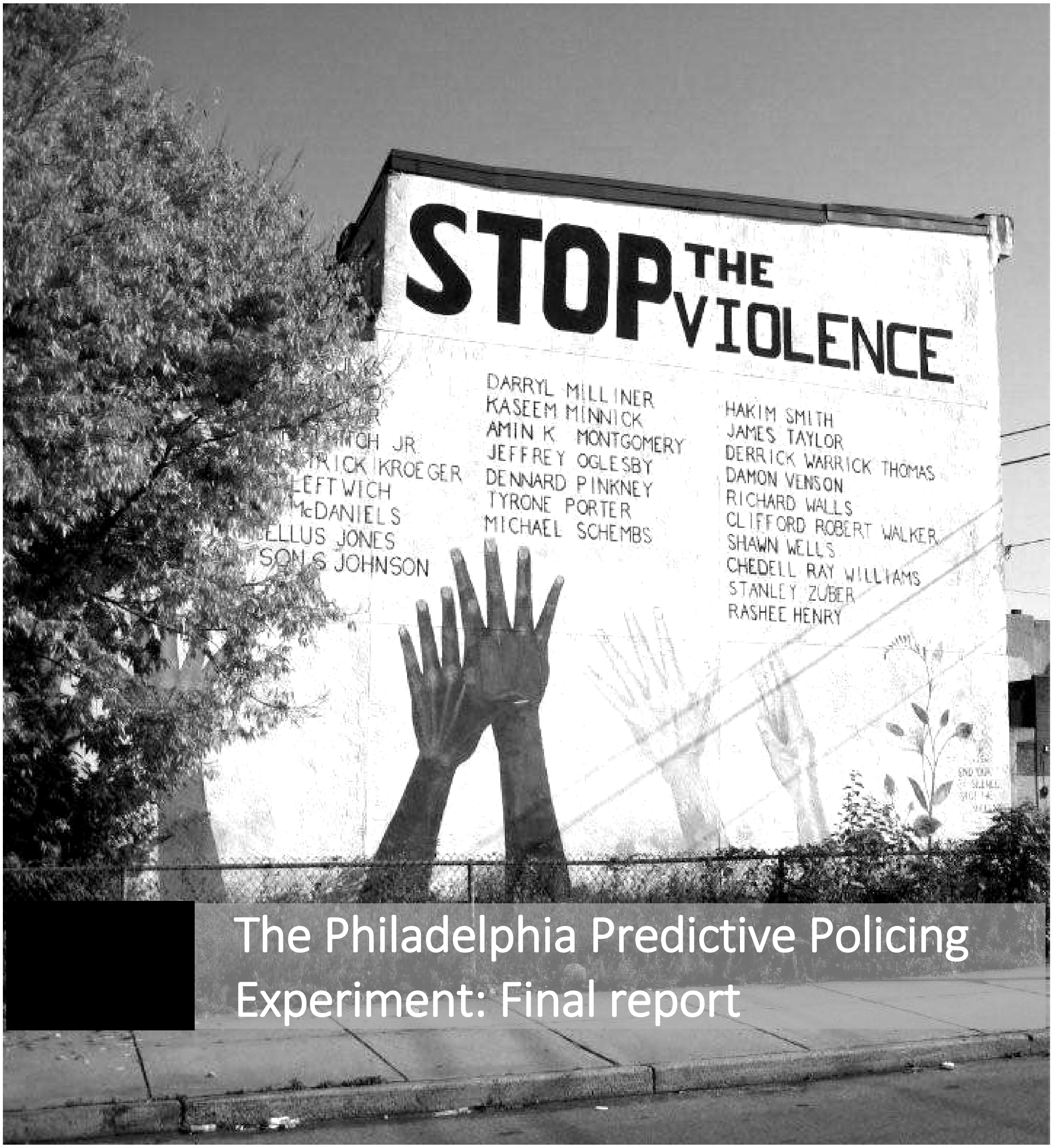
TEMPLE UNIVERSITY

Center for Security and Crime Science



The Philadelphia Predictive Policing Experiment: Final report

Report Title: The Philadelphia Predictive Policing Experiment: Final report

Award number: 2014-R2-CX-0002

Authors: Jerry H. Ratcliffe, PhD, Ralph B. Taylor, PhD, Amber P. Askey, PhD, Ryan Fisher, MA and Josh Koehnlein, MA.

Contact:

Dr. Jerry Ratcliffe
Center for Security and Crime Science
Department of Criminal Justice
Temple University
5th floor, Gladfelter Hall
1115 Polett Walk
Philadelphia PA 19122
215-204 7918

www.cla.temple.edu/cj

jhr@temple.edu

January 2018

NOTES

This final report follows the Final Technical Report Guidelines laid out at http://www.nij.gov/funding/pages/final-technical-report-guidelines.aspx

Parts of this report are copied from existing publications written by the research team, and in particular the following publications:

- Taylor, RB, Ratcliffe, JH & Perenzin, A (2015) Can we predict long-term community crime problems? The estimation of ecological continuity to model risk heterogeneity, *Journal of Research in Crime and Delinquency*, 52(3): 635-657.
- Ratcliffe, JH (2016) *Intelligence-Led Policing*, Routledge: London. Second edition.

ACKNOWLEDGEMENTS

The authors are immensely grateful to the officers of the Philadelphia Police Department for their dedication to the safety of the city, and for their assistance and help with this project. In particular, officers in marked and unmarked cars were unwaveringly welcoming to our researchers during ride-alongs. We would like to specifically recognize Police Commissioner Charles Ramsey (ret.), Commissioner Richard Ross and Deputy Commissioner Kevin Bethel (ret.) without whom this project would not have succeeded. We have a huge debt of gratitude to Nola Joyce, Kevin Thomas, John Grasso and Shawn Hagan for their immeasurable efforts to make this project a reality. At Temple University, we would like to recognize the work of Ryan Sentner, Mackenzie Tyler, Megan Van Wallendael, and all of the graduate students who worked on this project.

We are indebted to Jeremy Heffner and his colleagues at Azavea for their technical assistance throughout the project. Finally, we would like to recognize and thank Dr. Joel Hunt from the National Institute of Justice for his help and careful stewardship of this project. Opinions or points of view expressed in this report are those of the authors and do not necessarily reflect the official position or policies of the agencies, companies or individuals mentioned here.

ABSTRACT

Since the NIJ's first predictive policing symposium was conducted in Los Angeles in 2009, a number of police departments have adopted predictive policing technologies and strategies, with varying degrees of success. Notwithstanding its growing popularity, empirical research on predictive policing is still relatively scarce. The Philadelphia Predictive Policing Experiment was a place-based randomized control trial borne out of a collaboration between Temple University researchers and the Philadelphia Police Department. It explored the crime reduction effects of three different operationally-feasible police patrol strategies driven by a predictive policing software program.

The experiment was conducted in two phases; the property crime phase ran for one shift only for 90 days from June 1, 2015 through August 25, 2015. After a scheduled break, the violent crime phase ran for one shift only for 92 days from November 1, 2015 through January 31, 2016. Predicted crime areas were generated by the HunchLab program designed by Azavea, which predicted three 500 x 500 square foot high-crime grid cells, for each eight-hour experimental shift, occurring once per day, for each district. Twenty Philadelphia police districts were randomly assigned to one of four experimental conditions using block randomization with a 1:1:1:1 allocation ratio across the four conditions. These conditions include (1) a control condition, with a business-as-usual patrol strategy, (2) an awareness condition where officers were made aware of the predicted high crime areas and asked to focus on those areas when able, (3) an enhanced awareness model with a dedicated marked patrol car to exclusively patrol the predicted crime areas, and (4) an enhanced awareness model with a dedicated unmarked vehicle to exclusively patrol the predicted areas.

Graduate research assistants and the primary investigators of the project deployed as field researchers to collect observations during ride-alongs with officers to assess the implementation of the different patrol strategies and monitor treatment integrity. Mixed effects models of property crime in spatially buffered treatment areas found the marked patrol car condition resulted in a 31 percent reduction in property crime counts, or a 36 percent reduction in the number of cells experiencing at least one crime. In addition, during the eight hours following the

property crime patrols, the marked car districts were associated with a reduction in property crime counts that were 41.6 percent lower and expected crime occurrences that were 48.1 percent lower compared to the control districts. There were no crime reduction benefits associated with the violent phase of the experiment, nor the property crime awareness or unmarked car interventions. The relative rarity of crime on an hour-by-hour basis in limited geographic areas hindered making confident inferences about any crime reductions regarding the experimental conditions. In other words, while the percentages were substantial, the results were not statistically significant due to floor effects. While the percentage changes sound substantial, the specific numbers are small. This translates to a reduction in three crimes over 3 months for an average city district patrolling around three grids. To extrapolate, if each of the 21 geographic districts in Philadelphia dedicated a marked car to three grids for an 8 hour shift each day, we estimate the result would be a reduction of 256 Part I property crimes per year.

The qualitative data highlighted the challenges inherent in innovation implementation within a major metropolitan police department. Beyond the institutional inputs which remain crucial to implementation, there is a clear need to address the buy-in and attitudes of front-line officers regarding the innovation. That said, this is the first robust empirical evidence derived from a randomized experiment of a tactic tied to a predictive policing implementation demonstrating some crime reduction. At present, there is little available evidence regarding how best to use shift-based patrol police officers beyond hot spots policing and this experiment provides an additional avenue for patrol commanders. We conclude that a marked police car targeted to a limited number of predicted crime areas is a promising crime reduction strategy, but police commanders should be realistic about the limited evidence they are likely to see and may need to manage this strategy alongside other resource constraints and public and officer expectations.

CONTENTS

NOTES 3
ACKNOWLEDGEMENTS 4
ABSTRACT....5
TABLES....9
FIGURES.... 11
EXECUTIVE SUMMARY 12
SYNOPSIS OF THE PROBLEM AND THE RESEARCH PURPOSE 12
RESEARCH DESIGN AND RESULTS 15
CONCLUSIONS AND IMPLICATIONS 18
INTRODUCTION.... 19
STATEMENT OF THE PROBLEM 19
LITERATURE REVIEW.... 21
STATEMENT OF HYPOTHESIS OR RATIONALE FOR THE RESEARCH 24
METHODS.... 26
EXPERIMENTAL DESIGN.... 26
STUDY SETTING 26
PARTICIPANTS.... 27
INTERVENTIONS 27
OPERATIONALIZATION OF THE INTERVENTION 29
OUTCOMES AND DATA COLLECTION 33
ADDITIONAL OUTCOMES.... 34
RANDOMIZATION 35
GAUGING INITIAL PROBABILISTIC EQUIVALENCE ACROSS ASSIGNMENT CONDITIONS 41
CONCEALMENT AND BLINDING.... 42
QUALITATIVE FIELDWORK.... 42
STATISTICAL METHODS 46
SAMPLE SIZE.... 46
CROSS-TABULATION.... 46
INTENTION TO TREAT 47
MULTIVARIATE ANALYSIS OF CONTIGUOUS AREAS 48
ADDITIONAL ANALYSES 50
RESULTS: MICRO-LEVEL QUANTITATIVE FINDINGS 51
SOFTWARE FORECAST EFFICACY 51
PROPERTY CRIME PHASE DESCRIPTIVE STATISTICS.... 53
PROPERTY CRIME PHASE MULTILEVEL MODELS 56
VIOLENT CRIME PHASE DESCRIPTIVE STATISTICS.... 60
PROPERTY CRIME POST-TREATMENT EFFECTS.... 63
VIOLENT CRIME POST-TREATMENT EFFECTS.... 66
SUMMARY OF QUANTITATIVE RESULTS.... 68
A NOTE ON FLOOR EFFECTS 69

RESULTS: DISTRICT LEVEL QUANTITATIVE FINDINGS 72
UNDERPINNINGS 72
DESCRIPTIVE PRELIMINARIES 76
CONTRASTING TWO WAYS TO THINK ABOUT TRENDS OVER TIME 80
SUMMARY ON GLOBAL VERSUS LOCAL PREDICTIONS ABOUT TIME AND CRIME ACROSS CONDITIONS 91

RESULTS: DISTRICT-LEVEL TREATMENT SPECIFIC FINDINGS 93
PREDICTORS AND OUTCOME VARIABLES 93
MODEL SEQUENCE 94
SUMMARY 100

RESULTS: DAILY REPORT LOGS 102
PROPERTY CRIME EXPERIMENT PHASE (OFFICER FORMS) 102
PROPERTY CRIME EXPERIMENT PHASE (SUPERVISOR FORMS) 107
VIOLENT CRIME EXPERIMENT PHASE (OFFICER FORMS) 112
VIOLENT CRIME EXPERIMENT PHASE (SUPERVISOR FORMS) 116

RESULTS: FIELDWORK OBSERVATIONS 121
FIELD OBSERVATION FORMS 121
INSTITUTIONAL ENABLING MECHANISMS 124
INNOVATION UTILITY AND ACCEPTANCE 130
SUMMARY 136

CONCLUSIONS 138
DISCUSSION OF FINDINGS 138
LIMITATIONS 139
COSTS 141
GENERALIZABILITY 141
INTERPRETATION 141

DISSEMINATION OF RESEARCH FINDINGS 143
JOURNAL ARTICLES 143
CONFERENCE PRESENTATIONS 143
WEB SITES 144

REFERENCES 145

APPENDICES 154

TABLES

Table 1 Census variables that contributed to the blocked design. 38
Table 2 Temporal weighting periods for pre-intervention analysis. 39
Table 3: Kendall's Tau correlation test results for equivalency 40
Table 4 Final qualitative field codes 45
Table 5. Relationship between crime and software forecasts for first and second order contiguous grid cells in control districts only. 52
Table 6 Descriptive (during shift) district-day crime statistics for the property crime phase. 54
Table 7 Predicting crime count during treatment 57
Table 8 Predicting dichotomized property crime during treatment 59
Table 9 Descriptive (during shift) district-day crime statistics for the violent crime phase. 61
Table 10 Predicting property crime count in the eight hours post-treatment 64
Table 11 Predicting dichotomized property crime in the eight hours post-treatment 65
Table 12 Predicting violent crime count in the eight hours post-treatment 67
Table 13 Predicting dichotomized violent crime in the eight hours post-treatment 68
Table 14 District weekly property crime counts: January 1, 2014 – end of May, 2016 77
Table 15 Descriptive statistics, weekly district-level property crime counts, during property experiment 96
Table 16 During property experiment: discrepancies with control condition districts when predicting weekly counts 97
Table 17 Property experiment less first and last weeks: discrepancies with control condition districts when predicting weekly counts 99
Table 18 How much additional patrol time do you think your officers were able to provide? 103
Table 19 How busy were the predicted areas? 104
Table 20 How accurate were the predicted areas? 104
Table 21 What was the major activity in the predicted area during the shift? 105
Table 22 What was the major strategy in the predicted area during the shift? 106
Table 23 How many times did officers exit vehicle? 106
Table 24 How much additional patrol time do you think your officers were able to provide? 108
Table 25 How busy were the predicted areas? 109
Table 26 Based on your experience, do you think the predicted areas were in correct places? 110
Table 27 What was the major activity in the predicted area during the shift? 111
Table 28 How much additional patrol time do you think your officers were able to provide? 112
Table 29 How busy were the predicted areas? 113
Table 30 How accurate were the predicted areas? 114
Table 31 What was the major activity in the predicted area during the shift? 115

Table 32 What was the major strategy in the predicted area during the shift? 115
Table 33 How many times did officers exit vehicle?...116
Table 34 How much additional patrol time do you think your officers were able to provide? 118
Table 35 How busy were the predicted areas?...118
Table 36 Based on your experience, do you think the predicted areas were in correct places? 119
Table 37 What was the major activity in the predicted area during the shift?...120
Table 38 Field codes for location and activity...122

FIGURES

Figure 1 Deputy Commissioner Kevin Bethel (now retired) briefing PPD commanders on details of the Philadelphia Predictive Policing Experiment (April 21, 2015). 30
Figure 2 Dr. Jerry Ratcliffe briefing PPD mid-level commanders on details of the Philadelphia Predictive Policing Experiment (April 21, 2015). 31
Figure 3 Senior Analyst John Grasso and Director Kevin Thomas from the PPD field phone calls from police districts on the first day of the Philadelphia Predictive Policing Experiment. 32
Figure 4 Administrative Corporal from the 22nd District discusses the Philadelphia Predictive Policing Experiment on the first day of the study with Dr. Jerry Ratcliffe (off camera). 33
Figure 5 Professor Ralph B. Taylor leads a graduate researcher training session on field observations for the Philadelphia Predictive Policing Experiment. 43
Figure 6 Differing mission sizes. 49
Figure 7. District weekly reported property crime counts: Comparison to theoretical Poisson and negative binomial distributions 78
Figure 8 Reported district level weekly property crime counts 79
Figure 9 District weekly property crime counts: Quadratic and LOWESS smoothing results 81
Figure 10 Marked condition only: Weekly district-level reported property crime counts 83
Figure 11 Marked condition only: Global and local smoothed relationships between time and property crime counts 85
Figure 12 Unmarked condition only: Weekly district-level reported property crime counts 86
Figure 13 Unmarked condition only: Global and local smoothed relationships between time and property crime counts 87
Figure 14 Awareness condition only: Weekly district-level reported property crime counts 88
Figure 15 Awareness condition only: Global and local smoothed relationships between time and property crime counts 89
Figure 16 Control condition only: Weekly district-level reported property crime counts 90
Figure 17 Control condition only: Global and local smoothed relationships between time and property crime counts 91
Figure 18 Dedicated car time from field observations (property) 123
Figure 19 Dedicated car time from field observations (violent) 124

EXECUTIVE SUMMARY

Synopsis of the problem and the research purpose

We present the results of a randomized, controlled experimental field test of different policing strategies in response to potential crime locations estimated by predictive policing software. Not only has our aim been to study the crime reduction link between different theories of crime control and predictive policing, but a secondary aim has been to better understand the operational connections/disconnections between predictive estimates of crime locations and day-to-day operations in an urban area policed by a large metropolitan department. Since advances continue in the development of predictive policing software, this study evaluates three different policing responses.

