

Evidence that curtailing proactive policing can reduce major crime

Christopher M. Sullivan^{1*} and Zachary P. O’Keeffe²

Governments employ police to prevent criminal acts. But it remains in dispute whether high rates of police stops, criminal summonses and aggressive low-level arrests reduce serious crime^{1–7}. Police officers target their efforts at areas where crime is anticipated and/or where they expect enforcement will be most effective. Simultaneously, citizens decide to comply with the law or commit crime partly on the basis of police deployment and enforcement strategies. In other words, policing and crime are endogenous to unobservable strategic interaction, which frustrates causal analysis. Here, we resolve these challenges and present evidence that proactive policing—which involves systematic and aggressive enforcement of low-level violations—is positively related to reports of major crime. We examine a political shock that caused the New York Police Department (NYPD) to effectively halt proactive policing in late 2014 and early 2015. Analysing several years of unique data obtained from the NYPD, we find that civilian complaints of major crimes (such as burglary, felony assault and grand larceny) decreased during and shortly after sharp reductions in proactive policing. The results challenge prevailing scholarship as well as conventional wisdom on authority and legal compliance, as they imply that aggressively enforcing minor legal statutes incites more severe criminal acts.

In the last few decades, proactive policing has become a centerpiece of ‘new policing’ strategies across the globe^{8,9}. The logic, commonly associated with the broader theory of order maintenance policing (OMP; also known as broken windows), is that rather than wait for citizens to report criminal conduct, law enforcement should proactively patrol communities, maintaining order through systematic and aggressive low-level policing^{10,11}. According to proponents, increasing police stops, quality-of-life summonses, and low-level arrests deters more serious criminal activity by signalling that the area is being monitored and that deviance will not be tolerated^{12,13}. As a corollary, following a phenomenon termed the Ferguson effect, disengaging from proactive policing emboldens criminals, precipitating spikes in serious crime¹⁴.

But while elected officials commonly justify proactive policing by pointing to the enforcement of legal statutes, the strategy’s efficacy continues to be debated^{5,15,16}. A serious concern is that proactive policing diverts finite resources and attention away from investigative units, including detectives working to track down serial offenders and break up criminal networks^{8,17}. Proactive policing also disrupts communal life, which can drain social control of group-level violence¹⁸. Citizens are arrested, unauthorized markets are disrupted, and people lose their jobs, all of which create more localized stress on individuals already living on the edge^{19,20}. Such strains are imposed directly through proactive policing, and thus are independent from subsequent judgments of guilt or innocence²¹. Inconsistency in aggressive low-level policing across community

groups undermines police legitimacy, which erodes cooperation with law enforcement^{11,20}. The cumulative effect increases ‘legal cynicism’—individual reliance on extra-legal sanctions and informal institutions of violence as a replacement for police^{22,23}. Reflecting these mechanisms, we propose that sharply reducing proactive policing in areas where it had been deployed pervasively may actually improve compliance with legal authority, thereby reducing major crimes.

To assess these claims, our study analyses an aberration in NYPD strategy, in which police sharply limited foot patrols, criminal summonses and low-level arrests in a manner unrelated to the city’s underlying crime rate. In the midst of a political fight between Mayor de Blasio, anti-police brutality protesters and the city’s police unions, the NYPD held a work ‘slowdown’ for approximately seven weeks in late 2014 and early 2015. Within New York City (NYC), the most proximate cause of protests against the NYPD was the strangling death of Eric Garner in Staten Island. While there was considerable fallout from the incident itself, the conflict intensified when a grand jury declined to indict the involved officers on 4 December 2014. Thousands of protesters marched across the Brooklyn Bridge, while others blocked portions of the West Side Highway as well as the Lincoln and Holland tunnels. Then, two weeks after the non-indictment decision, two NYPD officers, Wenjian Liu and Rafael Ramos, were fatally shot by an anti-police extremist. Because they are legally prohibited from striking, NYPD officers coordinated a work-to-rule strike. Officers were ordered to respond to calls only in pairs, leave their squad cars only if they felt compelled, and perform only the most necessary duties. The act was a symbolic show of strength to demonstrate the city’s dependence on the NYPD. Officers continued to respond to community calls for service, but refrained from proactive policing by refusing to get out of their vehicles to issue summonses or arrest people for petit crimes and misdemeanours.

Emblematic of the slowdown’s effects (and the change from proactive to responsive policing), zero summonses were issued for quality-of-life violations on New Year’s Eve 2014, while just the week before, two officers were fatally shot responding to a reported robbery. Eventually, under pressure from the media as well as growing demands for city revenue, Commissioner Bratton conceded to the ‘self-initiated’ slowdown in proactive policing, before publicly ordering his officers to return to work by 16 January.

The change in tactics appears particularly stark when compared with the aggressive strategy of proactive policing the NYPD pursued during the preceding decades. Correspondence between the introduction of proactive policing in New York and the city’s historic drop in major crime has been heralded as *prima facie* evidence of the strategy’s effectiveness¹³. As a result, cities across the globe adopted the NYPD’s protocols and practices, which suggests not only that proactive policing strategies are presumed to deter major

¹Louisiana State University, Baton Rouge, LA 70803, USA. ²University of Michigan, Ann Arbor, MI 48109, USA. *e-mail: csullivan@lsu.edu

