805 \times 2 = 379{,}610$ images were acquired.

The images were then fed into the open-source zero-shot object detection algorithm Owlv2 (https://huggingface.co/google/owlv2-base-patch16-ensemble) along with a text prompt. Owlv2 identifies objects based on the input text prompt. The prompt chosen for our detection task is "surveillance camera" due to its better performance in the pilot trials than the other prompts, including "camera", "cameras", "cctv", "cctv camera", "cctv cameras", "cam", "surveillance cameras", and "security camera". The model outputs the bounding box (four sets of x- and y-coordinates) where it detects a CCTV camera, along with a certainty score.

This detection process was partitioned into two rounds. The first round used a 0.15 certainty score, a low bar that identifies all the possible objects that could be a CCTV. The second round zoomed in on the box by adjusting the heading, FOV, and camera pitch to focus on the object, then performed another round of API calls to obtain the street-view image for each potential object. This zoom-in process could provide a clearer picture of the object and, thus, higher detection accuracy, given that, no matter how high the original panorama's resolution is, the picture the API provides is always 640 by 640 pixels. In this second round, the certainty score threshold was raised to 0.25 to be counted as a camera.

Figure A4: Two Examples of CCTV Detection from Street View Images.



The described process was conducted in batches using four computing nodes, which collectively utilized 28 graphics processing units (GPUs) in parallel on a high-performance computing (HPC) server. The entire operation took approximately five days to complete. All certainty scores for the detections were saved, and some examples of the detection results are shown in Figure A4. The red boxes indicate the areas where the model identifies a surveillance camera. Given the small size of these cameras and the presence of many other objects in the images, we believe the detection algorithm is sufficiently sensitive to identify CCTV cameras.

8.2 Limitations

The limitations of this procedure for measuring changes in camera density over time should be made clear. They can be categorized into three sources: the detection model, the captured street-view images, and Google's panorama update policy.

Firstly, we found that the Owlv2 model sometimes identifies street lights, hollow poles, floodlights, or external exhaust ports as CCTV cameras, especially when these objects are distant or captured from the side. Unfortunately, we did not have enough resources to conduct human verification for all the identified cameras from 379,610 images. Nonetheless, assuming these other objects do not undergo a citywide systematic change over time, and that the algorithm, in general, still has greater specificity for CCTV, changes in the detected cameras over time can still be informative about the adoption *trend*.

Second, CCTV cameras may be obscured by uncontrollable objects in the acquired images, leading to an undercount of the actual number of cameras. The most common form of blockage is from taller vehicles, such as vans, buses, and trucks, as well as trees and sidewalk sheds. Figure A5 provides two examples of lights being detected as cameras. Besides, privacy-concerned individuals could also request that Google blur their home or vehicle, which can also block CCTV cameras attached to the building (https://support.google.com/maps/answer/15439776). Figure A6 shows an example in which the right area is completely blurred and only one camera is detected.

We checked the same location on Google Maps from another year without the blurring patch and found another camera in the blurred area, symmetrical to the detected camera's position, with the balcony as the center.

Figure A5: Two Examples of Failed CCTV Detections.



Another challenge comes from the image's improving resolution over time, which could lead to more CCTV cameras being detected, even though they may have already been at the same spot but not detected due to poor image quality. As we mentioned in the main text, we suspect the low detection rate for 2007 is due to poor panorama quality, not solely to lower CCTV density (see

Figure A7). For instance, Figure A8 shows a street view image taken in 2007. The upper-right corner of the building appears to have a white object attached, which is very likely a CCTV camera. However, the resolution is too low, and zooming in or using different angles does not help, even for human eyes. Based on the 2011 photos, this appears to be a camera at the same location, though the CCTV cameras seem to have been replaced with a different type (black dome).

Nonetheless, if this change in resolution is an unaccounted-for factor that leads to an overestimation of the trend, it provides even stronger evidence supporting our claim that the rapid growth in the use of rapid video evidence in gunshot investigations cannot be attributed to the rather mild change in street camera density.