With reference to geographic patrol operations, predictive policing is “the use of historical data to create a spatiotemporal forecast of areas of criminality or crime hot spots that will be the basis for police resource allocation decisions with the expectation that having officers at the proposed place and time will deter or detect criminal activity" (Ratcliffe, 2014b: 4). While predictive policing has been discussed in detail since the First Predictive Policing Symposium held in Los Angeles in 2009, it remains the case that “the policing aspect is mostly overlooked in evaluations of predictive policing” (Rummens, Hardyns, & Pauwels, in press: 7).

Rummens and colleagues go on to note that “aspects such as map usability, *police resource limitations and response strategies* have an important influence on the ultimate successful implementation of predictive policing” (in press: 7, emphasis added). This experiment has been cognizant of police resource constraints and the operational responses were designed by the Philadelphia Police Department to explore a number of response strategies that are sustainable into the long term.

The experiment explained in this report includes three different interventions and a control strategy of ‘business as usual’ as the counterfactual. Our most limited intervention is to make all available patrol officers in a district aware of the predicted crime locations, but provide them with no further direction. In other words, this ‘awareness’ model examines the impact of a change in awareness among the general patrol force. A more direct intervention is a dedicated marked patrol car in predicted crime and disorder areas. A theoretical groundwork for crime prevention in crime hot spots is our impetus for the use of uniformed patrol officers in marked police cars during this intervention. As Perry and colleagues (2013) point out, the theoretical foundations that allow for crime prediction are ably supported by routine activity theory (L. E. Cohen & Felson, 1979), the rational choice perspective (Cornish & Clarke, 1986) and crime pattern theory (Brantingham & Brantingham, 1981-2)—collectively known as opportunity theories.[1] The police response to these concentrations of crime is generally described by the hot spots policing literature (Braga, 2005; Weisburd & Braga, 2006). A Campbell Collaboration review reported that 80% of police interventions focused on crime hot spots "reported noteworthy crime and disorder reductions" (Braga, Papachristos, & Hureau, 2012: 6); however, a number of the interventions described requiring significant resource commitments from the police department, such as dedicated foot patrols of multiple officers for 16 hours a day (Ratcliffe, Taniguchi, Groff, & Wood, 2011), or sufficient directed patrol resources to 'saturate' a crime hot spot (B. Taylor, Koper, & Woods, 2011). The application of uniform–marked—patrol resources to predicted crime areas during this study was grounded in a more sustainable and realistic appreciation of the operational resource constraints that police chiefs confront on a day-to-day basis.

The research evidence regarding incapacitation of high-rate offenders motivates the use of plain-clothes/unmarked police units (our final intervention). It is well-established that a small group of offenders are responsible for a large proportion of the crime suffered by a community (Wolfgang, Figlio, & Sellin, 1972) with estimates that about 6% of the population are responsible for about 60% of the crime (Martinez, Lee, Eck, & O, 2017; Ratcliffe, 2016). Martinez and colleagues draw on a meta-analysis to determine that of the offending population, about 15 percent commit 50 percent of the crime (Martinez et al., 2017). The spatial context is also relevant: Johnson and colleagues determined that a spatial scale between meso and micro is appropriate for patrolling choices and that at this scale "such patterns are overwhelmingly a reflection of the activity of individual offenders" (Johnson, Bowers, Birks, & Pease, 2009: 177).

---

[1] Perry et al. (2013) consolidated these approaches into what they called *blended theory* but environmental criminologists usually refer to these as the *opportunity theories*.

Empirical evidence appears to support this (Bowers & Johnson, 2004; Short, D'Orsogna, Brantingham, & Tita, 2009), and some researchers have found that increasing *targeted* arrests (rather than an increase in random arrests) is tied to reductions in the crime rate (Chilvers & Weatherburn, 2001a, 2001b; Farrell, Chenery, & Pease, 1998).

Focused police operations that have deliberately employed arrest and incapacitation have been successful in reducing burglary (Farrell et al., 1998), drug activity (Nunn, Quinet, Rowe, & Christ, 2006), and gang crime (Kent, Donaldson, Wyrick, & Smith, 2000) though success depends on the ability of police systems to identify appropriate offenders (Townsley & Pease, 2002). An incarceration effect (when measured directly as opposed to via deterrence theory) was among the strongest correlates of macro-level crime reduction predictors when assessed in a 2005 meta-analysis (Pratt & Cullen, 2005)—at least at the macro level (Spelman, 2000). Therefore, focusing police attention on key recidivist offenders in predicted crime areas may be a viable tactic in a specific deterrence mode. Recent research in Philadelphia suggests significant benefits at a crime hot spot level, based on experimental results from a randomized field trial (Groff et al., 2015).

Across these three interventions we deploy resources using the HunchLab software from Azavea. Previous randomized studies that have examined the effect of police on crime hot spots focused on hot spots determined by analysis of a year of crime prior to the experiment (L. W. Sherman & Weisburd, 1995) or even longer (Ratcliffe et al., 2011; B. Taylor et al., 2011). Yet beyond these chronic hot spots, Gorr and Lee have recently shown that short-term, temporary hot spots (such as might be determined by predictive policing methods) can be more effective for crime prevention than large chronic hotspots, and provide for greater equity of crime reduction resources (W. L. Gorr & Lee, in press). HunchLab is able to incorporate both long-term and short-term factors in its predictions. The current study tackles two different types of high-volume street crime, pertinent because previous research has shown that street crime (Chainey, Tompson, & Uhlig, 2008) and high-volume crime patterns (W. Gorr, Olligschlaeger, & Thompson, 2003) are more predictable than other types of crime.

The goal of our study is to investigate the crime reduction link between certain hot spot policing tactics associated with predictive policing and guided by a predictive policing algorithm, and related concepts of crime control grounded in offender focus and opportunity theories. Specifically, we test whether greater awareness among general duties patrol officers of the predicted crime areas will be sufficient to deter crime, whether a dedicated uniform patrol attendance in predictive areas will increase visible police presence sufficiently in the local area to deter crime, or if dedicated plain-clothes units performing surveillance and unmarked patrol will increase interdiction and offender incapacitation sufficiently to reduce crime.

Research design and results

20 Philadelphia police districts were randomly assigned to one of four experimental conditions using block randomization with a 1:1:1:1 allocation ratio. These conditions include (1) a control condition, with a business-as-usual patrol strategy, (2) an *awareness* condition where officers were made aware of the predicted high crime activity areas at roll call, (3) an awareness model treatment enhanced with a dedicated *marked* patrol car with uniformed officers to exclusively patrol the predicted crime areas, and (4) an awareness model treatment supplemented with dedicated officers in an *unmarked* vehicle to exclusively patrol the predicted areas. This initial design was outlined and supported by the Philadelphia Police Department at the start of the experiment and there were no significant changes to the initial trial design.

This study took place in Philadelphia. Philadelphia is the 6th largest city in the country with a population of 1.5 million. The Philadelphia Police Department (PPD) is the nation's fourth largest police department, with over 6200 sworn members and 800 civilian personnel. While the city has dozens of police departments operating within its borders, the PPD is the primary law enforcement department responsible for serving Philadelphia County, conterminous with the City of Philadelphia, which extends over 140 square miles.

The property crime phase ran for 90 days from June 1, 2015 through August 25, 2015. A break was scheduled during the month of September while Philadelphia prepared for, and hosted, the World Meeting of Families (September 22, 2015 – September 27, 2015) and a visit from Pope Francis. Extensive road closures and increased levels of security required that the police department dedicate all of its available resources to this event. The study resumed on November 1, 2015 for the violent crime phase of the experiment which ran for 92 days through January 31, 2016.

As stated earlier, predicted crime areas were generated by the HunchLab prediction program, designed by Azavea. Starting with a grid of 500 foot by 500 foot cells projected across the entire city, the software forecast three of these grid cells as potential crime areas for each pertinent shift, each day, for each district. This was the same for both the violent crime and property crime phases of the experiment. To enable the experimental conditions, Azavea adapted the software to generate three predicted 500 feet square grids per district for each treatment shift. This allowed the project to create the necessary experimental conditions by limiting the output of the software to generate three predicted grids (500 feet by 500 feet) in every district of the city each day. It is important to note therefore that the experiment artificially reduced the efficiency of the software in three ways; 1) it forced the software to choose grids in low crime districts, 2) the program was constrained to assign only a limited number of grids in high crime districts, and 3) Azavea included a partial randomization component to reduce the possibility that the same grid cells were predicted for several days in a row.

Results show that the software predicted on average twice as much crime as one would expect if crime were spread uniformly across the spaces within each district. It did this even though artificially constrained by our experiment to be less effective than designed, both in the program's between-district and within-district selections.

Crosstabulation tables revealed significant variation across conditions for property crime but not violent crime. Mixed effects models of property crime in spatially buffered treatment areas find notably lower property crime counts and property crime presence in marked as compared to control condition districts during the treatment shift and the post treatment shift. Impaired statistical power associated with a floor effect, however, means these notable crime prevention benefits prove statistically non-significant.

When examining both mission grids and their contiguous areas, during shifts when the experimental treatments were scheduled to be deployed, the marked cars effected a reduction in 2.3 property crimes for every thousand days' worth of shifts. While not significant given impacts of floor effects on statistical power, this property crime reduction translates to a 31 percent reduction if examining counts of crime, or a 36 percent reduction in the number of cells experiencing at least one crime. In the eight hours after the property treatment, the marked car districts were associated with a reduced crime compared to the control areas, such that expected crime counts were 41.6 percent lower and expected crime occurrences were 48.1

percent lower. For violent crime, cross-tabulations showed no significant link between experimental conditions and violent crime.

A district level analysis was also conducted. A descriptive graphic analysis using data smoothing contrasting the global relationship between time and weekly district level property crime counts with the local relationship suggested that in the marked car condition the local estimation for district property crime counts deviated significantly below the global prediction during the weeks when the property crime experiment was in operation. In other words, this suggested that there may have been a district level dynamic at work in the marked condition that might have been affecting the entire district property crime levels during the experiment.

Statistical models (specifically cross-sectional panel design generalized estimating equation models) contrasting each treatment condition with the control condition during the experiment, and during the experiment focusing just on the weeks when it was most fully operational, tended toward supporting this scenario but again did not confirm it at the required levels of statistical significance. During the full property experiment, expected weekly property crime counts for the entire district were between three and six percent lower in the marked condition districts as compared to the control condition districts. If the first and last weeks of the property experiment are dropped out of the analysis, the expected weekly property crime counts in the marked condition were about eight percent lower compared to the control districts during the same timeframe. This contrasts with higher than expected crime counts in awareness versus control and in unmarked versus control districts.

Qualitative analysis identified a number of key themes related to operational strategy and predictive policing. The qualitative data highlighted the challenges inherent in innovation implementation within a major metropolitan police department. Beyond the institutional inputs which remain crucial to implementation, there is a clear need to address the buy-in and attitudes of front-line officers regarding the innovation. Some officers were supportive of the experiment; however, others felt that the study (and by implication the software program) constrained the freedom they had to patrol the district in the best way that they saw fit. Specifically, this research underscores the significance of unifying the craft-orientation of police with scientific and technological innovation and the strategic or organizational changes required for successful operationalization.

Conclusions and implications

Because of statistical limitations, we are unable to categorically state that the implementation of a dedicated marked car to property crime predicted micro-grid areas reduced crime; however, both the micro-level grid analysis and the broader district analysis are supportive of this implication. Reductions in property crime in grid areas in excess of 30 percent were identified. Furthermore, there is evidence of a temporal diffusion of benefits to the eight-hour shift immediately following the marked car, property crime district deployment. While property crime provided some challenges, violent crime is even more problematic. Violent crime levels are so low on an hour-by-hour basis at the street level, even in a relatively high crime city, that making confident inferences about a crime reduction process is extremely challenging.

Our implementation did cause some consternation among officers because the change in work practices at the street level had implications for the workload of officers across the district. This caused some resentment. Police leaders made considerable efforts to inform frontline officers about the experiment and its potential value for the police department, yet it remains the possibility that other departments in the future may be able to offset some of the officer antipathy with a different dissemination strategy.

We conclude that a marked police car targeted to a limited number of predicted crime areas is a promising crime reduction strategy, but police commanders should be realistic about the limited evidence they are likely to see and they may need to manage this strategy alongside other resource constraints and public and officer expectations.

INTRODUCTION

Statement of the problem

This document presents the design and results of a randomized, controlled experimental field test of different policing strategies in response to potential crime locations estimated by predictive policing software. Our aim has been to study both the crime reduction link between different theories of crime control and policing, and the operational connections/disconnections between predictive estimates of crime locations and day-to-day operations in an urban area and police department. Since advances continue in the development of predictive policing software, this study evaluates three different policing responses.

Predictive policing has been defined as “the application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions” (Perry et al., 2013: xiii). Given the specific spatial focus of much patrol policing, another definition more germane to street policing is “the use of historical data to create a spatiotemporal forecast of areas of criminality or crime hot spots that will be the basis for police resource allocation decisions with the expectation that having officers at the proposed place and time will deter or detect criminal activity" (Ratcliffe, 2014b: 4).

Whichever definition is used, it is not surprising that since the NIJ’s first predictive policing symposium in Los Angeles in 2009, a number of police departments have sought to move to the next evolution of patrol. Attendees of the technical breakout session at that meeting[2] identified numerous potential applications of predictive policing, the primary use being the *time and*

[2] For purposes of transparency it should be noted that one of the principal investigators of this report (Ratcliffe) was one of the attendees.

location of future incidence in a crime pattern or series.[3] While noting that numerous other applications of predictive policing exist, this study is primarily concerned with this first priority given its relevance to nearly every police department in the country as well as its pertinence to advancing our theoretical understanding of predictive policing and a related concept, hot spots policing.[4] Hot spots policing is relevant because studies have shown for a decade that "focused police resources on crime hot spots provided the strongest collective evidence of police effectiveness that is now available" (National Research Council, 2004: 250).

While clear that hot spots policing is important to crime control, unfortunately, the research is less clear on what specifically police officers should be doing in hot spots to most effectively reduce crime. Concepts such as the Koper Curve (Koper, 1995) or strategies such as focused foot patrol (Ratcliffe et al., 2011) may only work for long-term, chronic hot spots and may prove less effective at addressing dynamic day-to-day projections identified by modern predictive policing algorithms. If we consider predictive policing strategies to be related to hot spots policing, then two recent observations from Weisburd and Telep are relevant: 1) there are numerous strategies that have not yet been rigorously tested, and 2) much more needs to be learned about the impact of new technology on policing effectiveness (Weisburd & Telep, in press).

Our research grounds itself in the current fiscal reality of police departments across the country. It does not propose swamping predicted crime areas with a massive amount of resources (saturation patrol). Such an option is impossible to sustain on an ongoing basis for all but the largest departments. Instead, it asks whether crime can be reduced in predicted crime areas through greater patrol awareness or with the application of modest additional resources. Fiscal realities have fundamental relevance to police chiefs across the United States. Many agencies have seen reductions in budgets or reductions in personnel, leading them to seek greater benefits from patrol, and to maximize the value of any additional assets.

[3] http://www.nij.gov/topics/law-enforcement/strategies/predictive-policing/symposium/Pages/technical-breakout.aspx
[4] "Predictive policing…is hot-spot policing, significantly enhanced by technology" *[Bill Bratton, NPR interview (Around the Nation), November 26, 2011]*.

The research also answers an additional but closely related question. If modest additional resources are available, are they better deployed in uniform patrol – corresponding to a general ecological deterrence theoretical perspective – or in a plain clothes capacity – corresponding to an incapacitation theoretical perspective?

Literature review

This research is conducted at the intersection of two significant ideas in modern policing; first, how and why the spatial distribution of crime is (to a degree) predictable, and second whether policing strategies can be employed to act on those predictions and either deter criminal activity or capture and incapacitate offenders.