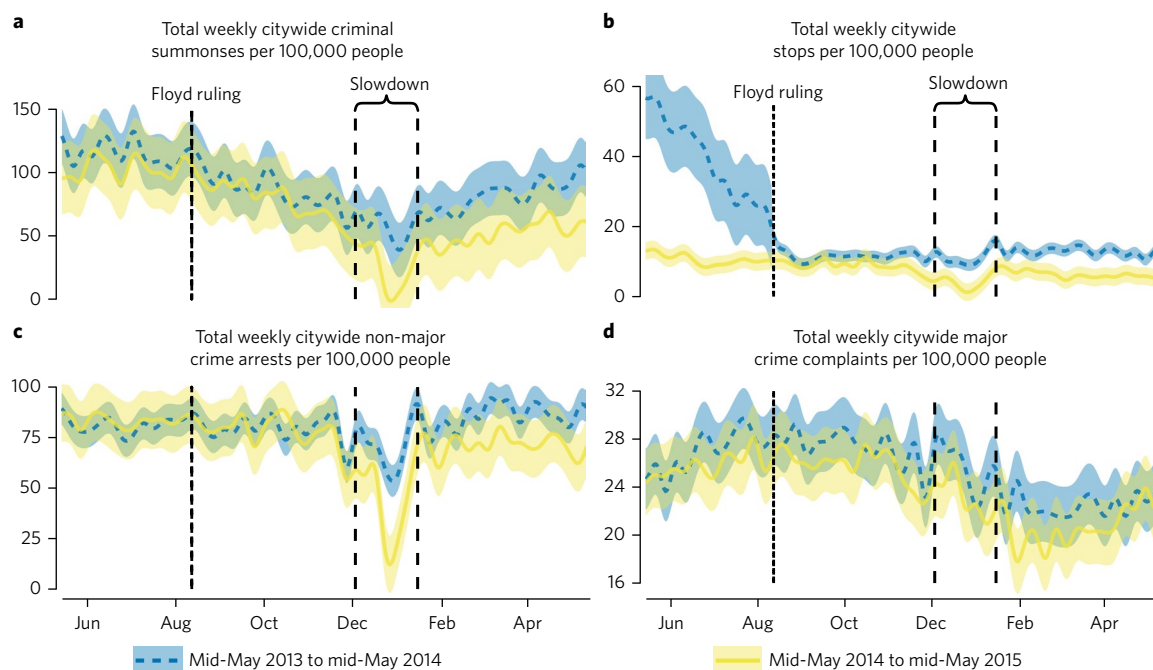


Fig. 1 | Temporal variation in policing and crime complaints in NYC. **a–d**, Graphs showing total weekly citywide activity over time. The titles refer to y axes; the x axis is time; the original unit is one week, but days are plotted. The line colours and types correspond to different series: the dashed blue lines run from 15 May 2013 to 14 May 2014; the solid yellow lines run from 15 May 2014 to 14 May 2015. The blue and yellow lines are from a natural cubic spline fit through all weekly citywide data points (aggregated from 76 precincts), with each week being a knot. Fifty-two knots are plotted per series per model, derived from an original 7,904 precinct-week observations per variable. The long-dashed black lines delineate the NYPD slowdown weeks (1 December to 19 January), which is the primary comparison period of interest between the two series. The short-dashed black lines indicate the calendar day of the ‘Floyd versus City of New York’ ruling, 12 August. The shaded ribbons represent one standard deviation in the variable above and below the interpolated value. For models **a**, **c** and **d**, separate standard deviations are calculated by series (N per series per model = 52). In model **b**, separate standard deviations in per capita stop, question and frisks are calculated for the 13 weeks before, and 39 after, the 12 August 2013 ‘Floyd’ ruling in the first (blue) series, and for all 52 points in the second (yellow) series. Criminal summonses are misdemeanour and summary offences. Major crimes are murder, rape, robbery, felony assault, burglary, grand larceny and grand theft auto; non-major crimes are all other arrestable crimes.

crime in NYC, but also that these policies are widely thought to work in other contexts as well^{5,24}.

If, as would seem to be the case, the slowdown was unrelated to the city’s underlying crime rate, this makes for a unique natural experiment to identify the causal effects of changing police practices. While Garner was being arrested for a misdemeanour offence, and the killings of Liu and Ramos were homicides, these three crimes neither reflect nor predict citywide (nor precinct-wide) crime. And while anti-police brutality protests and the ensuing political conflict were tied to policing practices across the country, it is difficult to argue that the protests were caused by NYC’s crime rate.

To assess the slowdown’s effects, we filed a series of Freedom of Information requests soliciting a comprehensive set of NYPD CompStat reports from 2013–2016 (see Supplementary Fig. 1 for an example). CompStat (short for computer statistics) was introduced in New York as part of a series of reforms to target proactive policing at ‘hotspots’ in which crime was most concentrated^{5,24}. The reports document weekly activity in each NYPD precinct. On the basis of findings from earlier research, we are confident that CompStat data represent the best available source of disaggregated information on police behaviour and crime, and correlate strongly with the underlying reality (see further discussion in the Methods section)^{4,24–26}. Perhaps the best evidence of their validity comes from the fact that the NYPD uses CompStat reports to allocate police resources and develop strategy in real time²⁷.

Examining citywide time series, we find evidence of the timing of the NYPD slowdown, as well as preliminary indications of its effects (Fig. 1). Several policing measures are considered.

‘Criminal summonses’ are charges issued for summary Penal Law Violations (that is, quality-of-life violations, including, most commonly, public consumption of alcohol and disorderly conduct, but not ticketable parking fines or moving violations). ‘Stop, question and frisks’ (SQFs) are temporary street detentions and searches of individuals for contraband. Use of SQFs dropped precipitously to a new baseline in anticipation of the judgment in *Floyd versus City of New York*, which ordered a series of reforms to prevent unconstitutional racial profiling. ‘Non-major crime arrests’ are arrests for all crimes and misdemeanours, excluding the NYPD’s ‘seven major crimes’—murder, rape, robbery, felony assault, burglary, grand larceny and grand theft auto. It includes arrests made by members of the precinct as well as officers from the Transit and Housing Departments and two specialized bureaus: the Organized Crime Control Bureau (OCCB) and the Detective Bureau. According to annual NYPD statistics, misdemeanour arrests represented 92% of all non-major crime arrests in 2014¹².

Our indicator of legal compliance, ‘Major crime complaints’, measures civilian reports of any of the ‘seven major crimes’ indexed by the NYPD. We focus on major crime complaints for several reasons. First, the major premise behind proactive policing is that increasing police stops, criminal summonses and low-level arrests will prevent these types of major crime. As expressed by two of proactive policing’s chief architects, ‘A neighbourhood where minor offenses go unchallenged soon becomes a breeding ground for more serious criminal activity and, ultimately, for violence’¹³. Second, the NYPD pays particular attention to these offences and tracks them consistently across time and space²⁴. Indicative of the measure’s