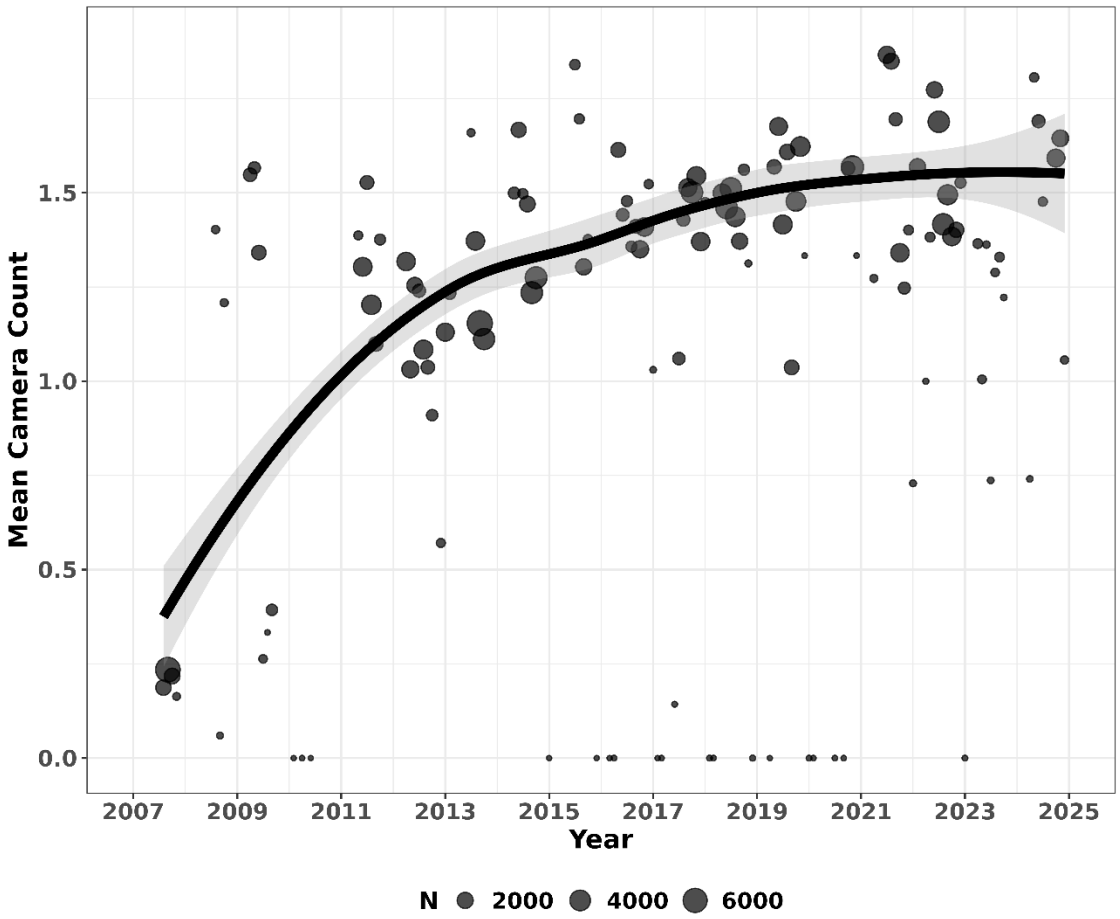
Lastly, although our location sampling was random, Google's sampling of the street views is not. The frequency of panorama collection at the same queried location is also neither fixed nor random, and it varies even across New York City. Some places may be over- or under-sampled relative to others in our trend estimation, and the geographical composition of each year-month estimate shown in Figure 1 in the main text varies substantially. These limit our ability to trace an unbiased change in camera density.

Figure A6: An Example of a Blurred Street View Image.



Note: The image was collected at 40.75984°, -73.99195° in July, 2017. One dome CCTV camera is detected (marked in red). The blurring patch could be the result of a privacy request, in which we find suggestive evidence of another camera through manual checks on Google Maps.

Figure A7: Mean Camera Detected Per Queried Location from 2007–2024.



Note: This is a non-truncated version of Figure 1 in the main text. It shows the most dramatic change between 2007 and 2011 (the next wave of Google's panorama collection). We believe this is an artifact of image quality improvement between the two time points.