The case for a place focus

A theoretical foundation for crime prevention in crime hot spots is the impetus for the use of uniformed patrol officers in marked police cars during our experiment, while the research evidence regarding incapacitation of high-rate offenders motivates the use of plain-clothes/unmarked police units. We discuss first the predictability of spatial crime patterns.

Scholars have previously commented on the gulf between crime data mining techniques (similar to some predictive policing implementations) and criminological theory (Marshall & Townsley, 2006), but a significant body of research demonstrates that crime is unevenly distributed among places and victims (Felson, 1987; L. W. Sherman, Gartin, & Buerger, 1989; Weisburd & Eck, 2004), and a number of spatial theories of crime can explain short-term changes in crime risk for small, local areas (Chainey & Ratcliffe, 2005). That crime is concentrated in a few places is unsurprising, given that many phenomena are concentrated, not just crime (Eck, Lee, O, & Martinez, 2017). But in order to explain this, as Perry and colleagues (2013) point out, the theoretical foundations that allow for crime prediction are ably supported by routine activity theory (L. E. Cohen & Felson, 1979), the rational choice perspective (Cornish & Clarke, 1986) and crime pattern theory (Brantingham & Brantingham, 1981-2)—collectively known as opportunity theories.[5] While these theories explain predictable patterns of criminal behavior, they do not immediately suggest a specific type of police or crime prevention response.

Previous randomized studies that have examined the effect of police on crime hot spots focused on hot spots determined by analysis of a year of crime prior to the experiment (L. W. Sherman & Weisburd, 1995) or even longer (Ratcliffe et al., 2011; B. Taylor et al., 2011). Yet beyond these chronic hot spots, Gorr and Lee have recently shown that short-term, temporary hot spots (such as might be determined by predictive policing methods) can be more effective for crime prevention than large chronic hot spots, and provide for greater equity of crime reduction resources (W. L. Gorr & Lee, in press). There is a growing empirical foundation to support the notion that predictive policing approaches such as risk terrain modeling (Caplan, Kennedy, & Miller, 2011; Kennedy, Caplan, & Piza, 2011; Moreto, Piza, & Caplan, 2014), techniques that utilize short-term event patterns (W. L. Gorr & Lee, in press; Johnson et al., 2009) and processes such as the Epidemic Type Aftershock Sequence method (a nonparametric self exciting point process, see Mohler, Short, Brantingham, Schoenberg, & Tita, 2011) are all competing with traditional crime mapping approaches in the prediction of short-term crime hotspots. The current study tackles two different types of high-volume street crime, pertinent because previous research has shown that street crime (Chainey et al., 2008) and high-volume crime patterns (W. Gorr et al., 2003) are more predictable than other types of crime.

The police response to concentrations of crime is generally described by the hot spots policing literature (Braga, 2005; Weisburd & Braga, 2006). A Campbell Collaboration review reported that 80% of police interventions focused on crime hot spots “reported noteworthy crime and disorder reductions” (Braga et al., 2012: 6); however, a number of the interventions described require significant resource commitments from the police department, such as dedicated foot patrols of multiple officers for 16 hours a day (Ratcliffe et al., 2011), or sufficient directed patrol resources to “saturate” a crime hot spot (B. Taylor et al., 2011). The application of patrol resources to predicted crime areas during this study was grounded in a more sustainable and

[5] Perry et al. (2013) consolidated these approaches into what they called *blended theory* but environmental criminologists usually refer to these as the *opportunity theories*.

realistic appreciation of the operational resource constraints that police chiefs confront on a day-to-day basis.

The case for an offender focus

The theoretical foundation for crime prevention in crime hot spots is the impetus for the use of uniformed patrol officers in marked police cars during our experiment; however, the case for using unmarked police units is grounded in research on high-rate offenders. It is well-established that a small group of offenders are responsible for a large proportion of the crime suffered by a community (Wolfgang et al., 1972) with some estimating that 6% of the population are responsible for 60% of the crime (Ratcliffe, 2008). A recent meta-analysis found a robust range of studies that support this generalized conclusion (Martinez et al., 2017).

The spatial context is relevant: Johnson and colleagues determined that a spatial scale between meso and micro is appropriate for patrolling choices and that at this scale "such patterns are overwhelmingly a reflection of the activity of individual offenders" (Johnson et al., 2009: 177). Empirical evidence appears to support this (Bowers & Johnson, 2004; Short et al., 2009), and some researchers have found that increasing *targeted* arrests (rather than an increase in random arrests) is tied to reductions in the crime rate (Chilvers & Weatherburn, 2001a, 2001b; Farrell et al., 1998). Focused police operations that have deliberately employed arrest and incapacitation have been successful in reducing burglary (Farrell et al., 1998), drug activity (Nunn et al., 2006), and gang crime (Kent et al., 2000) though success depends on the ability of police systems to identify appropriate offenders (Townsley & Pease, 2002). An incarceration effect (when measured directly as opposed to via deterrence theory) was among the strongest correlates of macro-level crime reduction predictors when assessed in a 2005 meta-analysis (Pratt & Cullen, 2005)—at least at the macro level (Spelman, 2000). Therefore, focusing police attention on key recidivist offenders in predicted crime areas may be a viable tactic in a specific deterrence mode. Recent research in Philadelphia suggests significant benefits at a crime hot spot level, based on experimental results from a randomized field trial (Groff et al., 2015).

It is worth noting that the choice of policing strategy is not just tied to a theoretical foundation but also to operational capacity. Recent research in Philadelphia has shown that while there are chains of connected, near-repeat, street robbery events, they have a short half-life. This requires any police department to be extremely responsive if a predictive intervention is to be successful (Haberman & Ratcliffe, 2012; Ratcliffe & Rengert, 2008; Wyant, Taylor, Ratcliffe, & Wood, 2012).

Haberman and Ratcliffe (2012) concluded any organization seeking to invest in predictive policing needs to have a proficient surveillance and analysis mechanism to identify crime patterns as well as a capable decision-making framework that supports operational flexibility, a theme reflected in other work (Johnson et al., 2009). The current study tests both the results of the theory-driven response as well as the operational capacity strains on the Philadelphia Police Department.

Statement of hypothesis or rationale for the research

The goal of our study is to investigate the crime reduction link between certain hot spot policing tactics associated with predictive policing and guided by a predictive policing algorithm, and related concepts of crime control grounded in offender focus and opportunity theories. Because advances continue in the development of predictive policing software and there is growing law enforcement interest in the ideas surrounding predictive policing, this study is important to help guide appropriate policing responses. It also provides an experimental test of different theorized benefits of law enforcement intervention. The experiment was designed to test a number of hypotheses related to the implementation of strategies designed to conduct predictive policing.

> Hypothesis 1: Greater awareness among general duties patrol officers of the predicted crime areas will be sufficient to deter crime.

It may be that simply making the general duties patrol officers in a police district aware of the predictive crime areas during roll-call will be sufficient to increase the amount of time they spend in the areas such that potential offenders are deterred from committing crime during that tour of duty with resultant measurable levels of crime reduction. General duties patrol officers in Philadelphia are assigned to large patrol areas and are predominantly tasked with responding to calls for service from the public, supporting and backing up colleagues, and initiating proactive police work or general patrolling during times when they are not assigned other duties or tasks. During this experiment the relevant shift officers were made aware of the predictive policing grid areas during their initial roll-call at the start of their tour of duty.

> Hypothesis 2: A dedicated uniform patrol attendance in predictive areas will increase visible police presence sufficiently in the local area to deter crime.

The main difference between this and the first hypothesis is in the dosage, focus, and quality of the patrol intervention. In hypothesis 1, any additional patrol in the predicted areas stems from patrol officers assigned to response policing in the district paying additional attention to the predicted crime areas during the time when they are not assigned to response policing activities, such as radio calls or backing-up other officers. In this hypothesis (2) the same awareness model of making all officers on roll-call aware of the predicted areas is supplemented by reassigning one patrol vehicle to focus specifically on the predicted policing areas to the exclusion of most other police work. By employing a dedicated visible patrol to the predicted areas, this hypothesis tests whether an increase in dosage supported by a general patrol focus can significantly deter crime in the area of the predicted grids.

> Hypothesis 3: Dedicated plain-clothes units performing surveillance and unmarked patrol will increase interdiction and offender incapacitation sufficiently to reduce crime.

Rather than deter crime through visible patrol presence, it may be possible to reduce crime through the surveillance and interdiction, arrest and incapacitation of key offenders suspected of committing crime in predicted crime areas. While it has been estimated that there is little discernable benefit to incapacitation (Bhati, 2007), and notwithstanding the recognized conceptual and practical challenges in disentangling the effects of deterrence and incapacitation (Piquero & Blumstein, 2007), incapacitation efforts at the local neighborhood level are generally not well understood. The use of unmarked police vehicles provides officers with greater opportunity to observe criminal activity without being identified and as such increase the likelihood of catching an offender in the act.

> Hypothesis 4: The previously described interventions will cause temporal displacement and an increase in crime in the predicted areas in the hours following an intervention.

For the current experiment, each predicted crime estimate is valid for an eight hour shift, which leaves up to 16 hours before the next experimental treatment period. Because predicted crime areas could overlap to some degree from day-to-day, it is not possible to test for spatial displacement. Nevertheless, because of the time period where each district reverts to a business-as-usual state it is possible to establish if there is a detectable increase in offenses during this non-experimental time block, which would be suggestive of temporal crime displacement (Bowers & Johnson, 2003). Alternatively, there might be a reduction in crime, suggestive of a temporal diffusion of benefits from the intervention (Weisburd & Green, 1995).

METHODS

Experimental design

20 Philadelphia police districts were randomly assigned to one of four experimental conditions using block randomization with a 1:1:1:1 allocation ratio. These conditions include (1) a control condition, with a business-as-usual patrol strategy, (2) an *awareness* condition where officers were made aware of the predicted high crime activity areas at roll call congruent to hypothesis 1, (3) an awareness model treatment enhanced with a dedicated *marked* patrol car and uniformed officers to exclusively patrol the predicted crime areas consistent with hypothesis 2, and (4) an awareness model treatment supplemented with dedicated officers and an *unmarked* vehicle to exclusively patrol the predicted areas, as per hypothesis 3. This initial design was supported by the Philadelphia Police Department throughout the experiment and there were no significant changes to the initial trial design.

Study setting

This study took place in Philadelphia. Philadelphia is the 6th largest city in the country with a population of 1.5 million. The Philadelphia Police Department (PPD) is the nation's fourth largest police department, with over 6200 sworn members and 800 civilian personnel. While the city has dozens of police departments, the PPD is the primary law enforcement department responsible for serving Philadelphia County, conterminous with the City of Philadelphia, which extends over 140 square-miles.

The property crime phase ran for 90 days from June 1, 2015 through August 25, 2015. A break was scheduled during the month of September while Philadelphia prepared to host the World Meeting of Families (September 22, 2015 – September 27, 2015) and visit from Pope Francis. Extensive road closures and increased levels of security required that the police department dedicate all of its available resources to this event. The study resumed on November 1, 2015 for the violent crime phase of the experiment which ran for 92 days through January 31, 2016.

Participants

The Philadelphia Police Department comprises 22 police districts. Each district is commanded by a Captain and while each district fulfils a number of functions, the primary role of district personnel is patrol. While districts are aggregated for administrative purposes into divisions, and the divisions are also aggregated by two Regional Operations Commands (North and South), these are administrative alignments that are not germane to this project. One of these districts is the city's international airport, which was excluded from this study, leaving 21 eligible districts. In order to allocate the districts to four treatment conditions using a 1:1:1:1 ratio, the district with the lowest crime frequency in the assignment phase was dropped from the study. When districts were ranked using property crime, this excluded district number 7. When districts were rank ordered using violent crime, this excluded district number 5.

Interventions

Predicted crime areas were generated by the HunchLab prediction program, designed by Azavea. Starting with a grid of 500 foot by 500 foot cells projected across the entire city, the software forecast three of these grid cells as potential crime areas for each day for each district. This was the same for both the violent crime and property crime phases of the experiment. HunchLab is a web-based predictive policing system that accesses real-time Philadelphia police data to produce crime forecasts for the city. The crime prediction algorithm (Azavea, 2014) incorporates statistical modeling that combines components of risk terrain modeling (Caplan et al., 2011), near repeats, collective efficacy, and self-exciting point processes (see Mohler et al., 2011). Officers at police district buildings in all experimental conditions (except the control and excluded districts) could log in and print out maps of the forecast grids each day just prior to the experimental 8-hour shifts. Azavea adapted the software at the request of the Philadelphia Police Department and the academic researchers to generate three predicted 500 feet square grids per district for each treatment shift. This allowed the project to create the necessary experimental conditions by limiting the output of the software to generate three predicted grids (500 feet by 500 feet) in every district of the city each day. It is important to note therefore that the experiment artificially reduced the efficiency of the software in three ways. Firstly, it forced the software to choose grids in low crime districts. Further, it could assign only a limited number of grids in high crime districts. Finally, Azavea also included a slight randomization component to reduce the possibility that the same grid cells were predicted for several days in a row.

Each district in a matched block of four was randomly assigned to one of four conditions; control, awareness, marked, and unmarked.

In '**control'** districts, police personnel did not have access to the crime prediction software. The predictive policing algorithm calculated three predicted target areas for each shift but these predicted areas were not available for any district personnel to view. Absent any knowledge of the predicted areas, the target areas received the standard police patrolling and response in that district as determined by local command staff and officers.

In '**awareness'** districts, officers were informed at roll-call of the predicted target areas for that shift and were asked to pay attention to the areas when they were able, within the constraints of their operational duties. In other words, if they had spare time away from radio calls and other duties, they were instructed to patrol the areas. Perry *et al.* note in regard to predictive policing that "*situational awareness* among officers and staff is a critical part of any intervention plan" (2013: xviii, emphasis in original). This experimental condition represents a commonly-practiced patrol activity, especially in cities where resources are constrained. Philadelphia police leadership confirmed this patrol strategy as the most viable approach they would likely implement in the long term. The experiment did not impose any requirement to pay additional attention to the grids, nor did we dictate any order of grid attention or other constraints on these districts.

In '**marked'** car districts, in addition to having officers at roll-call aware of the predicted areas, two (though occasionally one) officers were assigned to a single marked vehicle creating a dedicated patrol of the predicted crime areas. This patrol was not assigned calls for service or other operational roles. They were given the opportunity to support other officers attending relevant calls for service in their prediction areas if the call related to the type of crime for the experiment, and they were permitted to leave their patrol zones should colleagues require immediate assistance.

In ‘**unmarked’** car districts the operational implementation was the same as in marked car districts with the exception that in these districts the officers always used an unmarked vehicle and were often dressed in plain clothes as well.

For the marked and unmarked car districts, officers were instructed to remain in the grids as much as possible, and to move between them as necessary during the tour of duty. The instructions to officers did not dictate any particular order of attendance to the three grids, nor instructed for how long and how frequently they attended the grids.

The experiment was conducted in two phases; a property crime phase and a violent crime phase. The property crime phase ran for 90 days from June 1, 2015 through August 25, 2015. The violent crime phase started on November 1, 2015 and ran for 92 days through to January 31, 2016. The property crime grid predictions were active from 8:00am to 4:00pm, and the violent crime predictions were active for an eight hour period from 6:00pm to 2:00am. District experimental assignment in the property crime phase did not necessarily carry over to the second phase. A totally separate block randomization of districts for the violent crime phase took place after the property crime phase was completed.

Operationalization of the intervention

Information regarding the experiment was disseminated as follows. The Deputy Police Commissioner responsible for patrol policing held a command briefing (Figure 130), a briefing that also featured one of the academic researchers (Ratcliffe, see Figure 2) for the all available mid-level and senior district level leaders.

Figure 1 Deputy Commissioner Kevin Bethel (now retired) briefing PPD commanders on details of the Philadelphia Predictive Policing Experiment (April 21, 2015).

Furthermore, instructional information was sent to each district, and two short instructional videos of a few minutes were made available to both crime analysts responsible for accessing the crime predictions and personnel assigned to marked and unmarked cars. Ratcliffe also wrote an article for the PPD newsletter; however it was never disseminated because the newsletter project apparently fell into hiatus.

Figure 2 Dr. Jerry Ratcliffe briefing PPD mid-level commanders on details of the Philadelphia Predictive Policing Experiment (April 21, 2015).

In the first day of the property phase of the experiment, some districts were not able to get online to the HunchLab software, due to some confusion with the logins and passwords. In anticipation of possible problems, academic researchers and analysts from the police department were available during the first week to field phone calls from the districts (Figure 3) or to directly attend districts in person to answer questions from district personnel (Figure 4).

Figure 3 Senior Analyst John Grasso and Director Kevin Thomas from the PPD field phone calls from police districts on the first day of the Philadelphia Predictive Policing Experiment.