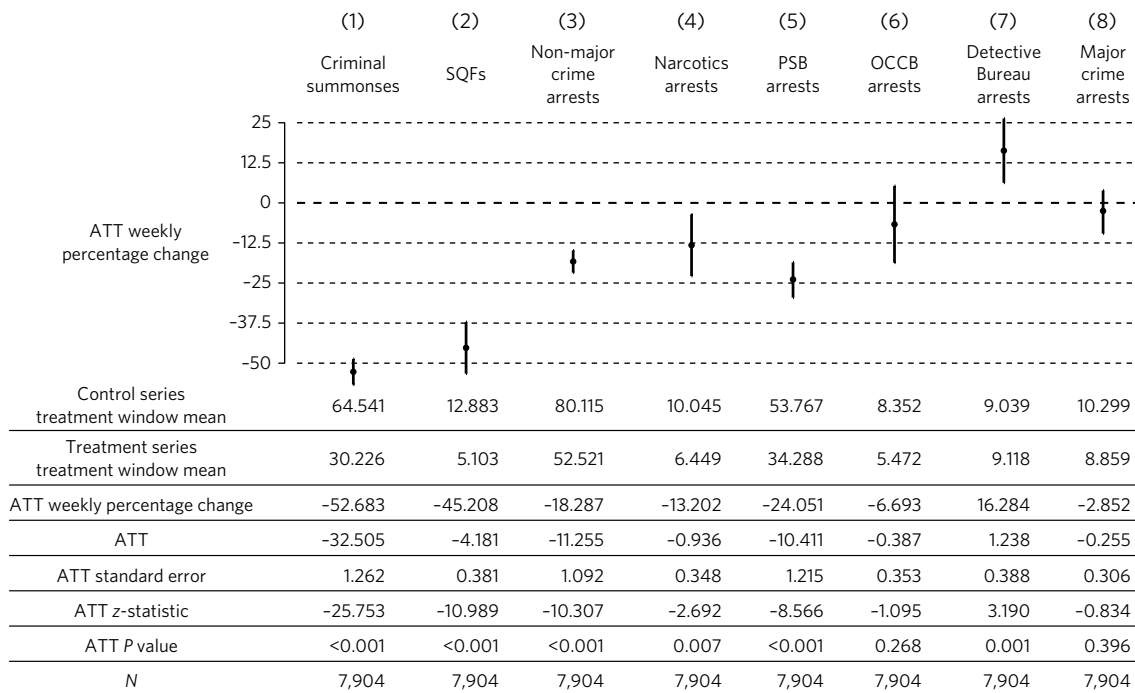


Fig. 2 | Effects of slowdown on police behaviour. The column headings of models (1)–(8) indicate outcome variables, measured at the precinct-week level. SQFs, stop, question and frisks. PSB, Patrol Services Bureau. OCCB, Organized Crime Control Bureau. Criminal summonses are misdemeanor and summary offences. Major crimes are murder, rape, robbery, felony assault, burglary, grand larceny and grand theft auto; non-major crimes are all other arrestable crimes. All models use a difference-in-differences design and negative binomial (NB2) regression. All models include all controls from the base specification (model (1) in Fig. 3), except (2), which substitutes a pre- and post-‘Floyd versus City of New York’ ruling indicator variable for ‘SQFs’. (3)–(8) also control for all arrests. (8) additionally controls for major crime complaints. The control series is 20 January 2013 to 19 January 2014. The treatment series is 19 January 2014 to 18 January 2015. The control series treatment window is 1 December 2013 to 19 January 2014. The treatment series treatment window is 30 November 2014 to 18 January 2015. ATT (average treatment effect on the treated) represents the mean predicted precinct-weekly change in the outcome during the slowdown. Mean predicted weekly percentage change = $100 \times \text{ATT} / (\text{mean of predicted counterfactual values})$. ATT standard error (s.e.) is clustered by precinct and calculated using the delta method, where the gradient is the exponentiated ‘Intervention’ coefficient. ATT z-statistic = $\text{ATT} / (\text{ATT s.e.})$. ATT P value represents the P value from a two-tailed z-test of the null hypothesis that the ‘Intervention’ coefficient (and thus ATT) is 0. N represents the number of precinct-week observations in regression. The vertical bars are 95% confidence intervals; the filled circles are point estimates.

validity, the NYPD employs the same index of major crime complaints when assessing tactical effectiveness²⁷. Third, focusing on major crime complaints is relatively standard within the literature, largely because these statistics are the most reliable across time and space⁵. Research auditing the NYPD’s major crime complaints data validates the statistics: patterns found in independent sources of crime data, including victims’ surveys, coroners’ reports and insurance losses, appear identical to major crime complaints²⁴.

Our analyses identify the effects of the 2014–2015 NYPD slowdown using a cross-sectional weekly time series of proactive policing and major crime complaints in 76 NYPD precincts. Our identification strategy uses difference-in-differences (DiD) to compare police and criminal behaviour before, during and after the slowdown with similar patterns observed during the same period the year before. For our primary analyses, we examine the period from mid-January 2013 through mid-January 2015 ($N=7,904$). In our DiD design, the ‘Treatment series’ includes precinct-weeks from mid-January 2014 to mid-January 2015. The ‘Control series’ is the same, but for 2013 to 2014. Drawing on the evidence above, our study defines the ‘Treatment window’ as 1 December through 19 January. The ‘Intervention’ (that is, the slowdown) is the seven-week period during the ‘Treatment window’ of the ‘Treatment series’. Effects are expressed as average treatment effects on the treated (ATTs), which represent the average predicted weekly change in the outcome induced by the slowdown. Our base specification uses negative binomial regression, but we also report results from replications using

Poisson and ordinary least squares, and interrupted time series (ITS) instead of DiD. In the analyses, we control for a variety of demographic characteristics, measures of police capacity and strategy, elements of concentrated disadvantage, season and weather indicators, time trends, and spatial–temporal lags of our dependent variables. Details on our measurement and identification strategy are contained in the Methods section.

We first estimate the slowdown’s effects on police behaviour (Fig. 2 and Supplementary Table 1). Following the procedures of a recent NYPD assessment of OMP, our approach ‘acknowledges that disorder reduction may not always require issuing summonses or making misdemeanor arrests, and may include other police activities like...situational crime prevention or problem-oriented policing strategies’, while limiting analyses of proactive policing to ‘focus exclusively on quality-of-life enforcement as a crime reduction tactic rather than these other forms of disorder reduction’¹². We find that, compared with other policing tactics, ‘Criminal summonses’ and SQFs decreased most precipitously during the slowdown, supporting earlier claims that the slowdown particularly affected low-level policing. ‘Non-major crime arrests’—all arrests apart from those for the seven major crimes—also declined significantly and by substantively meaningful amounts. Because CompStat data do not allow the study to exclude felony offences and violent crimes other than the seven major crimes from non-major crime arrests, we consider additional evidence locating the effects of the slowdown on proactive policing. ‘Narcotics arrests’, which are all