Figure A8: An image collected in 2007.



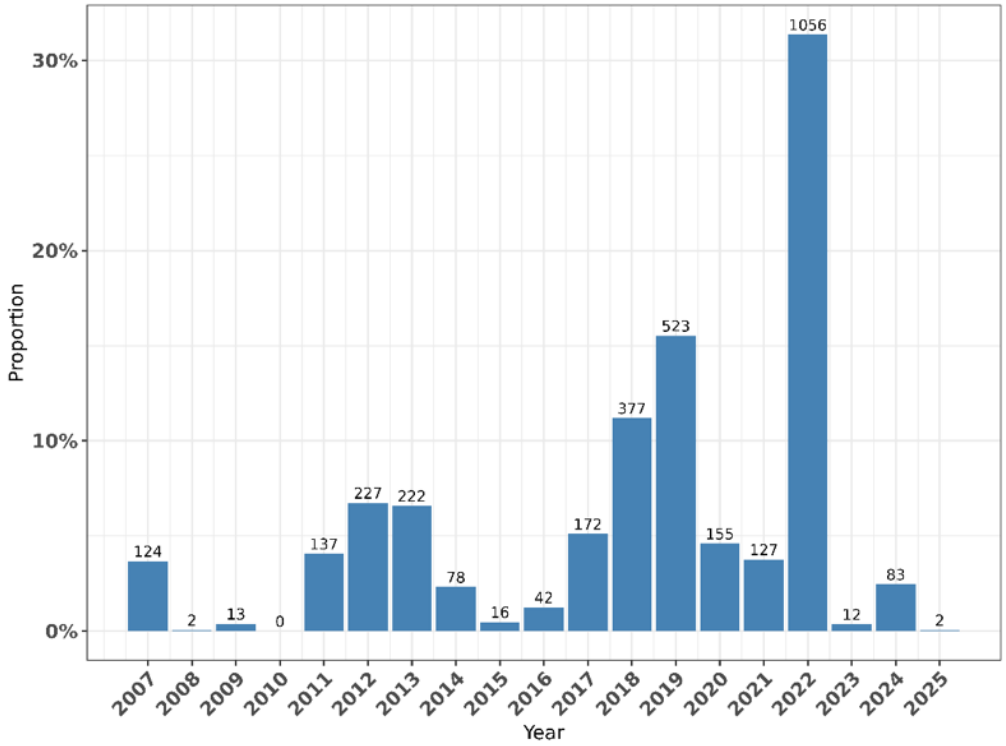
Note: The image was taken at 40.85359°, -73.90407° on July 15, 2007. The red circle was manually added to the photo where we suspect a CCTV camera is located, but we cannot confirm due to poor image quality. Note that the watermark on the image, "©2019 Google", does not mean that the Street View image was taken in 2019.

To illustrate, we quantify the frequency of panorama updates based on our queried coordinates. 12.5% of locations in our samples with at least one panorama had never received an update by August 2025.

Figure A9 shows that over 30% of these single-panorama locations were taken in 2022. More than 10% were taken in 2018 and 2019. For these locations, we cannot trace changes in the camera, given that there are no images of the same location over time.