Some of the problems encountered included;

- No clarity as to what the dedicated cars were supposed to do, sometimes because district command staff had not clearly conveyed instructions to their subordinates;
- Unable to log into the system because the login names included an underscore, which caused some confusion as to how to include this in the login;
- Difficulty accessing dedicated unmarked cars;
- Some Sergeants were reluctant to comply with the orders regarding the experiment as these orders originated from headquarters and had not been confirmed to them directly by their Captain.

We were able to convey additional information about the experiment through our field researchers who conducted observations throughout the experiment and accompanied police officers in the marked and unmarked cars on the ride-alongs.

Figure 4 Administrative Corporal from the 22nd District discusses the Philadelphia Predictive Policing Experiment on the first day of the study with Dr. Jerry Ratcliffe (off camera).

Outcomes and data collection

The research design incorporated both process and outcome evaluation strands. Because this article focuses on the outcome evaluation, we concentrate here on the quantitative crime data available. All recorded police activity and crime incidents in Philadelphia are geolocated with the incident type. Outcome indicators were daily crime counts in the corresponding categories in the predicted areas during the prediction period for each district. The research question we addressed was, *did treatment conditions reduce, relative to control districts, the instances of crime in the predicted areas during the tour of duty covering the prediction period (an eight hour shift)?* All pertinent data were drawn from the police department's incident database. This incident database is the city's record of officer-initiated activity, and recorded crime and calls for service that have been verified by a police officer.

Two outcome variables are examined, one for each experimental phase. For the property crime phase UCR part one property crimes, excluding arson, were included in the property crime analysis (specifically UCR codes 500-725). For the violent crime phase, only part one violent crimes were included (specifically UCR codes 111-116, 211, 231, 300-315, 350-399 and 411-416).

Other data collection relevant to measures of dosage involved officer record keeping. Officers assigned to dedicated vehicle patrol prediction areas (both marked patrol and unmarked cars) completed daily logs that were forwarded to the research team. The written logs (n=749 during the property crime phase, n=413 during the violent crime phase) were collected on a regular basis and were cross-checked for accuracy against the administrative data tied to the officers in relation to their pedestrian stops, arrests and other indicators of activity. These logs included details of officer activity in the predicted target areas, amount of time on site, officer observations, and other pertinent information.

Additional outcomes

During the experiment, trained researchers accompanied officers on just over 100 dedicated car assignment patrols (either marked or unmarked cars). It was clear from these observations that it was either impossible or unrealistic for officers to exclusively patrol the 500 feet square predicted grid areas. In Philadelphia the average street block length is a little more than 400 feet, so a grid cell often comprised not much more than a street and a couple of intersections. Beyond necessary travel between the grids during a shift, in many cases the grids encompassed one-way streets, necessitating driving through the surrounding street blocks to return to the grid assignment. In other cases, it was clear that for unmarked cars to remain entirely within the grid cells would have 'burned' the cars and given away the presence of police. Because the officers had to navigate the surrounding streets to return to the assigned grids, we estimated that any intervention effect would likely impact the streets immediately surrounding the predicted grids as well as the grid locations themselves. Grid cells immediately adjoining predicted grids were therefore included in our analysis of the experimental outcomes. We do this for the grids nearest to each predicted grid (in technical terms, *first order queen contiguity*).

In an additional related analysis, we examine the efficacy of the software and its ability to predict crime. This examination focuses on the control districts only, since the control sites, barring some type of wide-ranging contamination effect, would not have been affected by focused police activity resulting from the experiment. Stated differently, the control districts are a good indication of the efficacy of the software program absent any experimental impact. Looking at the predicted grids, and those grids combined with immediately adjoining grids, the spatial concentration of crime is compared to what would be expected if crime were distributed uniformly across all space in a district. The software is considered effective to the extent that the predicted areas, and their immediate surrounds, end up hosting a higher fraction of district crimes for the shift in question compared to their fraction of district area."

Randomization

Blocked experimental designs have been around for close to a century and are generally credited to statistician R. A. Fisher (Box, 1980; Fisher, 1935). If treatments are randomly assigned to units that have been grouped into blocks of units that are somewhat similar to one another on a factor likely to influence the outcome, analyses can separate the influence on the outcome of these other factors taken into account with the blocking from the influence of the treatment itself. "The advantages of block designs were not apparent until, with the coming of the analysis of variance, it became possible to isolate the variance to be ascribed to block differences and see how it could be eliminated in the analysis" (Box, 1980: 4).

Medical researchers conducting randomized clinical trials have found that blocked designs could cope with shifts over time in the composition of potential subjects available for assignment to treatment or control conditions, or studies drawing potential subjects from a variety of centers (Lachin, Matts, & Wei, 1988). As a result, "In clinical trials ... the *permuted-block* procedure ... commonly known as blocked randomization, has been widely employed" (Lachin, 1988: 290).

Permuted-block randomization is a special case of stratified randomization (Lachin et al., 1988). In stratified randomization the researcher stratifies the potential subjects by a covariate or combination of covariates thought likely to influence the outcome, and then randomly assigns persons or units within a particular range of scores on the covariate. For studies with relatively small sample sizes stratified randomization may be desirable to insure balanced scores on covariates across different treatment conditions. “With unstratified randomization the probability of covariate imbalances decreases as the sample size increases” so such imbalances are of potential concern when small numbers of units are randomly assigned (Lachin et al., 1988: 290). In the current study, the probability of covariate imbalances merit attention because there are less than two dozen police districts available for assignment. In criminal justice research, blocked randomization has been proposed as one way to partially address some known limitations of randomized trials. “If x is suspected of altering how the treatment affects the response, and if x is measured as part of the experiment, x`s role can be examined. In perhaps the most straightforward manner, block randomization can be employed and an average treatment effect can be estimated separately within each block” (Berk, 2005: 425).

Criminal justice researchers using spatial units of analysis like police districts, drug markets or hot spots have been drawn to blocked randomization for two reasons. First, it elevates the likelihood that random assignment even with small numbers of spatial units will achieve initial probabilistic equivalence across treatment groups, or across treatment and control groups. Given concerns about failing to achieve initial probabilistic equivalence through randomization, meta-analysts trolling through criminal justice experiments archived with the Campbell Collaborative have argued for a minimum of at least 100 (treatment plus control) units (Braga, Welsh, Papachristos, Schnell, & Grossman, 2014: 3). Consequently, they have questioned the contribution of a number of experiments using organizational or spatial units which number well under a hundred.

In light of the potential benefits of blocked randomization, however, the stance of the Campbell Collaborative researchers may be softening. Braga et al. (2014: 3) have acknowledged that blocked randomization may lessen the need for such a minimum number of study units and say requiring 50 minimum control units and 50 minimum treatment units “is not a hard-and-fast rule.” The softening stance of the Campbell Collaborative researchers on the minimum number of units needed is partly in response to Gill and Weisburd’s (2013) re-analyses of Jersey City drug market experiment data, plus simulated data. In light of their re-analyses Gill and Weisburd (2013: 159) concluded that “even with a limited number of cases, block randomized studies can achieve high levels of pretest equivalence and improve statistical power.” Improved statistical power would be the second reason to use blocked randomization designs.

Block randomization, however, carries potential penalties. Depending on how the blocking is done and what the blocking variable is, there may be non-independence within the blocks. If so, this would require statistical adjustments (Boruch et al., 2004: 613). A non-zero intragroup correlation coefficient would require an analysis recognizing the blocking (Matts & Lachin, 1988: 330). A sizable intragroup correlation coefficient seems more likely when blocks are based on time-linked segments of the population of units recruited (Matts & Lachin, 1988).

The current study employed block randomization for two reasons: to increase initial probabilistic equivalence across the four groups of districts on crime harm, and on community demographic factors strongly linked to crime. The current project investigates impacts of predictive policing strategies on short term crime changes. Previous work in Philadelphia has demonstrated the impacts of community socioeconomic status (SES), and percent white, non-Hispanic population on crime changes taking place within the next year (Ralph B. Taylor, Ratcliffe, & Perenzin, 2015). So one purpose of block randomization here was to increase the chances that the districts in the different conditions would be comparable on these background factors.

A second purpose was to insure that at least one instance of each treatment condition, including the control condition, would occur at different ranges of current and recent crime harm levels (Ratcliffe, 2014a). For the first experiment concerned with property crime, property crime harm levels were used rather than property crime rates. Constructing crime harm levels involves, in part, weighting crime occurrences by their associated gravity, as reflected for example in sentencing guidelines (see for example Ignatans & Pease, in press; Ratcliffe, 2015a, 2015b; L. Sherman, Neyroud, & Neyroud, 2016; L. W. Sherman, 2007).

Demographic indicators were used to create equivalence in demographic factors across like districts. Census variables capturing socio-economic status, race, and the total residential population were used in the blocked design. The specific variables are listed in Table 1. These variables were used to create variables measuring SES, race and the total residential population for each police district.

Table 1 Census variables that contributed to the blocked design.

Unique ID	Variable description
B19001001	Total population of households
B19001002	Household income less than $10,000
B19001003	Household income $10,000-$14,999
B19001004	Household income $15,000-$19,999
B19001011	Household income $50,000-$59,999
B19001012	Household income $60,000-$74,999
B19001013	Household income $75,000-$99,999
B19001014	Household income $100,000-$124,999
B19001015	Household income $125,000-$149,999
B19001016	Household income $150,000-$199,999
B19001017	Household income greater than $200,000
B25077001	Median home value
B19013001	Median income
B03002003	White non-Hispanic population
B03002001	Total population

The SES index included four variables: percentage households reporting income less than US$20,000; percentage households reporting income greater than US$50,000; median house value (natural logged after adding 1); and median household income (natural logged after adding 1). Each variable was z-scored and then averaged to create the SES index; higher scores indicate higher SES.

In addition to the demographic measures, a crime harm index was also used in the block randomization process. First, each crime incident within each crime category was weighted by a gravity score; this was done for property crime as well as for violent crime. These gravity scores were informed by the Pennsylvania Sentencing Guidelines. The median sentence length from the repeat felony offender category was used as a multiplier to weight serious offenses more heavily than less serious offenses. Then, the gravity-weighted crime total across all crime types was multiplied by 100,000, and divided by the total resident population. Three-year crime harm rates were calculated by weighting recent crime more heavily than older crime, using the following weighting assignments (Table 2):

Table 2 Temporal weighting periods for pre-intervention analysis.

40% for crime within last 6 months
30% for 7-12 months prior
20% for 13-24 months prior
10% for 25-36 months prior

Using multiple years of crime data helped stabilize crime levels so that anomalous crime rates in a single period would not prove overly influential. At the same time, the most recent period for which crime data were available were weighted most heavily since this was the period for which the most recent pre-treatment data were available. Further, the two six month periods in the last year of the pre-treatment time frame together contributed to slightly less than three quarters (70%) of the crime rate index.

To create one stratification variable, each district's ranks on the relevant variables (SES index, percent population white non-Hispanic, crime gravity) were summed, counting the crime gravity rank twice. Each variable was constructed so that a higher score meant either lower crime, or higher SES, or a higher percentage of the population which was white and non-Hispanic. Districts were then ordered by this summed rank variable. This was done separately for property crime and violent crime.

For both crime type phases, one district was dropped from consideration to create allocation equity across experimental conditions. The chosen districts had the lowest summed rank and was excluded from the experiment. The remaining ordered 20 districts were separated into five blocks of four districts. A uniform random number was generated for each district in each block. Within each block of four districts the random assignment rules were as follows: lowest random number was assigned the *marked* car treatment group, the second lowest number was assigned to the *unmarked* car group, the third lowest number was assigned to the *awareness* group and the highest number was assigned to the *control* group. This resulted in one awareness, one uniform, one unmarked, and one control district within each block of four districts. In this way the 20 police districts were block randomized to the four conditions using a 1:1:1:1 allocation ratio.

Notes on the crime harm gravity

For each of the four different periods described in Table 2 on page 39 above, each crime incident within each crime category was weighted by a gravity score. These gravity scores were informed by the Pennsylvania Sentencing Guidelines (http://www.courts.phila.gov/pdf/criminal-reports/Sentencing-Guidelines-Matrix.pdf). The median sentence length from the repeat felony offender category was used as a multiplier to weight serious offenses more heavily than less serious offenses. Then, within each period, within each district, the gravity-weighted crime total, across all crimes, was multiplied by 100,000, and then divided by the total resident population. No correction was made for the 6 month time periods when dividing by population. The total residential population was used as the denominator for each time period. Each of the time periods were then differentially weighted (.1, .2, .3, and .4) as was done with the crime rates.

To make sure that the crime rates and gravity score 'rates' were rank ordering police districts in a similar way, an ordinal correlation test was run (Table 3). The results showed that the 3 year weighted average calculated with crime rates was extremely strongly correlated with the 3 year weighted average calculated with gravity rates. The strong correlation held regardless of whether the three years were simply averaged, the differential weighting favoring more recent periods was used, or an average based on 6 equal time periods (six month intervals) was used.

Table 3: Kendall's Tau correlation test results for equivalency

Method	Kendall's Tau
Equal weighting for 2012, 2013, 2014	.876
40%, 30%, 20%, 10% breakdown	.876
6 month intervals	.876

The final gravity-weighted crime score for property crime was constructed differentially weighting periods as had been done with crime:

> Weighted property crime = (2012 gravity rate*.1) + (2013 gravity rate*.2) + (2014 Jan-June gravity rate*.3) + (2014 July-Dec gravity rate*.4)

Gauging initial probabilistic equivalence across assignment conditions

To learn whether the block randomization procedure used created initial probabilistic equivalence on SES, racial composition, and gravity-weighted property crime scores, we used bootstrapped regression, because “When we use methods with unknown small-sample properties, bootstrapping provides methods for approximating sampling distributions” (Hamilton, 1992: 314). Since we just have 20 districts, bootstrapped regression provides a defensible way to estimate initial probabilistic equivalence across assigned conditions.

The procedure was as follows. Three dummy variables were constructed reflecting assignment to either awareness, uniform, or unmarked conditions. This left the control condition as the reference category whose score on the outcome was reflected in the constant. These were the predictors of interest. The *b* weight for each predictor reflected its discrepancy from the average for the control condition on the outcome.

Several outcomes were examined. Most important was the sum of the ranks on gravity weighted property crime (x 2) + SES + racial composition. Then, the effects of the assignment conditions on each individual ranking were examined (crime rank, SES rank, racial composition rank), as well as their effects on the gravity weighted property crime score. Finally, effects of assignment on SES index scores and percent white non-Hispanic were examined.

Five thousand bootstrapped samples of 16 districts each were requested. Usually, regression bootstrapped estimates were available on (5000-219=) 4,781 bootstrapped samples. Hamilton (1992: 314) recommends at least 2,000 bootstrapped samples if confidence intervals are to be estimated.

From the bootstrapped regression results, we examined, for each outcome, whether the confidence intervals for each dummy variable predictor overlapped with the confidence interval for each other dummy variable predictor, and for the constant. Overlap indicated no significant differences across the four assignment conditions. All means for different conditions, for every outcome, overlapped sizably with the control mean score, and mean estimates for different non-control conditions overlapped sizably with one another. For the dummy variables, z values were usually less than |.5|.

We can therefore say that the block randomization procedure, based on a combined stratification of gravity weighted property crime, SES, and white non-Hispanic racial composition, has created initial probabilistic equivalence for these factors across the four conditions.

Concealment and blinding

Because the prediction software is web-based and requires a login and password, we were able to electronically prevent access to the software platform for the control districts when they were acting as controls for a particular experimental phase. We were not able to conceal the existence of the experiment across the police department (indeed to achieve buy-in it had been widely discussed), but only a few individuals had access to the logins for each district and there is no evidence of control district personnel attempting to access crime predictions for their district.

Qualitative fieldwork

During the course of the experiment, graduate research assistants and the primary investigators of the project were employed as field researchers to collect observations during ride-alongs with officers who were assigned to patrol the grid areas for the entirety of their 8-hour shift. A total of 8 researchers were utilized as observers in the field over the entirety of the experiment. Field observations and interviews conducted during ride-alongs were used to assess the implementation of the different patrol strategies and monitor treatment integrity. The observation shifts were determined via random assignment for both the property crime and violent crime phases of the experiment. The field observations comprised of two elements; structured and systematic detailing of officer patrol behavior which was noted on a form in 15-minute increments and the use of open-ended ethnographic field notes. This form was filled out every 15 minutes by the researcher, with multiple options to describe the police activity (At station, car outside grids, car inside grids, car chat with citizens, on break, incident within the grid, incident responded to outside the grid, other). Furthermore, each form recorded the date, district, car type (uniform or unmarked), number of officers, and space to detail any other pertinent information relevant to the police activity on the shift. For the open-ended ethnographic field notes, the researchers were instructed to observe the actions of the officers and discuss topics related to the experiment and the software grid placement. Prior to the beginning of the experiment, a training session was held to discuss potential questions and themes to be mindful of during the ride-alongs (Figure 5).