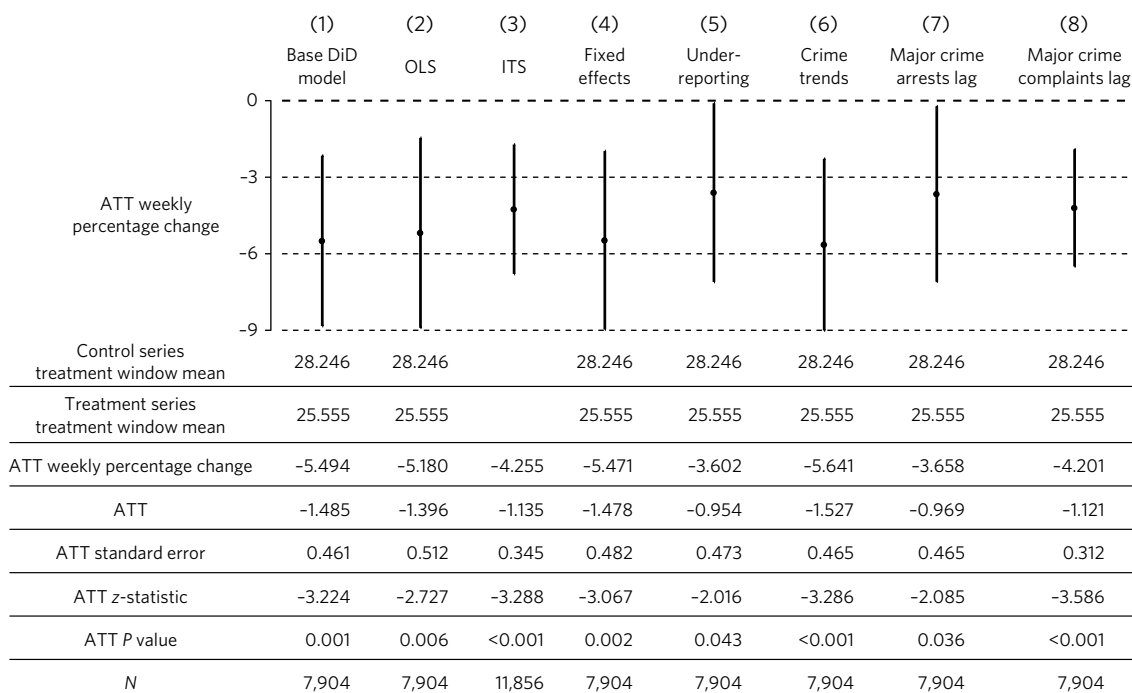


Fig. 3 | Effect of slowdown on major crime complaints. The outcome variable is the number of major crime complaints per week per precinct. All models (1)–(8) use negative binomial (NB2) regression, except (2), which uses ordinary least squares (OLS). For models using difference-in-differences (DiD), (1), (2) and (4)–(8), the series and treatment windows are the same as those in Fig. 2. The ITS model (3) specifies the ‘intervention’ as starting on 30 November 2014, and the ‘post-intervention’ period beginning on 19 January 2015. All models use all covariates described in the text for the base specification of model (1), except models (4) and (5), which exclude time-invariant predictors. Model (3) adds month dummies, and (4) and (5) add precinct dummies. Model (5) adds misdemeanour and violation complaints, and (6) adds the percentage change in weekly precinct major crime complaints between 2012 and 2011, and 2013 and 2012. Model (7) adds a one-week lag of major crime arrests, and (8) adds a one-week lag of major crime complaints. Standard errors for all models except (2) are calculated using the delta method, where the gradient is the exponentiated ‘intervention’ coefficient. For more information, see the note for Fig. 2.

charges relating to illegal drugs, dropped significantly during the slowdown. Alongside these measures, we consider arrests made by the Patrol Services Bureau (PSB), OCCB and Detective Bureau, conditioning our estimates on precinct-wide trends to locate any unique changes affecting the different bureaus. While the PSB engaged in significantly fewer arrests during the slowdown, there does not appear to have been a significant decline in the number of arrests by the OCCB. Replications examining arrests by officers in the Housing Bureau and Transit Bureau also returned non-significant results. In sharp contrast to this trend, evidence shows that arrests by the Detective Bureau increased significantly during the slowdown. This result is highly relevant to one of our theoretical mechanisms, since the Detective Bureau is charged with intensive investigations, rather than proactive policing. Further confirming that the slowdown’s effects were localized to proactive policing, we find no evidence that ‘Major crime arrests’ were significantly affected by the slowdown when we condition our estimates on ‘Major crime complaints’.

Having established that the slowdown significantly reduced proactive policing, we next estimate the slowdown’s effect on ‘Major crime complaints’ (Fig. 3 and Supplementary Table 2). Contradicting arguments that systematically decreasing proactive policing should correspond to increased crime (that is, the Ferguson effect), our results reveal that civilian complaints of major crimes declined by approximately 3–6% during the slowdown. Following these estimates, the decline in major crime caused by the cessation of proactive policing corresponds roughly to the relative decline in crime that earlier research attributed to the effects of mass incarceration²⁸. Replicating the analysis using alternative model

specifications, including ordinary least squares and interrupted time series specifications, produced substantively identical results (Fig. 3, Supplementary Tables 5 and 8 and Supplementary Fig. 3).

One might worry that under-reporting during the slowdown may be confounding our estimates of declining major crime complaints. Concerns of under-reporting do not nullify the identified decline in major crime complaints, but they do complicate a strict causal interpretation of our results. Perhaps officers were less likely to learn of crimes because they were staying in their squad cars, rather than patrolling the streets and speaking with victims about their experiences. Or trust in police may have fallen due to tensions between protesters and police. Recent findings show that high-profile cases of police violence suppress police-related 911 calls²². Anecdotal evidence also suggests that trust in police was down during this period, although trust had been declining since the summer. Further complicating questions about under-reporting, there is evidence that calls for NYPD service are significantly lower in areas with the highest rates of police stops and police use of force²⁹. Individuals may be less likely to report crimes when they think they are going to be stopped, questioned and potentially arrested in the process²⁰.

In our analyses, we examine how crime under-reporting may bias the results. We employ precinct fixed-effects to address time-invariant sources of under-reporting, such as communities’ varying histories of police distrust. We then model time-variant sources of under-reporting biases, such as those caused by the killing of Eric Garner and/or the heightened conflict between protesters and police. Model (5) in Fig. 3 controls for the number of community complaints reported in each precinct-week for misdemeanours and criminal violations. Assuming that time-variant sources of