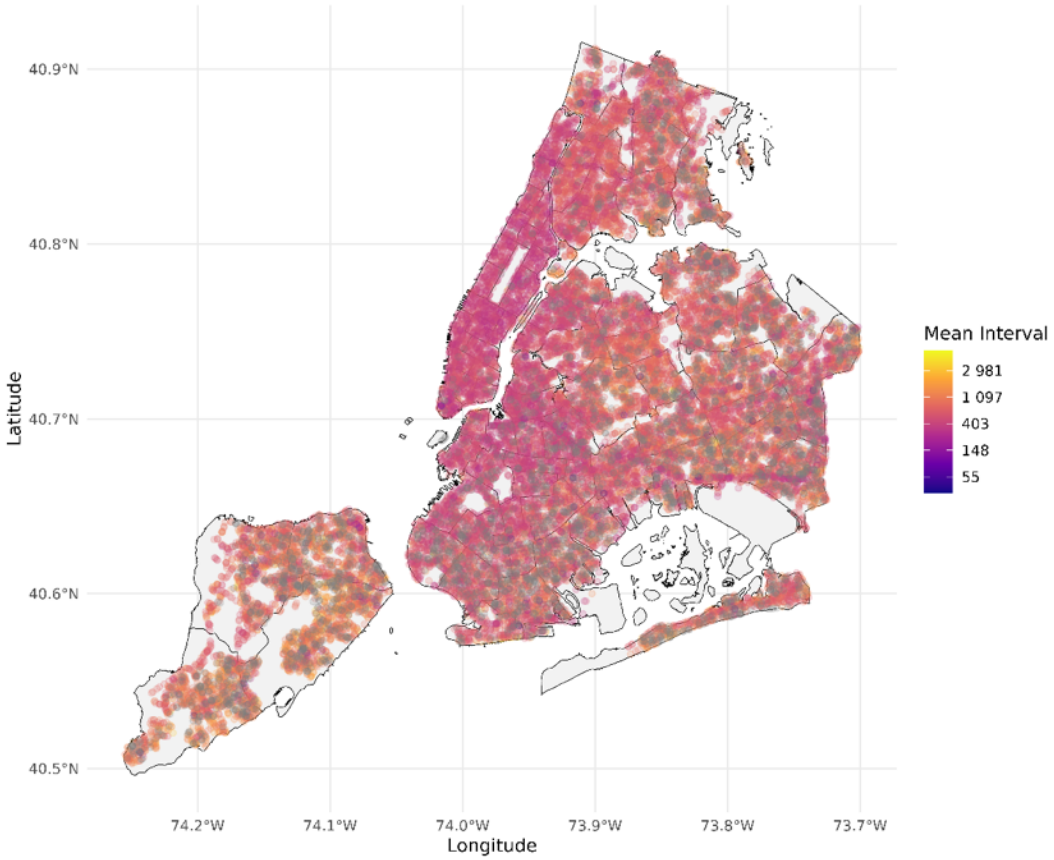
The rest have mean and median panorama update intervals of 23 and 17 months, respectively, suggesting that Street View updates usually take more than one year. Using mean update intervals by day, Figure A10 shows that some places have more frequent panorama updates than others. Manhattan apparently has the most frequent panorama updates, as do the waterfronts of Brooklyn and Queens. A traffic artery like the Belt Parkway, which connects southern Brooklyn and Queens, also receives more frequent updates, as shown by the purple vein in the plot. Figure A11 plots the distribution of the panorama collected by year. The coverage rates for Manhattan, Brooklyn, and the southeast corner of Queens were low in 2007. Fuller coverage of Manhattan appears only after 2009, after 2011 for Brooklyn, and after 2012 for the Southeast corner of Queens, while many other locations already had two or three rounds of panorama collections (Manhattan).

Figure A9: Proportion of Queried Locations Without Updates.