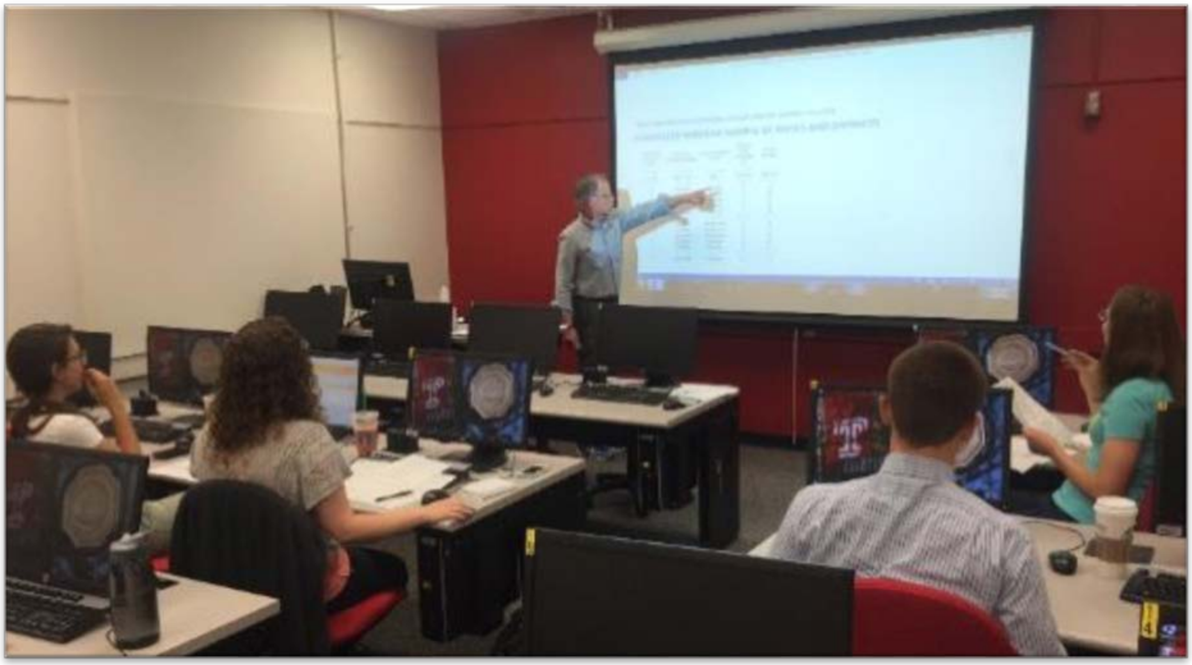


Figure 5 Professor Ralph B. Taylor leads a graduate researcher training session on field observations for the Philadelphia Predictive Policing Experiment.

In particular, observers were told to keep some of the following themes in mind during the shift: activity level of police, particular strategies employed during patrol, the level of overall activity in the grid areas, citizen interaction, reaction to the grids, and any unusual events that took place. Researchers were allowed to explore these themes either through direct questioning of the officers or supervisors encountered on the shift or simply via the natural course of dialogue about the experiment with the officers and supervisors.

The field observers recorded jottings while in the field which were then synthesized upon the completion of the ride-along. In total, there were 101 individual field note entries that were completed through both phases of the experiment. There were 79 of them originating from the property crime portion, and the remaining 22 from the violent crime portion. In addition, each researcher constructed a final write-up, addressing the main themes, lessons, and take-aways from their ride-alongs.

We should note that the significant difference between the numbers of field observations for the two phases was caused by a serious incident that took place in the city unrelated to the experiment. At 11.40pm on Thursday January 7th, 2016, Edward Archer shot Officer Jesse Hartnett while he was in his patrol car at the intersection of 60th and Spruce Street in West Philadelphia. Archer told investigators that he targeted an officer because police defend laws that are contrary to the Quran. Fearing possible copycat violence or additional incidents, police leadership decided to curtail the ride-along program for the remainder of the experiment.

The analysis of the field notes was primarily informed by the 'grounded theory' approach of Strauss and Corbin (1998). This was undertaken in a two-stage process whereby two researchers independently read through and descriptively coded the field notes in the program ATLAS.ti. These codes were then further refined through direct comparison and focused discussion before taking this revised coding list and applying it to the full set of field notes. The first 10 documents were coded by each researcher independently to determine the level of interrater reliability. Once this was established, researchers took a random set of 10 documents and coded these independently before being revised again. Each researcher took this set of final revised codes and applied them to the final 81 field note documents. Finally, one researcher took the final coded documents and explored the main themes and relationships between them, sometimes distilling them into overarching conceptual ideas. Table 4 on page 45 shows the final codes used in the analysis along with the total number of individual codes made by each researcher and the independent prevalence of the codes that were not coded by both researchers.

Table 4 Final qualitative field codes

Codes	Researcher A	Researcher B	Both Coded	Total	Prevalence
Answering Calls/ Normal Duties	29	55	14	84	70
Buy-In	17	32	7	49	42
Departmental Skepticism	36	75	23	111	88
Deterrence	6	3	0	9	9
Friendly interaction	25	43	23	68	45
Interference with 'regular duties'	7	26	2	33	31
Manpower Issues	12	35	7	47	40
Miscommunication	29	45	19	74	55
Police 'craft'	46	123	25	169	144
Recommendations	15	36	15	51	36
Respectful Assertion of Authority	34	22	8	56	48
Software Problems, Information Quality	27	51	16	78	62
Time Spent in Grid	56	55	39	111	72
Timeliness, Usefulness, Ease of use	8	36	2	44	42
Unmarked Cars		28		28	28
Visibility or Presence in unpatrolled areas		7		7	7

STATISTICAL METHODS

Sample size

The sample in this study is termed the 'district day'—that is, an eight-hour shift during each day nested within districts. For each experimental condition there were five districts across a three-month experimental phase. This equates to 450 district days (five districts x 90 days = 450) in each property crime experimental condition and 460 days for the violent crime phase (five districts x 92 days = 460). Numerous policing experiments have employed similar sample sizes, such as the Jersey City problem oriented policing at violent crime hotspots experiment (Braga et al., 1999) and the Lowell Massachusetts disorder policing experiment (Braga & Bond, 2008). Weisburd and Gill have recently demonstrated that block randomization methods provide significantly stronger equivalence than simple randomization, in terms of absolute mean differences between treatment and control groups on pre-intervention dependent variables and other factors related to outcomes. Limiting the treatment to one high crime 8-hour shift within a 24-hour period (as we do here) has precedent in numerous significant and influential randomized controlled experiments (L. W. Sherman & Weisburd, 1995) and it allows the treatment to be focused at the most appropriate time.

Cross-tabulation

For each experimental phase, crosstabulations and descriptive data on crime counts provided a preliminary assessment of links between each experimental condition and their effects on crime in the predicted areas. The assessments considered just the three mission areas per district for each day's treatment shift for each experimental condition. Four crosstabulations were conducted for each experimental phase (property crime and violent crime). In the first, treatment condition (columns) were crossed with the numbers of district days with different numbers of reported crimes. Each data point represented the daily count of crime for the shift in question in each district's combined three mission areas. In the second crosstabulation, the predicted grid areas were spatially expanded to include first order queen contiguous grids cells. These are the neighboring cells that adjoin the mission cell when that mission area is buffered out (we refer to these cells *as spatially buffered* below). In the third and fourth crosstabulations, these two tables were rebuilt but just using crime presence vs absence (two rows) rather than crime counts. Finally, to investigate temporal displacement (hypothesis 4), another set of four crosstabulation tables were created using the eight-hour time period (post shift) immediately following the treatment shift.

For the property crime phase, during the treatment shift, all four crosstabulations generated significant Likelihood Ratio χ^2 values (count/mission grids, $p = .03$; dichotomized/mission grids, $p=.03$; count/spatially buffered, $p =.005$; dichotomized/spatially buffered, $p = .001$). For the post-shift period, significant Likelihood Ratio χ^2 values appeared for the tables using spatially buffered counts ($p=.008$), and spatially buffered dichotomized data ($p=.001$). So for the property crime phase of the experiment: crime counts, buffered crime counts, dichotomized crime, and buffered dichotomized crime all varied significantly across conditions during the treatment shift. For the post shift, only spatially buffered crime counts or dichotomies varied significantly by condition. These cross-tabulations confirm that there is *some* type of connection between condition and crime for this phase. Multivariate models will specify those connections more closely.

For the violent crime phase of the experiment, none of these tables generated significant Likelihood Ratio χ^2 values. These results indicate that during the violent crime phase of the experiment, reported crime did not vary by experimental condition. This means the variables are independent; the experimental conditions were unrelated to crime both during the treatment shift and the eight-hour post shift. Given this lack of any overall relationship, with either crime counts, dichotomized data, in either spatially buffered or unbuffered form, either during the shift or immediately following, we do not provide detailed reports on the multivariate models for violent crime.

Intention to treat

Note that because crimes reported during the entire eight hour shift are included in the treatment period, even though officers were not always actively patrolling during the entire time, the approach here is tantamount to an 'intention-to-treat' analysis, which is the best way to "obtain an unbiased estimate of the effect of selecting one treatment over another" (Detry & Lewis, 2014: 85) in a randomized control trial. It also reflects the reality of the difference between policing intent from headquarters and policing reality of actions on the ground.

Multivariate analysis of contiguous areas

The analysis just explained was repeated for the within-district areas contiguous to the three predicted grids, so that grid cells adjoining mission grids were also treated as a unit of analysis. Unfortunately, the cross-tabulation results with the adjoining areas are potentially misleading for the following reason. Even though the stand-alone mission areas are all of the same size, with three 500' x 500' grid cells in each district each day, the collective size of the mission areas plus the adjoining cells is not constant. This is because first order queen contiguity adjoining areas could potentially overlap if the mission grids were near each other or conterminous. Furthermore, it is possible that the original mission grids could be close to a boundary with another district (Figure 6). Because the officers usually operated with strong operational discipline and remained in their district, any adjoining cells that were in other districts were not included in our analysis. Therefore, even though each mission cell would normally have eight adjoining cells surrounding it, in reality, exclusion or overlap varied the actual number of actively-examined cells in each district.

To address this limitation with the adjoining cell data, mixed effects count models using the number of grid cells (the three original mission grids plus a varying number of contiguous cells) as an exposure variable are run. In addition, mixed effects logit models with the adjoining areas are run as well, and include the number of actively-examined grid cells as a predictor. These models control for random effects (random variation) across districts, recognizing that days are nested within districts.

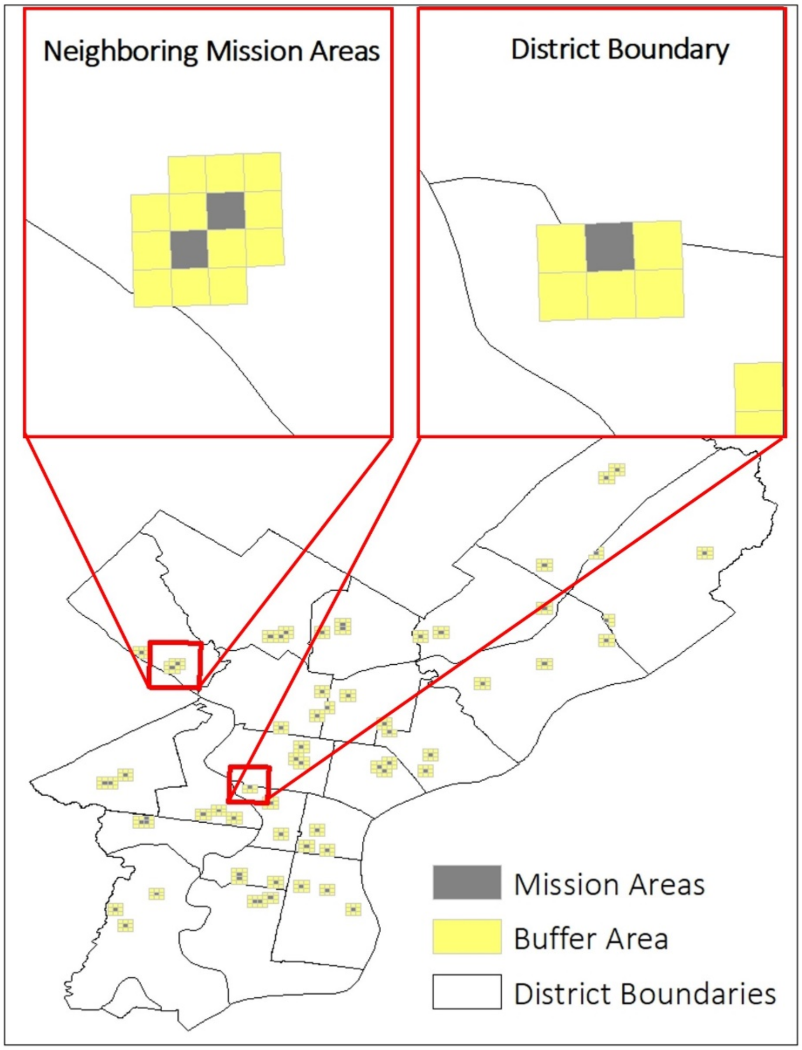


Figure 6 Differing mission sizes.

Because the number of grid cells in these combined (target and contiguous) mission areas varied across days and districts, the number of grid cells was included as an exposure variable in the models. In the multivariate analysis, this shifts the interpretation of the constant to the expected crime count per grid cell (predicted mission grid or contiguous cell), rather than per (three grid cell) mission area.

Additional analyses

One additional analysis is included in this study, to address the question of the efficacy of the software to predict crime. As explained earlier, we examined the efficacy of the software using the untainted control districts and considering not just the predicted grids, but also (and more realistically) the surrounding contiguous grids as well. We do this for the grids nearest to each predicted grid (in technical terms, *first order queen contiguity*), and for grids within two cells of each predicted grid (*second order queen contiguity*). In the results that follow we report descriptive statistics that examine the ratio of crimes identified in forecast areas compared to a baseline of the crimes being distributed uniformly across each district. This is the first analysis we present.

RESULTS: MICRO-LEVEL QUANTITATIVE FINDINGS

Software forecast efficacy

As explained earlier, the analytical unit is the district-day, i.e. the combined three mission grids in each district per day for the single shift in question. But as mentioned above, experimental conditions on the ground suggest that it was reasonable to additionally include the areas immediately adjoining the mission grids in our analysis. To estimate the efficacy of the software absent any experimental influence, we explore the results for control districts only. For the property crime phase there are 90 experimental days and five control districts, creating 450 single shift district days. With 92 days in the violent crime phase, the number of single shift district days increases to 460. As shown in Table 5 for the control sites, 274 of the potential 450 total district days (60.1%) had at least one crime in the *entire* district during the shift. For property crime in these 274 district days—extending out from the original mission grids to the first order contiguous grids—the software forecast nearly 14 percent of the crime available to be forecast in these districts would take place in just 6.8 percent of the area of the five districts. When extended up to two grid cells away from the predicted grids, it predicted 30.3 percent of the property crime while highlighting only 15.3 percent of the area of the district. It is important to reiterate that these forecasts were not optimized – the software was deliberately hobbled to enable the experimental design. The results are therefore conservative as estimates of how well the software predictions could have worked. Also note that the results differed by district and crime type, as shown in the table where district identifiers have been replaced by letters. In summary, the software was able to predict twice as much crime as we would expect if crime were spread uniformly across the districts, even when artificially constrained by our experiment to be less effective than designed.

Table 5. Relationship between crime and software forecasts for first and second order contiguous grid cells in control districts only.

Phase / district	District days	Crime days in district[1]	Total crimes[2]	Within one grid (% total)[3]	% area of district[4]	Within two grids (% total)[5]	% area of district[6]
Property	450	274 (60.1%)	489	68 (13.9%)	6.8%	148 (30.3%)	15.3%
A	90	61 (67.7%)	127	11 (8.6%)	5.1%	31 (24.4%)	11.4%
B	90	46 (51.1%)	64	12 (18.7%)	6.1%	22 (34.4%)	12.9%
C	90	74 (82.2%)	157	22 (14%)	5.7%	34 (21.6%)	14.2%
D	90	59 (65.5%)	92	15 (16.3%)	12.9%	38 (41.3%)	28.6%
E	90	34 (37.7%)	49	8 (16.3%)	2.6%	23 (46.9%)	4.9%
Violent	460	137 (29.8%)	176	24 (13.6%)	5.1%	43 (24.4%)	12.1%
F	92	49 (53.2%)	68	7 (10.3%)	5.8%	18 (26.5%)	13.9%
G	92	32 (24.4%)	40	4 (10%)	5.9%	6 (15%)	14.4%
H	92	19 (20.6%)	20	4 (20%)	7.4%	10 (50%)	16.2%
I	92	27 (29.3%)	37	8 (21.6%)	2.6%	8 (21.6%)	6.5%
J	92	10 (10.8%)	11	1 (9.1%)	1.4%	1 (9.1%)	3.6%

Note. Individual districts referenced by letter.