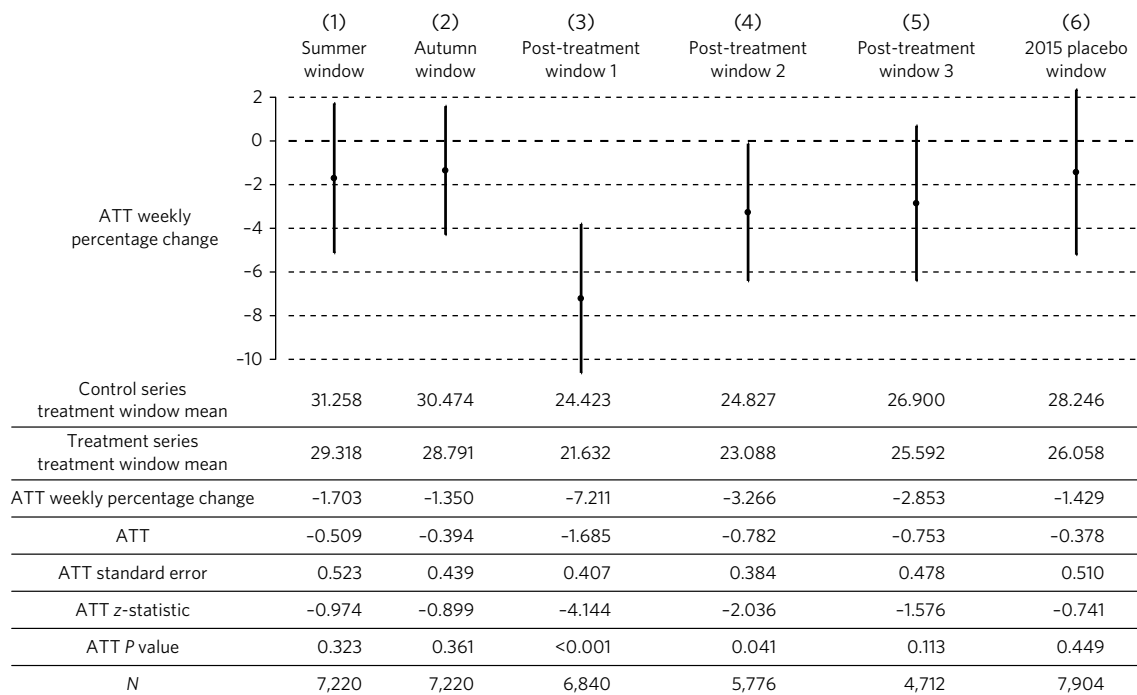


Fig. 4 | Alternative treatment specifications for changes in major crime complaints. The outcome variable is the number of major crime complaints per week per precinct. All models (1)–(6) use a difference-in-differences design and negative binomial (NB2) regression, as well as all covariates from the base specification in Fig. 3 model (1). The control (treatment) series by model are weeks 1–48 of 2013 (2014) in models (1) and (2), 11–48 of 2013 (2014) and 4–10 of 2014 (2015) in model (3), 18–48 of 2013 (2014) and 11–17 of 2014 (2015) in model (4), 25–48 of 2013 (2014) and 18–24 of 2014 (2015) in model (5) and, 4–52 of 2013 (2015) and 1–3 of 2014 (2016) in model (6). Control (treatment) series treatment windows by model are weeks 29–35 of 2013 (2014) in model (1), 37–43 of 2013 (2014) in model (2), 4–10 of 2014 (2015) in model 3, 11–17 of 2014 (2015) in model (4), 18–24 of 2014 (2015) in model (5), and 49–52 of 2013 (2015) and 1–3 of 2014 (2016) in model (6). Values for the last two weeks of the treatment series in model (6) are from 2015 (for results without imputed values, see Supplementary Fig. 5 model (5)). For more information, see the note for Fig. 2.

under-reporting are correlated across crime types, this model is robust to slowdown-induced under-reporting bias. While we cannot entirely rule out the effects of under-reporting, our results show that crime complaints decreased, rather than increased, during a slowdown in proactive policing, contrary to deterrence theory. Additional tests show the results are robust to specifications including controls for long-term trends in crime (Fig. 3 model (6)), lagged ‘Major crime arrests’ (Fig. 3 model (7)) and lagged ‘Major crime complaints’ (Fig. 3 model (8)). We report results from more robustness checks in Supplementary Fig. 5.

We also examined how the slowdown affected the different crimes constituting ‘Major crime complaints’ (Supplementary Fig. 6). While no category showed statistically significant increases during the slowdown, four complaint categories—murder, rape, robbery and grand theft auto—return statistically insignificant results, which we attribute to the relatively small number and high variance of such crimes. Robbery, the most common of the four nonsignificant categories, falls closest to statistical significance, but estimates appear highly sensitive to model specification. In light of earlier evidence, it is surprising that we find no robust increase in robbery complaints. One highly influential study finds that the strongest evidence supporting OMP exists in a ‘significant albeit modest association of disorder and officially measured robbery’¹⁶. And a recent analysis examining similar quasi-experimental conditions shows small increases in larcenies and robberies during the 1996–1997 NYPD labour negotiations strike³⁰. Our results belie these findings, as they show no statistically significant increase in complaints of any of the seven major crimes. Instead, evidence shows that the decline in major crime complaints identified during the slowdown was most affected by statistically significant reductions

in three high-volume categories: complaints of felony assault, burglary and grand larceny. Each week during the 2014–2015 slowdown, we estimate that 43 fewer felony assaults, 40 fewer burglaries and 40 fewer acts of grand larceny were reported.

Our analyses identify the timing and duration of the decline in major crime complaints by replicating the analysis using different operational definitions for the ‘Series’ and ‘Treatment window’ (Fig. 4, Supplementary Table 3 and Supplementary Fig. 4). The findings refute arguments that the decline in major crime complaints could have been affected by other factors emerging prior to the slowdown. No significant change in major crime complaints occurred following the death of Eric Garner (in July 2014) or in the months leading up to the slowdown. Additional tests confirm the timing of declines in major crime complaints aligns with the slowdown (Supplementary Fig. 8).

We also test whether the slowdown’s effect on crime complaints extended past its publicly announced end. Results from post-treatment analyses (Fig. 4 models (3) and (4)) show that statistically significant reductions in major crime complaints occurred seven and even fourteen weeks after sharp declines in proactive policing. While the study cannot address a principal concern of the law enforcement community—that reductions in proactive policing could increase criminality years later—it demonstrates substantial short-term reductions in crime that should prompt reflection on the mechanisms linking proactive policing to deterrence. Wilson and Kelling, for example, suggest that the benefits of proactive policing could be observed ‘in a few years or even a few months’¹⁰. Other studies point to the fact that crime rates remain plastic and highly volatile as evidence that persistent proactive policing caused NYC’s crime decline, rather than structural factors such as demography²⁴. As expressed by the NYPD, ‘Current crime levels

don't stay down by themselves...crime is actively managed in New York City everyday¹³. Further research will need to examine additional long-term effects. Within the short term, we estimate that the slowdown resulted in roughly 2,100 fewer major crime complaints. This estimate extrapolates from the ATTs for the seven weeks of the slowdown, plus the fourteen weeks of the two significant post-treatment windows. Tests of subsequent windows in the spring of 2015 return non-significant results, indicating that, as NYPD tactics returned to normal, the city's crime rate eventually reverted to its pre-treatment baseline. The nonsignificant results of a placebo test using a window spanning the seven weeks after the killing of Freddie Gray in April 2015 (Fig. 4 model (5)) support our conclusion that the results were not solely induced by the effects of police-related violence on under-reporting²². Finally, results from a placebo test (Fig. 4 model (6)) estimating the counterfactual scenario in which the slowdown took place during the subsequent year (2015–2016) prove nonsignificant, confirming that we have not misidentified our causal effect.

Findings from our study warrant a reconsideration of the assumptions guiding scholarship and practice related to enforcement and legal compliance. In their efforts to increase civilian compliance, certain policing tactics may inadvertently contribute to serious criminal activity. The implications for understanding policing in a democratic society should not be understated. It is well established that proactive policing is deployed disproportionately across communities, and that areas with high concentrations of poverty and people of colour are more likely to be targeted⁸. Our results imply not only that these tactics fail at their stated objective of reducing major legal violations, but also that the initial deployment of proactive policing can inspire additional crimes that later provide justification for further increasing police stops, summonses and so on. The vicious feedback between proactive policing and major crime can exacerbate political and economic inequality across communities³¹. In the absence of reliable evidence of the effectiveness of proactive policing, it is time to consider how proactive policing reform might reduce crime and increase well-being in the most heavily policed communities.