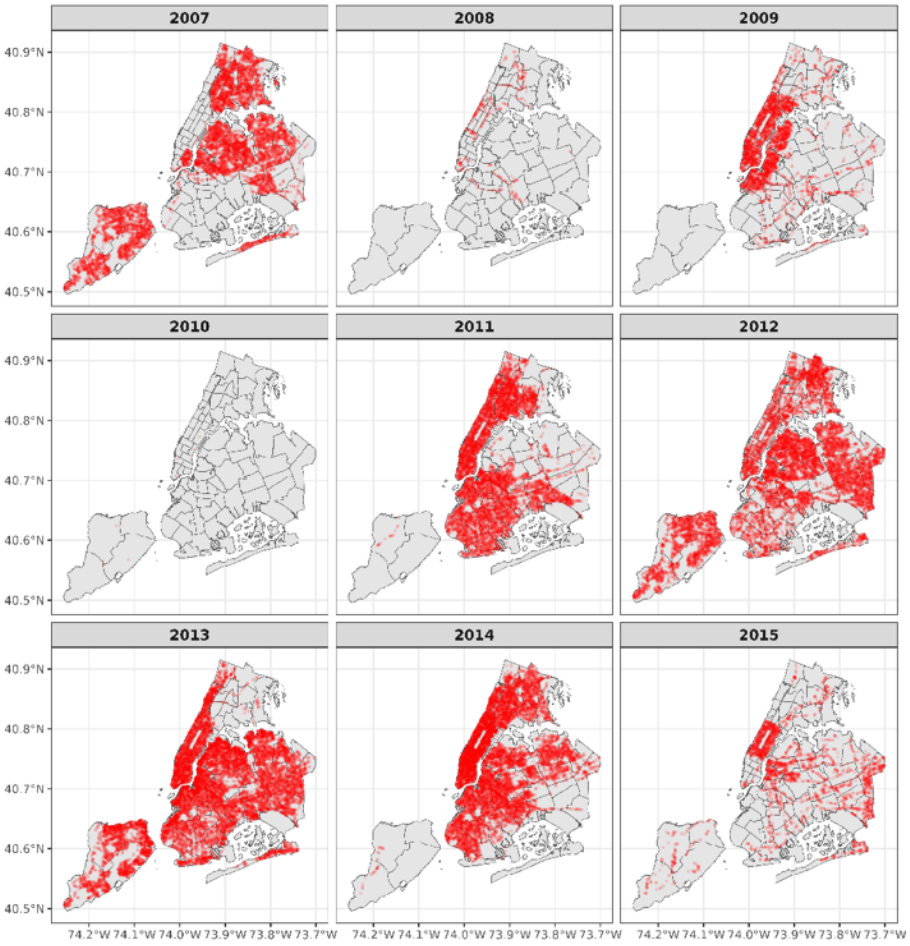
Note: The number above each bar is the number of locations queried.

Figure A10: Mean Update Intervals by Day.



Note: Each dot is a queried location with at least two panoramas collected at different times, suggesting at least one update. The legend shows the mean update interval for the location by day using the 15th of each month. The color scheme is log-transformed. The reddish-purple areas have more frequent Street View updates.

Figure A11: Panoramas Collected from 2007–2015.



Note: The red dots are the queried locations where street view images of that specific year are available.

8.3 Summary

Overall, we want to stress that identifying the change of cameras from street view images through computer vision is an extremely challenging task, given that the sizes of the CCTV cameras only take a fraction of the whole panorama and that these street view images were not originally collected by Google for our research purposes, but only for an overview of the street.

To provide one last example of the complexity, Figure A12 shows an image with six objects that were identified as CCTV cameras by Owlv2. We find only the second and sixth (from right to left) to be actual cameras after manually viewing Google Maps and validating panoramas from different angles and times. The others are either floodlights or ornaments that are shaped like CCTV cameras. By looking at the original 640 by 640-pixel picture, we find that even humans have a hard time differentiating these objects. Moreover, while validating the detected objects on Google Maps, we found another camera behind the awning and before the sixth detected object. There is no empirical evidence to determine whether these cameras, farther from the street, have any investigative value for shooting investigations.

Figure A12: An Image Mingled with Wrong, Correct, and Missing Detections.



Note: The image was taken at 40.70255°, -73.81349° in November, 2016. Note that the watermark on the image, "©2022 Google", does not mean that the street view was taken in 2022.

Future research can develop a more accurate CCTV camera detection algorithm than the off-the-shelf zero-shot model we used in our study, or incorporate a human verification process to improve detection specificity. Still, we doubt that the challenges posed by image quality and the nonrandom geospatial sampling in Google Street View can be fundamentally solved. Even with these limitations, we still believe that Google Street View is the best available data source for researchers to approximate trends in CCTV camera density on the street over the past two decades.

Therefore, this exercise serves only as a basic overview of CCTV camera trends in the city. We do not recommend placing too much emphasis on the exact number detected by the model or using this data at a precise coordinate level to analyze the relationship between CCTV location and crime clearance. The low time resolution of Street View images, uncontrollable noise (e.g., blocking objects, blurring patches, changing image quality), and the uneven update frequency would significantly compromise researchers' ability to draw meaningful conclusions.