[1] Number of days when a crime took place in that district during the eight-hour patrol shift (8am-4pm for property crime phase, 6pm-2am for the violent crime phase).

[2] Total number of crimes that happened in the district during the eight-hour patrol shift during the entire experimental phase (90 days for property phase, 92 days for violent phase).

[3] Number of crimes that happened inside the predicted mission area or first order queen contiguous cells during the eight-hour patrol shift.

[4] The total area covered by the mission areas plus first order queen contiguity cells, calculated using the total area of the district.

[5] Number of crimes that happened inside the predicted mission area or second order queen contiguous cells during the eight-hour patrol shift.

[6] The total area covered by the mission areas plus second order queen contiguity cells, calculated using the total area of the district.

Property crime phase descriptive statistics

Table 6 shows property crime phase descriptive statistics for the areas comprising mission grids and then mission grids plus contiguous first order queen contiguous grid cells (spatially buffered), across all experimental conditions. The table indicates both raw counts of the number of crimes in each condition, as well as the *dichotomized* crime counts reflecting the number of district-days with any crime for the shift in question (the collapsed crime counts).

Table 6 Descriptive (during shift) district-day crime statistics for the property crime phase.

		Experimental Condition	Count	Min	Max	Sum	Mean	SD
Three mission grids only								
	Count							
		1 Control	450	0	1	15	0.033	0.180
		2 Awareness	450	0	1	21	0.047	0.211
		3 Marked	450	0	2	7	0.016	0.141
		4 Unmarked	450	0	2	14	0.031	0.186
		Total	1800	0	2	57	0.032	0.181
	Dichotomized							
		1 Control	450	0	1	15	0.033	0.180
		2 Awareness	450	0	1	21	0.047	0.211
		3 Marked	450	0	1	6	0.013	0.115
		4 Unmarked	450	0	1	13	0.029	0.168
		Total	1800	0	1	55	0.031	0.172
Mission grids plus contiguous grid cells								
	Count							
		1 Control	450	0	2	68	0.151	0.388
		2 Awareness	450	0	2	91	0.202	0.449
		3 Marked	450	0	2	50	0.111	0.342
		4 Unmarked	450	0	4	106	0.236	0.544
		Total	1800	0	4	315	0.175	0.440
	Dichotomized							
		1 Control	450	0	1	63	0.140	0.347
		2 Awareness	450	0	1	82	0.182	0.386
		3 Marked	450	0	1	46	0.102	0.303
		4 Unmarked	450	0	1	85	0.189	0.392
		Total	1800	0	1	276	0.153	0.360

Table 6 highlights how infrequently crime events occurred in the mission areas during the eight-hour treatment period. For all the day shifts in question, over the entire 90-day treatment period, across mission areas in all 20 districts, 57 property crimes were reported. Of these 57 crimes, 53 were solo district-day events, that is, they were the only crime reported in that district's mission areas on that day during the shift examined. There were only two district-days when two property crimes co-occurred in the same district on the same day during the shift in question. Of the 1800 available district-days, 1,745 (96.9%) district-days were crime-free in mission areas in a district for the shift in question, 53 (2.94%) district-days had one crime, and on two district-days there were two crimes in the mission areas (0.11%). Examining the dichotomized data, the chances that any property crime would occur appeared to vary across treatment condition. The chances were lowest (1.3% chance of any property crime per district-day for shift in question) for the marked condition, and highest (4.7% chance of any property crime per district day for shift in question) for the awareness condition. Bear in mind that counts from non-control areas may be susceptible to any intervention effect.

Examining the mission areas plus the contiguous grid cells (lower half of Table 6) increased the total number of property crimes about five times and the total mission area 6.9 times. The average contiguous mission areas in a district contained 20.74 cells (minimum = 8, maximum = 27; median = 21) compared to three cells for all the stand-alone mission areas. Property crimes during the shift in the contiguous mission areas totaled 315 compared to 57 in the unbuffered areas. The ratio of contiguous/stand-alone mission crime counts is roughly comparable to the ratio of contiguous/stand-alone areas (lower half of Table 6). In these 'mission grids plus contiguous grid cells' for the shift in question, there were about 17 property crimes per 100 district days (mean=0.175). In other words, on a daily basis in a district's mission and contiguous areas there was about a 17.5 percent chance that a property crime would be reported during the shift in question. The district day per shift average crime count varied by condition, ranging from a low of 11 property crimes per 100 district days in the marked condition for the treated shift (mean=.11) to a high of 24 per 100 district days in the unmarked condition (mean=.24) for the treated shift.

Property crime phase multilevel models

A null mixed effects negative binomial model, with only a random effect for districts, predicted crime counts during the treatment period for the mission areas plus their contiguous neighboring grid cells. The exposure variable is the natural log of the number of grid cells in the buffered mission areas. Results confirmed ($p < 0.05$) significant variation in expected counts across districts. The constant, after being exponentiated and converted to an incident rate ratio (IRR), reflects the expected number of reported property crimes, *per grid cell*, in a typical grid cell in a typical day in a typical district for the shift in question. The IRR generated by this model of 0.00721 anticipates that in mission-plus-contiguous areas, considering all 20 districts, on average about 7.2 reported property crimes were expected per cell per 1,000 district day shifts, or about one reported property crime per 2.7 years of daytime shifts in a typical cell in a typical spatially buffered mission area in a typical district.

The two right-most columns in this table show the results after adding the three dummy variables for awareness, marked, and unmarked conditions. The IRR associated with the marked condition (.69) indicates that expected crime counts were 31 percent lower in the buffered mission areas in the marked condition. This sizable impact translates to an expected crime count of (exp(-.4905 + -.271)) = .0051. This expectation of 5.1 crimes per thousand days' worth of day shifts in a typical grid cell in a typical mission area in a typical marked district contrasts with the expected count of 7.4 crimes in the same period in a typical grid cell in a typical mission area in a typical control district.

Table 7 Predicting crime count during treatment

	b	IRR	b / (se) / t	IRR
Predictor				
Awareness condition			0.110	1.117
			(0.373)	
			0.296	
Marked condition			-0.371	0.690
			(0.381)	
			-0.972	
Unmarked condition			0.171	1.187
			(0.373)	
			0.459	
Exposure (ln(number of cells)	1		1	
Constant (control condition)	-4.932	0.00721 (all)	-4.905	0.00741
ln_alpha	-2.129		-2.135	
Random effects				
Variance at district level	0.334		0.271	
(se of var)	(0.139)		(0.120)	
t	2.407		2.248	
Observations (district days)	1,800			
Number of groups (districts)	20			

Note. Results from mixed effects negative binomial model. Units = district days, one shift only. District days grouped within districts. Outcome = property crime counts. Counts summed across all spatially buffered mission areas in a district in a day. Period = during treatment shift (day shift). Since the control condition is uncoded, and the exposure reflects individual grid cells, the constant in the full model, In IRR (incident rate ratio) form, reflects the expected crime count, *in a typical individual grid cell in the control condition during a typical day shift*

With a dichotomized crime outcome, the frame shifts to the predicted probability that one or more reported property crimes will take place in a grid cell. See Table 8.

The left most columns in Table 8 report the results of a mixed effects logit model including only a random effect for districts. The outcome variation across districts proves significant ($p < .05$). The proportion of mission area grid cells where one or more crimes are expected is .0077, or 7.7 grid cells out of a thousand in a typical district on a typical day for the shift in question.

When the predictors for the three treatment conditions are entered, the constant (in odds ratio form) converts to the expected proportion of **control** district grid cells expected to have one or more reported property crimes during the shift examined. That expected proportion is .0081 or 8.1 grid cells with one or more expected crimes in a typical control district over a thousand district-days looking just at the treatment shift.

The impact of the marked condition (OR=.64) is to reduce the expected reported property crime count **by 36 percent**. This translates to an expected odds ratio of (exp(-4.811 + - .439)) = .00525 which translates to an expected proportion of (proportion = (OR/(1+OR)) = .00522 mission grid cells in a typical marked district experiencing one or more property crimes on a typical day for the shift in question. Stated differently, over a thousand days in a typical marked district five mission grid cells would be expected to experience at least one reported property crime during the shift in question.

Albeit sizable, and of considerable practical import, the contrast between marked condition districts and control condition districts on dichotomized crime counts fails to reach conventional levels of statistical significance (t=-1.03).

Table 8 Predicting dichotomized property crime during treatment

Predictor	b	OR	b / (se) / t	OR
Awareness condition			0.143	1.153
			(0.421)	
			0.340	
Marked condition			-0.439	0.644
			(0.429)	
			-1.025	
Unmarked condition			0.0945	1.099
			(0.422)	
			0.224	
Exposure (ln(number of cells))	1		1	
Constant (control condition)	-4.865	0.00771	-4.811	0.00814
Expected proportion		0.007651		0.008074
Random effects				
Variance at district level	0.417		0.347	
(se of var)	(0.173)		(0.152)	
Observations (district days)	1,800			
Number of groups (districts)	20			
Note. Results from mixed effects logit model. Units = district days, one shift only (day shift). Level 2 units = district. Outcome = dichotomized crime counts, (0 = none; 1 = 1 or more). Counts summed across all spatially buffered mission areas in a district in a day. Period = during treatment shift (day shift). Since the control condition is uncoded, and the exposure reflects individual grid cells, the constant in the full model reflects the predicted log odds, or odds (OR), of one or more crimes happening vs. no crimes happening, *in a typical individual grid cell in the control condition on a typical day shift.*				

Violent crime phase descriptive statistics

Table 9 shows violent crime phase descriptive statistics for the areas comprising mission grids, and again for mission grids plus contiguous first order queen contiguous grid cells, across all experimental conditions. The table indicates both raw counts of the number of crimes in each condition, as well as the dichotomized counts indicating the number of district-days with any crime (the collapsed crime counts).

Table 9 Descriptive (during shift) district-day crime statistics for the violent crime phase.

Experimental Condition	Count	Min	Max	Sum	Mean	SD
Three mission grids only						
Count						
1 Control	460	0	1	2	0.004	0.066
2 Awareness	460	0	3	9	0.02	0.180
3 Marked	460	0	2	9	0.02	0.154
4 Unmarked	460	0	3	10	0.022	0.197
Total	1,840	0	3	30	0.016	0.157
Dichotomized						
1 Control	460	0	1	2	0.004	0.066
2 Awareness	460	0	1	7	0.015	0.123
3 Marked	460	0	1	8	0.017	0.131
4 Unmarked	460	0	1	7	0.015	0.123
Total	1,840	0	1	24	0.013	0.113
Mission grids plus contiguous grid cells						
Count						
1 Control	460	0	4	24	0.052	0.283
2 Awareness	460	0	3	38	0.083	0.346
3 Marked	460	0	3	35	0.076	0.311
4 Unmarked	460	0	3	27	0.059	0.286
Total	1,840	0	4	124	0.067	0.307
Dichotomized						
1 Control	460	0	1	20	0.043	0.204
2 Awareness	460	0	1	30	0.065	0.247
3 Marked	460	0	1	30	0.065	0.247
4 Unmarked	460	0	1	23	0.05	0.218
Total	1,840	0	1	103	0.056	0.230

As with the property crime phase, data in Table 9 are summed across mission grids (and then mission and contiguous spatially buffered grids) in each individual district by condition on a daily basis for the treatment shift in question. Given that the violent crime experimental phase was running at a time of year that generally experiences lower crime (late fall and early winter) and given that reported property crime occurs more frequently relative to violent crime, a larger mean contrast was both expected and experienced. Over the 92-day experimental period, across all 20 districts in all four conditions, a total of 30 violent crimes were recorded, averaging about one and a half violent crimes per 100 district days' worth of evening shifts (mean=.016). This mean presents an interesting contrast with the mean property crime count (.032) in (noncontiguous) mission grids. And the dichotomized data show that in the mission areas, over 98 percent of district days were free of violent crime (1,816 out of 1,840).

Compared to the control districts, all of the experimental treatment districts experienced more violent crime. In the three treatment conditions the total number of recorded violent crimes in the mission areas were roughly equivalent, nine each in awareness and marked mission areas, and ten in unmarked mission areas for the shift in question. This contrasted with a total of two reported violent crimes in all the control condition districts' mission areas. The average number was about five times larger in the three treatment conditions (means between .020 and .022 in the three treatment conditions) compared to the control condition (mean = 0.004).

The picture is less pronounced with the missions-plus-contiguous-areas descriptive statistics, but the bottom line remains: crime in the control condition remained less likely than in any of the experimental sites. The expansion in violent crime counts was slightly smaller in magnitude, lifting the number of reported violent crimes from 30 to 124 as we switch from only mission areas to mission-plus contiguous—a fourfold expansion. This would seem to rule out spatial displacement of violent crimes from buffers to the immediately surrounding grids, and might even suggest some diffusion of benefits. This is suggested because the crime expansion is only about four times while the spatial expansion is almost seven times. These contiguous mission areas during a district day contained anywhere from 9 to 27 grid cells, averaging 20.4 (median = 21). Inclusion of contiguous cells expanded the number of available intervention grid cells 6.8 times, up to 37,535 from 5,520.

In the mission plus contiguous areas there were about 6.7 reported evening shift violent crimes per 100 district days across treatment and control conditions. The average was slightly lower in control districts (5.2 per 100 district days) and unmarked districts (5.9 per 100 district days), and slightly higher in marked (7.6 per 100 district days) and awareness (8.3 per 100 district days) districts. Switching to dichotomized data, the pattern of differences across different conditions was about the same. The overall mean across all districts and all conditions was about 5.6 district days with one or more evening shift reported violent crimes per 100 district days, with the average in awareness and marked conditions both averaging 6.5 district days per 100 district days with one or more reported evening shift violent crimes, placing them slightly above the overall average. By contrast the averages in the control (4.3 per 100 district days) and unmarked conditions (5 per 100 district days) were slightly below the overall average. The control condition outperformed all treatment conditions.

Property crime post-treatment effects

The modeling strategy adopted for property and violent crime during the experimental 8-hour periods was also used for the eight hours immediately following the experimental shift. For property crime, this time block referred to 4pm to midnight – the eight-hour period following the 8am to 4pm treatment period.

Starting with counts (Table 10), the random effects model shows significant variation across districts on counts ($p < 0.05$), and an expected count of 0.00508 in a typical mission-or-contiguous grid cell in a typical district on a typical day on a typical evening shift. Therefore, over a thousand district days' worth of evening shifts (4pm to midnight), one would typically expect 5 crimes within these areas over a thousand days. The model with the predictors for treatment condition transforms the (exponentiated) constant into an expected reported property crime count in a typical mission area grid cell in a typical control district of 0.0051 for a typical post-treatment shift. In the marked condition the predicted property crime counts were (1-0.584=) 41.6 percent lower. Therefore, in a typical marked district mission-or-contiguous area grid cell on a typical day the expected reported property crime count was 0.003 for the post-treatment shift.

Table 10 Predicting property crime count in the eight hours post-treatment

Predictor	b / (se) / t	IRR
Awareness condition	0.0976	1.102
	(0.386)	
	0.253	
Marked condition	-0.537	0.584
	(0.408)	
	-1.317	
Unmarked condition	0.0869	1.091
	(0.387)	
	0.225	
Exposure (ln[number of cells])	-5.282	0.00508
Constant (all / control condition)	0.0976	1.102
ln_alpha	-0.111	
Random effects		
Variance at district level	0.255	
(se of variance)	(0.125)	
Observations	1840	
Number of groups	20	

Turning to property crime absence/presence models (the dichotomized data model in Table 11), the random effects only model merely hinted ($p < 0.1$) at variation across districts in this dichotomized outcome. In a typical mission-or-contiguous grid cell 0.47 percent of these cells would expect one or more property crimes in a typical district on a typical day for the post treatment time frame.

When variables reflecting different treatment conditions were entered, the proportion of control district areas expected to have one or more property crimes on a typical day was 0.00513 (0.51%). In the marked districts, the odds of one or more crimes (compared to no crimes) was 48.1 percent lower. This translated to an expected proportion of 0.0027 mission-or-contiguous grid cells in the marked condition having one or more reported property crimes in a typical marked district on a typical day in the post treatment shift.

Turning to the question of statistical significance, the contrast between the marked and control condition difference on presence/absence of property crime during the post treatment time frame for mission-or-contiguous areas proves statistically significant with a one tailed test (t = -1.317; p (one tailed) = 0.043). This statistically confirms a temporally displaced diffusion of benefits for the marked treatment in the property crime experiment.