Methods

Data. For benchmarking purposes, each CompStat reports data both for the current year and for the same seven-day range in the previous year. Thus, the 2015 CompStats include 2014 data, and the 2014 CompStats include 2013 data, with weeks matched by their calendar start and end days. The 2014 data contained in the 2014 and 2015 CompStats do not perfectly align, however. Because of when the 52nd week of the previous year finished, week 1 of 2014 begins on 30 December 2013, and week 1 of 2015 begins on 29 December 2014. As a result, the weekly 2014 totals from the 2015 CompStats are off by one day compared with the weekly 2014 totals from the 2014 CompStats. While the choice of how to cut the data does not meaningfully impact our results, we constructed our data set in the following way. Necessarily, 2013 data are taken from the 2014 CompStats, and 2015 data from the 2015 CompStats. But because there are two observations for each week in 2014 (one from the 2014 reports, one from the 2015 reports), we are forced to adopt a rule for which values to use. Because our treatment and control series span multiple years by approximately three weeks in the beginning of January, we reasoned that the best criterion to use to subset the data is to maintain internal consistency within each series. To accomplish this, we used only the 2014 CompStats for all weeks measured as part of the control series, and only the 2015 CompStats for all weeks of the treatment series.

For several reasons, we are confident that the results are not affected by the one-day difference in the series. First, since the 'Intervention' is estimated by averaging over a seven-week period, days contained within the five middle weeks overlap completely, leaving only two weeks that are off by a day. Second, we replicated the analyses by averaging the 2014 weeks from the 2014 and 2015 reports. This approach yields comparable results, but because only a single data point is available for each week in 2013 and 2015, we prefer to maintain the data's internal consistency, rather than introduce another manipulation.

Since our data come from the NYPD, it is worth considering potential sources of bias in police reporting. Concerns have been raised about police data being influenced by the officers tasked with collecting statistics, as well as their superiors³. Still, we feel confident in the CompStat data for several reasons. First, police data are often strongly preferable to alternative sources. Because police records contain a more extensive listing of activity, they are often used to identify the form and extent of bias in other data sources. Second, to minimize the

biases associated with human error, the NYPD requires officers to apply a 'strict interpretation bias'. When reporting a crime complaint, an officer must enter the incident on the basis of the most serious crime described by the claimant, regardless of whether the officer believes the perpetrator can be tried or arrested for that offence. This procedure was put into place under the theory that strict interpretation bias would increase the willingness of individuals to come forward with crime complaints. As a result, the majority of errors in the categorization of a crime should lead to upgrading, rather than downgrading, the criminal classification²⁵.

Any remaining bias from manipulation by police officers would predispose the study towards identifying an escalation in major crime complaints. Prior to the slowdown, precinct commanders' interest lay in demonstrating continuing declines in crime. Professional incentives reversed during and after the slowdown insofar as commanders wished to demonstrate the necessity of the police force and the effectiveness of their policing strategies.

With regards to police protocol, there were three important changes in NYPD procedure during our time series worthy of mention. First, on 31 October 2013, an appellate court ruled on 'Floyd versus City of New York', ordering NYC to eliminate racial profiling in the NYPD's stop-and-frisk encounters. We display the corresponding sharp decline in these encounters in Fig. 1. Our analyses control for the effects of the 'Floyd' decision by including precinct-week counts of SQFs in models of other police and criminal behaviours, and a dummy for the pre- and post-'Floyd' periods in model (2) of Fig. 2, as SQFs are the outcome. Second, in July 2014, the Brooklyn District Attorney Ken Thompson declared that his office would no longer prosecute marijuana possession under certain conditions. Third, on 19 November 2014, NYC formally decriminalized marijuana possession, making it a summons rather than an arrestable offence. While these three procedural changes surely impacted policing practices, their causes are unrelated to the slowdown, and thus we should not expect them to impact our causal estimation. Indeed, our analyses in Fig. 4 and Supplementary Fig. 8 show that the timing of changing patterns of compliance corresponds to the period of the slowdown, rather than these earlier procedural changes.

With regards to data availability, we encountered missingness in two situations. The first concerns three of our measures of police strength and strategy.

The number of officers per precinct is from 2007, and the number of Civilian Complaint Review Board (CCRB) complaints is from 2013, and thus predate the formation of the 121st precinct, which became fully operational in July 2013. To address this, we imputed values for this precinct using data from the 120th and 122nd precincts, which were split to form the 121st. We weight these variables' covariates for all three precincts proportionately on the basis of geographic, and when appropriate temporal, coverage. We did the same for SQFs before July 2013.

The second site of missing data results from the fact that the NYPD has thus far failed to turn over CompStat reports from two weeks in January 2016. In spite of the NYPD's recalcitrance, we have no reason to suspect that the missing weeks (weeks 2–3 of 2016) impact the results. To empirically demonstrate that our results are not affected by missing data, we take a conservative approach when imputing data for the missing values. In the final column of Fig. 4, we fill in the missing 2016 data with the two weeks from the actual slowdown (in January 2015). We believe this is a better modelling strategy than multiple imputation, which can introduce bias when applied to nonlinear models³². Imputing the missing data using the actual slowdown values also presents a harder test for demonstrating nonsignificance in the placebo treatment as compared with the last observation carried forward. In the first four weeks of the slowdown, rates of major crime complaints were nearly 20% lower as compared with the same period the following year. Replicating the 2015 placebo tests without the imputed weeks produces comparable non-significant results (Supplementary Fig. 5 model (5)).

Modelling strategy. The systematic component of our econometric model is represented in equation (1). The DiD model estimates changes in police behaviour or civilian crime complaints (Y) as a function of dichotomous indicators of the 'Series' (S) and 'Treatment window' (T), an interaction of the two ('Series' \times 'Treatment window') representing the 'Intervention' period ($ST^T = I$), and a variety of covariates (X)^{33,34}.

$$E\{Y_i | S_i, T_i, I_i, X_i\} = r_i = \exp(\alpha + \gamma S_i + \lambda T_i + \delta I_i + X_i \beta) \quad (1)$$

A critical requirement of the DiD modelling strategy is the 'parallel trends' assumption. To reliably estimate differences during the treatment window, the data must follow the same pattern outside the window. Figure 1 confirms that the control series indeed provides a reliable baseline from which to measure any changes induced by the slowdown.

We estimate the models using a negative binomial specification ($Y_i \approx \text{NegBin}(r_i, p)$) because all outcome variables are overdispersed count data, as revealed by two types of dispersion test. For the base model using 'Major crime complaints' as the outcome variable (Fig. 3 model (1)), a likelihood ratio test of the null hypothesis that the Poisson model restriction of equal mean and variance is true is rejected with a $\chi^2(1)$ value of 2,377 ($P < 0.001$, two-sided). Results are comparable for all other models (see Supplementary Tables 1–3). Ordinary least squares is even less appropriate than Poisson, as the dependent variables are neither normally distributed nor interval. Furthermore, because observations within precincts

are not independent, and hence their errors are correlated, we calculate robust standard errors clustered by precinct.

While the slowdown in policing is arguably independent from precinct-level covariates, we include a number of controls in case these variables influence the precincts' responsiveness to the slowdown. Including controls such as population and other demographic characteristics helps to normalize the variance in the dependent variables across precincts. Because it lacks a residential population, all analyses exclude the Central Park Precinct. We use the most recent demographic data, which are taken from the 2014 five-year American Community Survey (ACS). Using the ACS, we identified each precinct's 'Population', as well as the crime-prone age group 'Percentage aged 15–24'.

We also include a number of key indicators of concentrated disadvantage. Using data from the ACS, we generated precinct-level measures of 'Average family income', 'Percentage of residents who are persons of colour', 'Percentage unemployed' and several household-level measures, including 'Percentage of households on public assistance', 'Percentage of households headed by women with children', 'Percentage of occupied housing units rented' and 'Percentage of households vacant'. Because these factors loaded poorly on a single dimension as well as on two dimensions, all analyses with covariates incorporate these variables individually (see Supplementary Tables 1–3). In Supplementary Fig. 5 model (3), we report a replication using our measure of concentrated disadvantage, which is defined as the mean of the standardized (that is, centred at 0 and scaled such that standard deviations are equal to 1) values of 'Percentage of residents who are persons of colour', 'Percentage unemployed', 'Percentage of households on public assistance' and 'Percentage of households headed by single women with children'. The model accordingly does not include these constituent variables individually to avoid collinearity.

Our models also control for precinct-level variation in policing capacity and behaviour. In addition to the total precinct-week SQFs mentioned earlier, we construct per capita precinct-level variables of the number of officers assigned to a precinct in 2007 ('Officers per 100,000 people') and complaints registered against the precinct with the CCRB in 2013 ('CCRB complaints per 100,000 people') using data from previous research³⁵. The CCRB data also provide an indicator of the distribution of complaints across racial groups, which we measure as 'Percentage of CCRB complaints by persons of colour/percentage of residents who are persons of colour'.

We further include three weather-related controls, each of which are weekly averages of daily measures for NYC from the National Weather Service: 'Mean temperature', 'Total rain accumulation' and 'Total snow/sleet accumulation'. To account for temporal autocorrelation and geographic spillover effects, all models include a one-week 'Spatial lag' of the dependent variable. To construct this, we identified adjacent, contiguous precincts for each precinct, and calculated the mean of the previous week's values. Because NYC is composed of multiple islands connected by bridges and tunnels, we deem this more appropriate than using an inverse distance weighted measure. The choice does not meaningfully alter the results.

Lastly, while Fig. 1 lends support to the parallel trends assumption necessary to DiD, we include two additional controls for temporal variance in policing and criminal behaviour. First, to adjust for the ongoing downward trend in crime, we include a 'Time counter', which counts the number of weeks since the first week of the time series, starting at 1. Second, alongside the three weather-related variables, we include dummy variables for 'Summer', 'Autumn' and 'Winter' to help control for seasonal effects.

Using our fitted model, we estimate our causal effects as shown in equation (2), where $\bar{\tau}$ represents the average treatment effect on the treated (ATT):

$$\begin{aligned} \bar{\tau} &= \hat{E}[Y_1 - Y_0 | I = 1] \\ &= \frac{1}{N_r} \sum_{i=1}^{N_r} \left[\exp(\hat{\alpha} + \hat{\gamma} + \hat{\lambda} + \hat{\delta} + X_i \hat{\beta}) - \exp(\hat{\alpha} + \hat{\gamma} + \hat{\lambda} + X_i \hat{\beta}) \right] \end{aligned} \quad (2)$$

Y_1 and Y_0 represent potential outcomes had the 'treatment' (e.g., the slowdown in Figs. 2 and 3) occurred versus had it not. N_r is the number of observations in the intervention period, which are indexed by i . The percentage change in the outcome induced by the 'Intervention' is:

$$100 \times \frac{\bar{\tau}}{\sum_{i=1}^{N_r} [\exp(\hat{\alpha} + \hat{\gamma} + \hat{\lambda} + X_i \hat{\beta})]} \quad (3)$$

The standard error can be found by applying the delta method to the exponentiated coefficient, multiplied by the average predicted counterfactual. Significance is assessed with z -tests.

In words, our procedure for calculating causal effects is as follows. We generate ATTs by averaging the precinct-week differences between the predicted value with the 'Intervention' set to 1 versus set to 0 during the 'Intervention' period (that is, the average difference between the values predicted for each precinct-week observation had the slowdown occurred versus had it not, all else equal)³³.

The ATT is converted to an average predicted weekly percentage change by dividing it by the mean predicted counterfactual and multiplying by 100. Figures 2–4 graphically present the average predicted percentage change and corresponding 95% confidence intervals with delta method standard errors clustered by precinct, as well as report the raw ATTs and their standard errors and P values. Statistical significance is determined using two-tailed z -tests. More detailed results for the models in Figs. 2–4 can be found in Supplementary Tables 1–3.

While we believe that DiD is the best modelling approach given our data and the nature of criminality and policing, we also ran ITS models using the entire time series. In the ITS analyses, we replicate the modelling approach adopted in earlier research on police slowdowns, with the addition of our precinct-level control variables³⁰. Results from the base specification comparing ITS with the DiD estimates are presented in Fig. 3. Results from a full replication of all models using ITS instead are displayed in Supplementary Figs. 2–4. We contend, however, that DiD is more appropriate primarily for three reasons. First, with ITS, the modeller must specify the functional form of the proposed trends before, during and after the 'treatment'. The most common assumption is that trends are linear. In Fig. 3 we do not include such additional trend shifts for simplicity, but doing so does not alter the results. In Supplementary Fig. 7 we show that the predicted values from a specification in which we include slowdown and post-slowdown trend shifts results in counterfactual predictions during the slowdown that closely mirror those of the base DiD model from Fig. 3 model (1). Second, ITS is especially sensitive to seasonal effects, and while there are different ways to control for them, none are perfect. Third, and most importantly, the treatment window overlaps very closely with (meteorological and astronomical) winter, exacerbating the previously mentioned issue, especially because it is a time of depressed crime in general. DiD does not require imposing as much structure, and controls for cyclical trends by design. Regardless, the ITS results are essentially the same.

Lastly, while interaction terms in nonlinear models are not usually equal to the product term, in the special case of difference-in-differences models, identification of the ATT is as simple as equation (2)^{33,34}. The same holds for ITS models that do not include trend shift variables. In such cases, the ATT is similarly derived from the level shift coefficient. Adding trend shift variables, however, introduces considerable complexity for two reasons: first, they are products of interacting level shifts with the time trend; and second, the total treatment effect is the combination of the multiple effects. While the point estimate of the ATT can be calculated using the average predicted change during the intervention period, standard errors are not as easily obtained. We ran such models and found that the ATTs and 'Intervention' level shift coefficients and standard errors were nearly identical to models without the trend shifts. Therefore, we focus on the simpler case in Fig. 3 model (3) and Supplementary Figs. 2–4.

Code availability. The computer code that support the findings of this study is available from the corresponding author upon reasonable request.