Table 11 Predicting dichotomized property crime in the eight hours post-treatment

Predictor	b / (se) / t	IRR
Awareness condition	0.124	1.132
	(0.354)	
	0.350	
Marked condition	-0.656	0.519
	(0.382)	
	-1.717	
Unmarked condition	0.188	1.207
	(0.353)	
	0.532	
Exposure (ln[number of cells])	1	
Constant (all / control condition)	-5.273	0.00513
Expected proportion	0.005104	
Random effects		
Variance at district level	0.186	
(se of variance)	(0.105)	
Observations	1800	
Number of groups	20	

In sum, the marked car experimental condition was associated with notably lower expected property crime counts during the post-treatment period. The temporally lagged diffusion of benefits resulted in expected crime counts that were 41.6 percent lower and expected crime occurrences that were 48.1 percent lower. Although these reductions are noteworthy, they do not achieve statistical significance. Again, this is partly due to the extremely low base rate of reported property crimes in grid cells in the examined areas. But if the focus shifts to property crime presence/absence, results confirm a statistically significant ($p = 0.043$, one tailed) temporal diffusion of benefit when marked and control districts' mission-and-contiguous areas are contrasted.

Violent crime post-treatment effects

The violent crime post-treatment examined crime counts in the period from 2am to 10am. Though there were occasions when the treatment period in the vehicle districts ended early (due to mechanical failure, arrest, or other operational conditions), this time period reflects the eight hours after the 'intention to treat' period.

Again, starting with counts (Table 12), we find an expected count of 0.001 in a typical mission-or-contiguous grid cell in a typical district on a typical day on a typical night shift. Therefore over a thousand district days' worth of night shifts (2am to 10am), one would typically expect only one crime within these areas over a thousand days. The model with the predictors for treatment condition transforms the (exponentiated) constant into an expected reported violent crime count in a typical mission area grid cell in a typical control district of 0.0051 for a typical post-treatment shift.

Table 12 Predicting violent crime count in the eight hours post-treatment

Predictor	b / (se) / t	IRR
Awareness condition	0.444	1.558
	(0.489)	
	0.907	
Marked condition	0.487	1.627
	(0.484)	
	1.006	
Unmarked condition	0.186	1.204
	(0.513)	
	0.363	
Exposure (ln[number of cells])	1	
Constant (all / control condition)	-6.925	0.0010
ln_alpha	2.257	
Random effects		
Variance at district level	0.0393	
(se of variance)	(0.192)	
Observations	1840	
Number of groups	20	

Table 13 Predicting dichotomized violent crime in the eight hours post-treatment

Predictor	b / (se) / t	IRR
Awareness condition	0.653	1.921
	(0.484)	
	1.348	
Marked condition	0.618	1.855
	(0.484)	
	1.276	
Unmarked condition	0.234	1.263
	(0.524)	
	0.446	
Exposure (ln[number of cells])	1	
Constant (all / control condition)	-7.158	0.0008
Expected proportion		0.0008
Random effects		
Variance at district level	0.0369	
(se of variance)	(0.151)	
Observations	1840	
Number of groups	20	

Summary of quantitative results

The software spatially predicted on average twice as much crime as one would expect if crime were spread uniformly across the grid cells within each district. It did this even though artificially constrained by our experiment to be less effective than designed, both in the program's between-district and within-district selections.

Crosstabulation tables reveal significant variation across conditions for property crime but not violent crime. Mixed effects models of property crime in spatially buffered treatment areas find notably lower property crime counts and property crime presence in marked as compared to control condition districts during the treatment shift, and during the post treatment shift. Impaired statistical power associated with a floor effect, however, means these notable crime prevention benefits prove statistically non-significant.

When examining both mission grids and their contiguous areas, during shifts when the experimental treatments were scheduled to be deployed, the marked cars effected a reduction in 2.3 property crimes for every thousand days' worth of shifts (results displayed in Table 7). While not significant given impacts of floor effects on statistical power, this property crime reduction translates to a 31 percent reduction if examining counts of crime, or a 36 percent reduction in the number of cells experiencing at least one crime. In the eight hours after the property treatment, the marked car districts were associated with less crime compared to the control areas, such that expected crime counts that were 41.6 percent lower and expected crime occurrences that were 48.1 percent lower.

For violent crime, cross-tabulations showed no significant link between experimental conditions and violent crime. In light of that, and in light of an even more extreme floor effect problem, detailed mixed effects results are not reported.

A note on floor effects

The floor effect refers to a well-known threat to internal validity (Cook & Campbell, 1979; Jones, 1990). It is an implementation difficulty that can crop up in either true experiments or quasi-experiments. The floor effect can become a problem "when numbers start out very low [as in the control condition] … and you expect for the numbers to go lower as a result of an experimental treatment [the three treatment conditions here], [but] it is hard for them to go lower…." (Ralph B. Taylor, 1994: 281). Thus, there is a practical concern. It is hard to drive very low numbers even lower.

But there is also a statistical concern. The concern can be framed in terms of a test of differences in proportions. Here, the proportions reflect crime free vs. crime afflicted district days at the grid cell level, and how those two proportions differ if we are comparing marked treatment to control districts. Simply put, statistical power—one's chances of detecting a statistically significant difference of a specified effect size at a specified alpha level—is compromised at extremely low proportions. These low proportions arise from the small spatial and temporal scales used here, and the resulting low crime base rates. These low base rates create the floor effect.

For example, take the expected proportion of day shifts with one or more reported per grid cell property crimes in mission-plus-adjacent grids, in the control condition: .008. The corresponding expected proportion in marked districts was .005.

Scholars recommend conducting a priori statistical analyses to gauge how large a design is needed. Such a priori estimation, however, requires available base rate information. *These numbers were not available a priori because no studies of which we were aware provided such detail.* Thus, an a priori statistical power analysis was not feasible.

Now that these two proportions are known, we can ask, focusing just on control vs. marked districts, with 900 district days total between these two conditions, what was the level of statistical power obtained in this analysis? G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) calculations ($p < .05$, two tailed, because the intervention could have made things worse), ignoring the multilevel nature of the data, show that the statistical power given these proportions was .025. This means that the chances of finding a 'real' and statistically significant difference between these two proportions given these parameters was 2.5 percent. If one did 100 exact replications of this experiment, this difference would prove statistically significant in two to three of those replications.

Typically, at least in psychology, one seeks statistical power of at least 80 percent, which corresponds to a Type II error rate of .20 (J. Cohen, 1988). Clearly, power here has been severely compromised.

If the control condition expected proportion was .24, and the thirty percent reduction corresponded to an expected crime proportion of .17 in marked districts, then with the current design statistical power would have been .71, close to the recommended minimum power level of .80. But returning to the observed proportions of .08 and .05; how many district day observations of individual cells (districts x days x mission-plus-contiguous cells) would we have needed to obtain an acceptable level of statistical power? We would have needed 11,920 cells in control district days and 11,920 cells in marked district days.

If we switch from individual mission cells to entire mission-plus-contiguous predicted high crime areas, the control district predicted areas experienced one or more property crimes on 14 percent of district days, and the marked predicted (mission-plus-contiguous) areas experienced one or more property crimes on 10.2 percent of district days. For acceptable power of 80 percent, we would have needed 1,201 district days in the control condition, rather than 450, and the same number, 1,201 in the marked condition. Given our 450 district days for each of these two groups, and the same difference in percentage district days with crime in these two conditions, effective statistical power was 37.8 percent. To get the power we “needed” would have meant convincing the Philadelphia Police Department to run the experiment for more than twice as long. For a range of implementation, research management and integrity, and organizational reasons, such a lengthy treatment period is extremely unlikely.

These extremely low base rates, and the attendant floor effect and low statistical power problems, arise in substantial part because the predictive policing scholarship is driving researchers to smaller and smaller geographical scales. This was explained in the introduction to the report. The floor effect/adverse statistical power impacts of such a micro focus represent a hitherto un-noticed adverse side effect of such a spatially constrained focus. When combined with a temporally appropriate, but also narrow time frame, like a shift, these adverse effects prove substantial.

In short, blocked randomized designs can increase statistical power (Gill & Weisburd, 2013), but narrowing the geographic scale of treatment target areas; focusing on realistic time frames like per shift crime reductions; and realistic limitations on the number of feasible target areas in a district; and on the length of time a large police organization can sustain a true experiment, have combined sizable adverse effects on statistical power. Low power, as is well known, means researchers cannot find significant differences even if those differences are sizable and “really there.” Fuller consideration of these tradeoffs merits attention.

RESULTS: DISTRICT LEVEL QUANTITATIVE FINDINGS

In this chapter we explore the weekly crime counts across **each entire district** before, during, and after the experiment. In doing so it frames the impact question in a different way. The question is whether there may have been district level patterns of impacts counterpointing the mission area results reported earlier in this report. The previous analyses suggest the marked condition associated with noticeably lower (albeit not always statistically significant) reported property crime counts within the mission areas and buffered mission areas. It seems worth exploring whether the entire district might have benefitted from participating in this condition, or any of the other treatment conditions. Possible relevant district level dynamics merit discussion below.

The following part of the study begins with a descriptive investigation that uses a full year and a half of weekly data for the period *prior* to initiating the property experiment, and for several months afterwards as well. This longer time frame permits gauging the global relationship between the passage of time, measured in weeks, and district level property crime levels, over a substantial time frame of over two years. This *global* relationship can be contrasted with a *local* relationship between time and district property crime levels *during* the time the experiment was operative.

Underpinnings

Ongoing temporal trends

Contrasting global relationships between district property crime and the passage of time with the local relationship when the experiment was operative is useful from a research methods vantage. Such a contrast attends to the potential threat to internal validity of history (Ralph B. Taylor, 1994: 275) whereby “an event other than the treatment… may have been responsible for the change observed." A staunch advocate of randomized control trials might counter that such concerns are irrelevant given the randomization protocol. In our case, this would be true if the experiment involved a large number of randomized units; however, given only five districts were assigned to each condition, situating the experimental timeframe in a broader temporal framework seems cautious and appropriate.

Auto-regression

As will be seen shortly, graphical examination does suggest local relationships existed between time and district-level weekly crime during the property experiment that were different from the global relationship. Subsequently, focusing just on the weeks the experiment was operative, statistical models are employed to investigate further. Because the focus is now on crime counts for entire districts, and because these correlate from week to week, serial auto-regression of errors merits consideration.

Previous scholarship leads to the strong expectation that a district's weekly property crime count is likely related to its crime count the week prior. Extensive work on near repeat dynamics at smaller spatial scales like hotspots (Townsley, Homel, & Chaseling, 2003), and 500 x 500 foot grid cells (Ralph B. Taylor et al., 2015), if it applies at the district scale, suggests this possibility as does the broader work in the area of time series (McCleary, McDowall, & Bartos, 2017).

Different analytics can be applied to handle auto-regression; for example correlated autoregressive residuals can be allowed, or temporally lagged outcome scores can be included as predictors (Twisk, 2013: 107-117). Twisk (2013: 116) recommends comparing results from different models since different models analyze "different parts of the longitudinal relationships." Extensive modeling was done using two analytic frames here: mixed effects generalized estimating equation models (GEE), and mixed effects negative binomial models. GEE permits directly modeling the autoregressive error structure. Mixed effects negative binomial models require adding temporally lagged outcome scores as predictors. The report provides results from the GEE models in detail. The results from the mixed effects models provided closely comparable coefficients and identical patterns of significance/non-significance. Given the relatively direct relationship expected between the treatment and the outcome, reporting the GEE models controlling for autoregressive errors would seem more appropriate (Twisk, 2013: 117). It is those we report in detail in the tables which follow.

Temporal and spatial scaling and theory

Community criminology continues to be bedeviled by important questions about the most appropriate temporal units to use and the best spatial units to use given a particular purpose (R. B. Taylor, 2015). These are important for policy and practice as well as theory. Shifting either the selected temporal or spatial scale for analysis can result in markedly different observed patterns. This can occur for theoretical reasons as well as simple mechanical quantitative reasons associated with aggregation or disaggregation of numbers (R. B. Taylor, 2015: 94-96).

Results for the property crime experiment have shown crime prevention impacts of the marked condition on property crime counts in spatially buffered mission areas. As was commented earlier, extremely low property crime counts given a restrictive timeframe – one eight-hour shift –and a restrictive spatial domain – a small number of 500 x 500 foot grid cells – each constrain the ability of any experiment to show a sizable crime reduction. Switching to a district level focus helps with those constraints.

A district focus proves theoretically appropriate as well. Extensive work on the ecology of policing documents district variation in norms (Klinger, 1997) and police practices like 'unfounding' (Taniguchi, 2010). Officers do not generally view parts of their area of responsibility in terms of single blocks. Herbert (Herbert, 1998) notes how LAPD officers distinguish between 'pro-police' and 'anti-police' areas, with associated behaviors in the latter areas including releasing seat-belts to aid movement, opening windows, and alerting dispatchers to an officer's location. Herbert's earlier work also reflected on the importance and respect given to district-level authority by officers from neighboring districts, whereby officers were expected to handle work in their area of responsibility as they saw fit (Herbert, 1997). In Philadelphia, police districts assigned to a condition required officers to print out daily mission grid maps, while other patrol officers not assigned a mission area knew about the ongoing experiments. District leadership and sergeants attended meetings about the experiment, though how and in what ways this information was passed on generally to other personnel in the district is not known. Nevertheless, given the clear importance attached to the experiment by Deputy Commissioner Bethel, it is plausible that at the district level participation in one of the two 'active' treatment conditions may have had an impact on policing across the entire district.

We witnessed in Philadelphia a reality that the intervention had an impact on the rest of the police district in a number of ways both potentially positive and negative: 1) marked and unmarked cars reduced available resources to address other district crime and disorder calls, 2) all officers at the district could be made precisely aware of the predicted grid locations, if they chose to be, 3) selection of officers for the cars may have biased the experimental impact (positively with 'hardchargers' or negatively with 'station queens' – to use Herbert's terms), and 4) the general focus on crime through the experiment might have changed behavior of other officers in a Hawthorne-by-proxy type of effect rather than a specific Hawthorne effect as was seen so many decades ago at the Western Electric assembly factory in Cicero, Illinois (see Homans, 1950).

Beyond suggestions of these effects through the qualitative research, any potential effects in these areas remain quantitatively unmeasured in the experiment. Nevertheless, in the case of this particular policing experiment, a plausible case can be made that in some conditions, more specifically, the marked and unmarked conditions, districtwide dynamics may have become engaged. Furthermore, this would seem particularly likely in the districts assigned to the marked condition since that was the one treatment assignment that most clearly suggested a crime reduction benefit.

Therefore, we might expect that districts participating in the marked condition would be the group most likely to evidence a district-wide disparity from the control condition districts during the time the property experiment was ongoing. It is in this group of districts where we would most likely expect lower weekly property crime counts *district-wide* compared to the control districts at the time the experiment was operative.

It is worth reiterating that we do not at this juncture have a clear idea of the relevant specific organizational dynamics at the district level should something be observed that seems linked with treatment participation. Nor were our qualitative data collection efforts attuned specifically to such potentialities. For example, for either the marked or unmarked conditions, we have no districtwide assessment about how cognizant patrol officers were at the district level about their district's participation and the nature of that participation. If there was broader district-wide knowledge among patrol officers not actually assigned to work the mission areas, how much did they know about participation details and/or administrative origins of the experiment itself? All of these are known unknowns at this point.

Nevertheless, despite these vagaries, we think it is reasonable to do three things. First, to descriptively explore what was happening districtwide before during and after the experiment.

That is, how were district-wide weekly property crime counts shifting before during and after the property experiment? To get a clue about whether something was happening, the global relationship between time and crime can be contrasted with the local relationship. If the local relationship diverges during the period the property experiment was operative, that would provide an initial suggestion something might have been going on.

Second, to contrast district-wide weekly property crime counts across the four conditions during the experiment. Of central interest is the contrast between districts in the marked condition versus those districts in the control condition during the experiment. Was property crime down districtwide in the marked condition on a weekly basis, paralleling the daily shift discrepancy already observed in buffered mission areas between control and marked districts? Third, if contrasting counts during the treatment timeframe are observed, what happens to those contrasts when temporal trends and autoregressive patterning during the experiment are also factored into the model?

Descriptive preliminaries

Descriptive statistics for district level weekly property crime counts appear in Table 14. These data are for 21 police districts, including district 7 which was left out of the property experiment. The airport district, 77, is excluded. Districts are unweighted, which means that each district contributes equally to the totals shown here regardless of variations in population size across districts. These counts are for the entire day, all shifts, for the entire district, for all days of the week. The period extends from the first week of January, 2014 up to the beginning of May, 2016.

Reported property crime counts ranged from none during the week up to 82. There are two outlying weekly property counts at around 80. District 12 located in the southwestern portion of the city, experienced a reported property crime count of 82 in the 31st week of 2014 at the end of July. District 24 in the Fishtown/Kensington/Port Richmond section of lower Northeast Philadelphia experienced a reported property crime count of 80 about four weeks before that. Except for these two weekly reported property crime counts, there were no other weeks in which any districts reported property crime counts exceeded 60. Given the gap between these two highest values and the rest of the crime count distributions, a Winsorized version of the count variable also was analyzed with these two highest values recoded to the third highest value, 59 (detailed statistics are not shown for the Winsorized count variable). All models described here were conducted with both the raw property count in the Winsorized property count variables.