Data availability. The data that supports the findings of this study are available from the corresponding author upon reasonable request.

Received: 22 February 2017; Accepted: 23 August 2017;
Published online: 25 September 2017

References

- Kubrin, C. E., Messner, S. F., Deane, G., McGeevrey, K. & Stucky, T. D. Proactive policing and robbery rates across U.S. cities. *Criminology* **48**, 57–97 (2010).
- Braga, A. A., Welsh, B. C. & Schnell, C. Can policing disorder reduce crime? A systematic review and meta-analysis. *J. Res. Crime Delinq.* **52**, 567–588 (2015).
- Cerdá, M. et al. Misdemeanor policing, physical disorder, and gun-related homicide: a spatial analytic test of "broken-windows" theory. *Epidemiology* **20**, 533–541 (2009).
- MacDonald, J., Fagan, J. & Geller, A. The effects of local police surges on crime and arrests in New York City. *PLoS ONE* **11**, e0157223 (2016).
- Harcourt, B. E. *Illusion of Order: The False Promise of Broken Windows Policing* (Harvard Univ. Press, Cambridge, MA, 2005).
- Weisburd, D., Wooditch, A., Weisburd, S. & Yang, S.-M. Do stop, question, and frisk practices deter crime? *Criminol. Public Policy* **15**, 31–56 (2016).
- Rosenfeld, R. & Fornango, R. The relationship between crime and stop, question, and frisk rates in New York City neighborhoods. *Justice Q.* **34**, 1–21 (2017).
- Fagan, J., Geller, A., Davies, G. & West, B. in *Race, Ethnicity, and Policing: New and Essential Readings* (eds Rice, S. K. & White, M. D.) 309–348 (New York Univ. Press, New York, 2010).
- Fagan, J., Braga, A. A., Brunson, R. K. & Pattavina, A. Stops and stares: street stops, race, and surveillance in the new policing. *Fordham Urban Law J.* **43**, 621–696 (2016).

10. Wilson, J. Q. & Kelling, G. L. Broken windows: the police and neighborhood safety. *Atlantic Monthly* **249**, 29–38 (1982).
11. Weisburd, D. Does hot spots policing inevitably lead to unfair and abusive police practices, or can we maximize both fairness and effectiveness in the new proactive policing? *Univ. Chic. Leg. Forum* **2016**, 661–689 (2016).
12. *Broken Windows and Quality-of-Life Policing in New York City* (Department of Investigation, The New York City Police Department, City of New York, 2015).
13. Bratton, W. J. & Kelling, G. L. Why we need broken windows policing. *City Journal* <https://www.city-journal.org/html/why-we-need-broken-windows-policing-13696.html> (Winter, 2015).
14. Pyrooz, D. C., Decker, S. H., Wolfe, S. E. & Shjarback, J. A. Was there a Ferguson effect on crime rates in large U.S. cities? *J. Crim. Justice* **46**, 1–8 (2016).
15. Fagan, J. & Davies, G. Street stops and broken windows: terry, race and disorder in New York City. *Fordham Urban Law J.* **28**, 457–504 (2000).
16. Sampson, R. J. & Raudenbush, S. W. Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. *Am. J. Sociol.* **105**, 603–651 (1999).
17. Leovy, J. *Ghettoside: A True Story of Murder in America* (Spiegel & Grau, New York, NY, 2015).
18. Hinkle, J. C. & Weisburd, D. The irony of broken windows policing: a micro-place study of the relationship between disorder, focused police crackdowns and fear of crime. *J. Crim. Justice* **36**, 503–512 (2008).
19. Geller, A., Fagan, J., Tyler, T. & Link, B. G. Aggressive policing and the mental health of young urban men. *Am. J. Public Health* **104**, 2321–2327 (2014).
20. Tyler, T. R., Jackson, J. & Mentovich, A. The consequences of being an object of suspicion: potential pitfalls of proactive police contact. *J. Empir. Leg. Stud.* **12**, 602–636 (2015).
21. Kohler-Hausmann, I. Managerial justice and mass misdemeanors. *Stanford Law Rev.* **66**, 611–693 (2014).
22. Desmond, M., Papachristos, A. V. & Kirk, D. S. Police violence and citizen crime reporting in the black community. *Am. Sociol. Rev.* **81**, 857–876 (2016).
23. Kirk, D. S. & Papachristos, A. V. Cultural mechanisms and the persistence of neighborhood violence. *Am. J. Sociol.* **116**, 1190–1233 (2011).
24. Zimring, F. E. *The City That Became Safe: New York's Lessons for Urban Crime and Its Control* (Oxford Univ. Press, New York, City of New York, NY, 2012).
25. Kelley, D. N. & McCarthy, S. L. *The Report of the Crime Reporting Review Committee to Commissioner Raymond W. Kelly Concerning CompStat Auditing* (New York Police Department, NY, 2013).
26. Rosenfeld, R. & Fornango, R. The impact of police stops on precinct robbery and burglary rates in New York City, 2003–2010. *Justice Q.* **31**, 96–122 (2014).
27. Eure, P. K. *An Analysis of Quality-of-Life Summonses, Quality-of-Life Misdemeanor Arrests, and Felony Crime in New York City, 2010–2015*. (Department of Investigation, City of New York, 2016).
28. Roeder, O., Eisen, L.-B. & Bowling, J. *What Caused the Crime Decline?* (Brennan Center for Justice, New York, NY, 2015).
29. Lerman, A. E. & Weaver, V. M. Staying out of sight? Concentrated policing and local political action. *Ann. Am. Acad. Pol. Soc. Sci.* **651**, 202–219 (2014).
30. Chandrasekher, A. C. The effect of police slowdowns on crime. *Am. L. Econ. Rev.* **18**, 385–437 (2016).
31. Lerman, A. E. & Weaver, V. M. *Arresting Citizenship: The Democratic Consequences of American Crime Control* (Chicago Univ. Press, Chicago, IL, 2014).
32. Li, X., Mehrotra, D. V. & Barnard, J. Analysis of incomplete longitudinal binary data using multiple imputation. *Stat. Med.* **25**, 2107–2124 (2006).
33. Puhani, P. A. The treatment effect, the cross difference, and the interaction term in nonlinear “difference-in-differences” models. *Econ. Lett.* **115**, 85–87 (2012).
34. Ai, C. & Norton, E. C. Interaction terms in logit and probit models. *Econ. Lett.* **80**, 123–129 (2003).
35. Rengifo, A. F. & Fowler, K. Stop, question, and complain: citizen grievances against the NYPD and the opacity of police stops across New York City precincts, 2007–2013. *J. Urban Health* **93**, 32–41 (2016).

Author contributions

C.M.S. developed the original study concept. C.M.S. and Z.P.O. gathered and analysed the data, and drafted and revised the manuscript. Z.P.O. wrote the computer code and generated figures and tables.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at doi:10.1038/s41562-017-0211-5.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.M.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.