A typical district in a typical week would report 17 or 18 property crimes; this translated to 2.64 reported property crimes a day, across all shifts, in a typical district during the entire timeframe. The interquartile range stretched from 12 to 23, indicating the middle half of the property crime counts were in this range. So half of the districts reported, on average over the timeframe, between 3.3 and 1.7 property crimes a day, across all shifts in that 24-hour stretch.

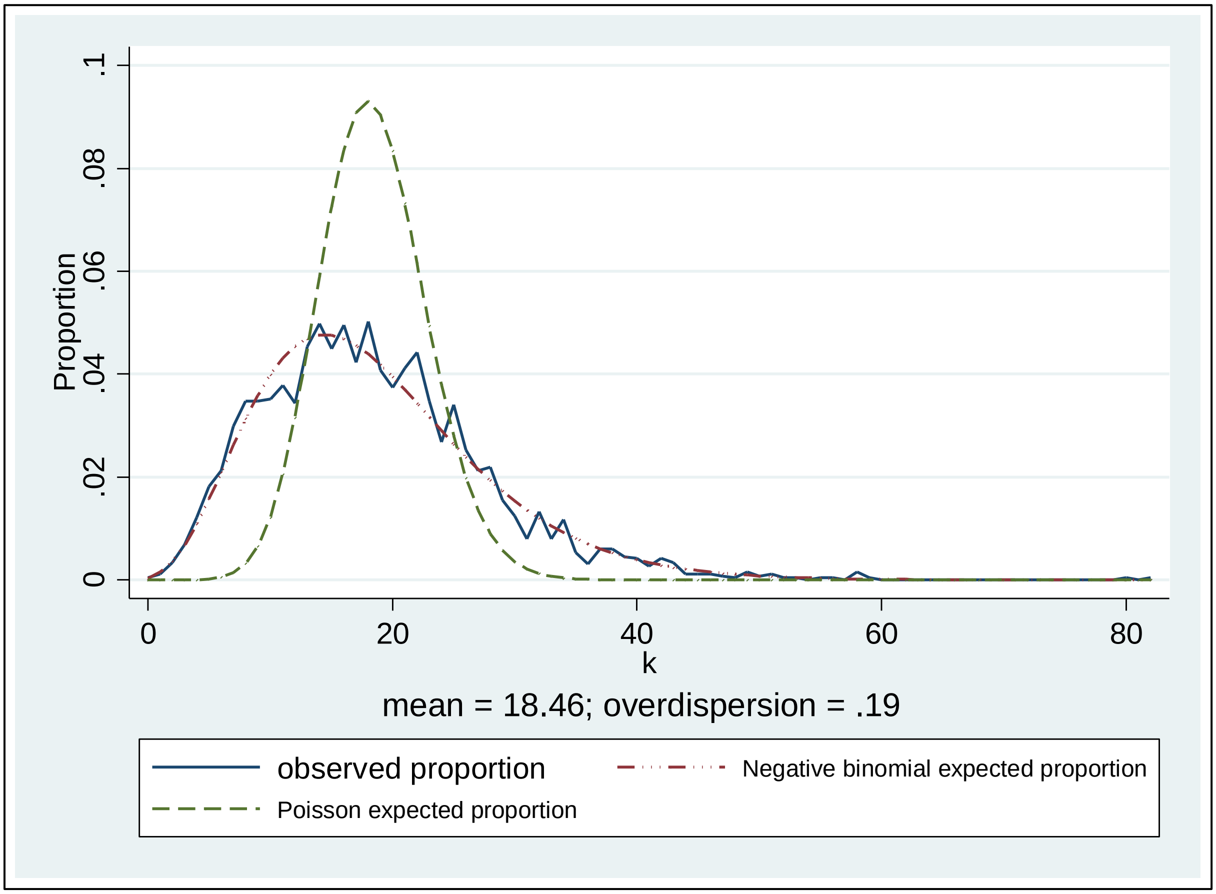
Table 14 District weekly property crime counts: January 1, 2014 – end of May, 2016

N	2,646
Min	0
Max	82
Mean	18.46
Median	17.00
SD	9.15
Variance	83.65
Skewness	1.02
10th percentile	8.00
25th percentile	12.00
75th percentile	23.00
90th percentile	30.00

Note. Airport excluded. Data for all remaining 21 districts, including District 7 which was excluded from the property experiment.

The variance, 84, far exceeds the mean, at 18, suggesting either an over-dispersed Poisson distribution or a negative binomial distribution. Figure 7 compares the observed proportions of counts at different values with the theoretical expectations of proportions based on either a Poisson distribution or a negative binomial distribution. The account distribution comes relatively close to matching theoretical expectations based on a negative binomial distribution of these reported crime counts. Consequently, that distribution was used in analyses.

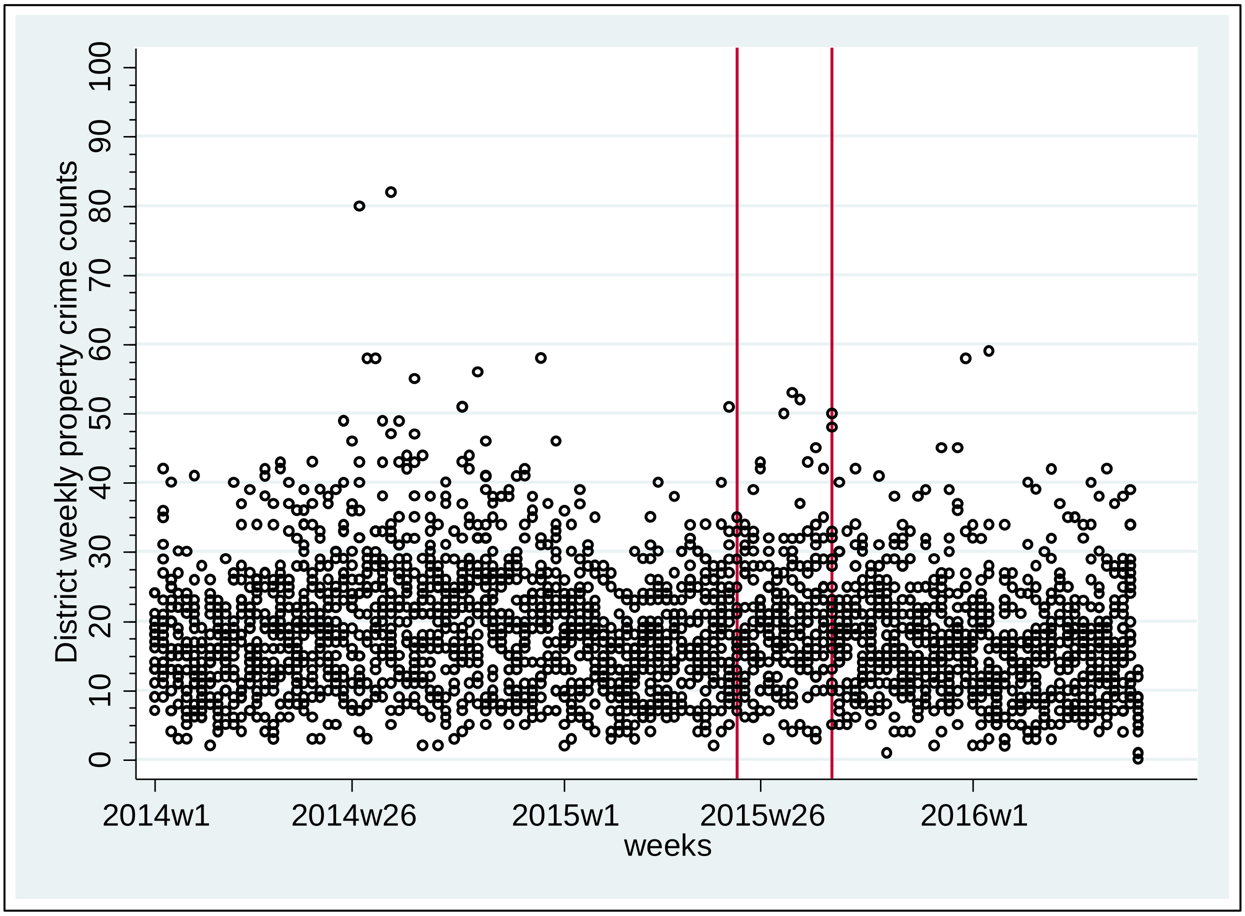
Figure 7. District weekly reported property crime counts: Comparison to theoretical Poisson and negative binomial distributions



Note. Weekly period extends from beginning of January 2014 to the beginning of May, 2016. Data shown for 21 districts, including the one district not participating in the property experiment. Airport excluded.

The district level weekly property crime counts are arrayed by week in Figure 8. Two vertical lines indicate the weeks that marked the beginning (June 1, 2015) and the ending of the property crime experiment (August 31, 2015). Each circle represents a weekly reported property crime count in one district.

Figure 8 Reported district level weekly property crime counts



Note. Airport excluded. District 7 included even though not part of the property experiment. Weekly period extends from beginning of January 2014 to the beginning of May, 2016. Red lines correspond to beginning and end of property experiment period.

As can be seen, save for the two counts of 80 and 82, no other reported weekly property crime counts at the district level exceeded 60. Further, the figure also shows only a small number of district weekly counts between 50 and 60; there were 13 in this range.

The two outlying counts and 80 and 82 appear in the first calendar year of the data series. There also seem to be few district weekly reported property crimes counts well above 40 in the first quarter of 2014. Further, in the last year of the data series, stretching from the leftmost red vertical line to the end of the series, there are only two counts closer to 60 than 50. And, following the end of the property experiment, that is, to the right of the right most red vertical line, there seem to be far fewer counts in the 40 to 50 range then was seen during the first year. So descriptively this suggests increasing crime counts from the first to the second quarter of 2014, followed by somewhat declining crime counts for the rest of this series.

Contrasting two ways to think about trends over time

One simple (and in some ways simplistic) way to model this variation over time is to fit a quadratic curvilinear line to these data. That quadratic line can also have attached to it an upper and lower 95% confidence interval. Data points within that confidence interval then roughly correspond to a *global* expectation of initially increasing and later decreasing crime counts; it is a global expectation because it aligns with a relationship based on the entire time frame.

The curvilinear smoothed curve counts each data point equally when making each predicted point on the curve. So when making the prediction for the first few weeks of the series, the weights of the data points in that immediate region are identical to the weights of the data points at the end of the series. Such a smoothed curve based on these predictions depicts the *global* relationship between time and the outcome counts.

But suppose we are interested in the local relationship between time and the outcome? That is, how are the temporal trends affected right around the time the property experiment is operative? Such smoothing, in making each predicted point, would prioritize values in the series that are temporally proximate, and de-emphasize values in the series that are farther away in time. The LOWESS (locally weighted scatterplot smoothing) procedure employs such smoothing and creates a locally weighted, robust regression line (Cleveland, 1979). “Lowess is a desirable smoother because of its locality —it tends to follow the data” (Stata Corporation, 2015: 1311). For our purposes here, the smoothing protocol is potentially helpful. ‘Locally weighted’ means count values from around the same time in the series.

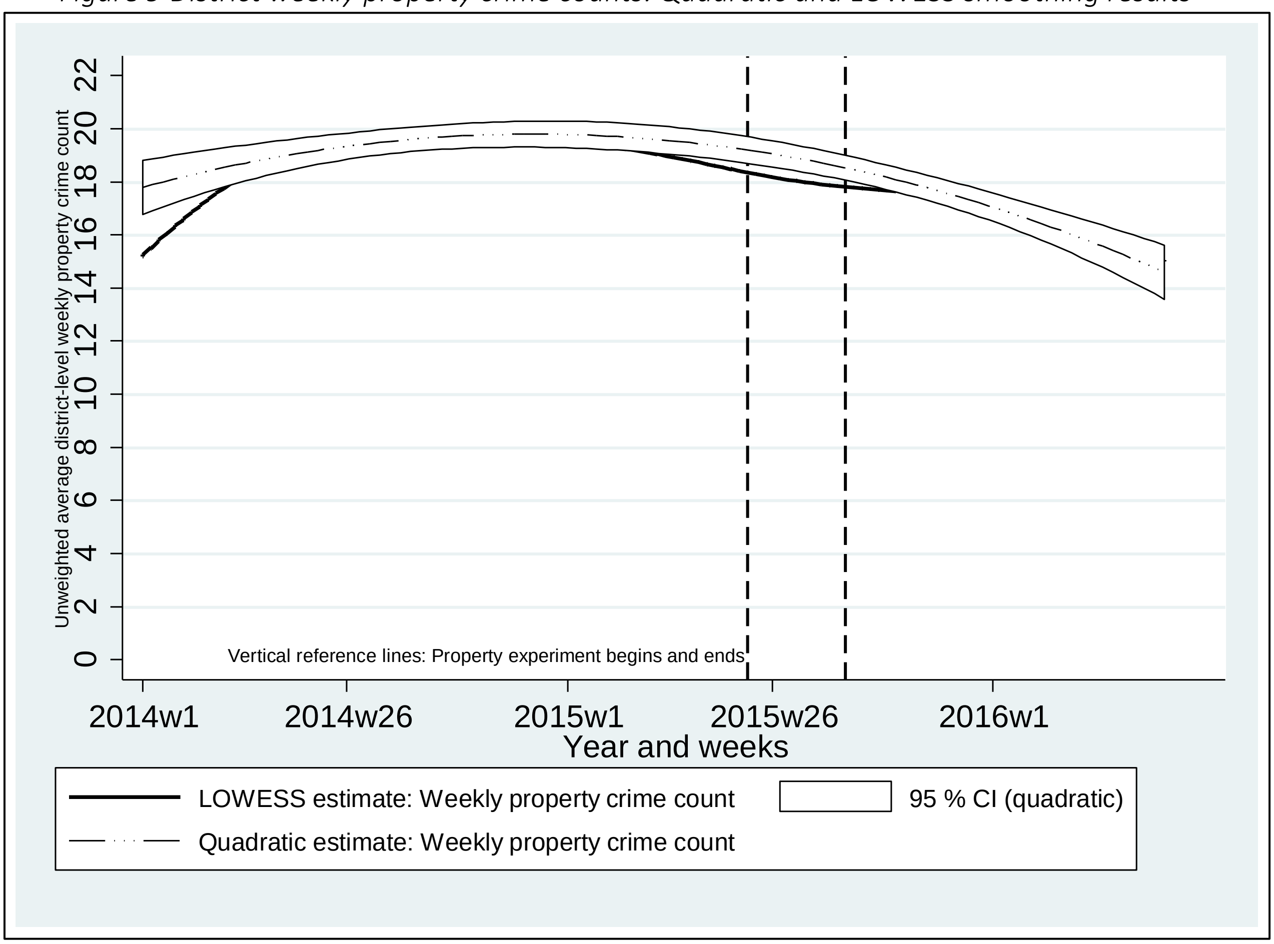
Of particular interest is whether a smoothed, locally weighted trend line deviates significantly below the smoothed quadratic line during the time the property experiment was operational. If it does, then further investigation by specific experimental condition would be appropriate.

Given the property crime results for mission areas and particularly buffered mission areas, we would expect to be most likely to see significant deviations of the LOWESS trend line from the smoothed quadratic line in the districts assigned to the marked condition. We would expect those counts to deviate below the trend line describing the global relationship between time and reported property crime counts.

All 21 districts

Using all 21 police districts, including district 7 which is not part of the property experimental phase, but leaving out the airport district, Figure 9 shows the relationship between time and weekly district reported property crime counts modeled in two different ways. The dashed-and-dotted line represents the predicted score based on quadratic fitting. This is the global relationship between district weekly property crime counts and time. The lines above and below the quadratic curve represent the upper and lower 95% confidence interval bounds of that curvilinear prediction.

Figure 9 District weekly property crime counts: Quadratic and LOWESS smoothing results



Note. Estimates using district-level weekly property crime counts from all 21 districts, *Philadelphia. Airport excluded. Vertical dashed lines indicate beginning and ending of property crime experiment.*

The LOWESS locally weighted and smoothed regression appears as a solid line. This depicts the local relationship between time and weekly district level property crime counts. That solid line is within the confidence interval of the curvilinear line except for two points in time. First, at the very beginning of the data series, in the first few weeks of 2014, the locally weighted regression line falls below the lower confidence bound of the curvilinear line. As was noted earlier, the first three months of 2014 appeared to be a time of quite low property counts. During the first three months of the series only a very small number of district weekly counts were 40 or higher. The smoothing function based on LOWESS prioritizes those nearby count values when making predictions for this segment of the series. The global quadratic smoothing function does not. Hence the divergence.

The second time the curvilinear line is outside of the confidence intervals is of more interest. This occurs just prior to the beginning of the property experiment to slightly after: here again the locally weighted regression line is outside the confidence bounds of the curvilinear line, and thus falls significantly below predictions based on the latter.

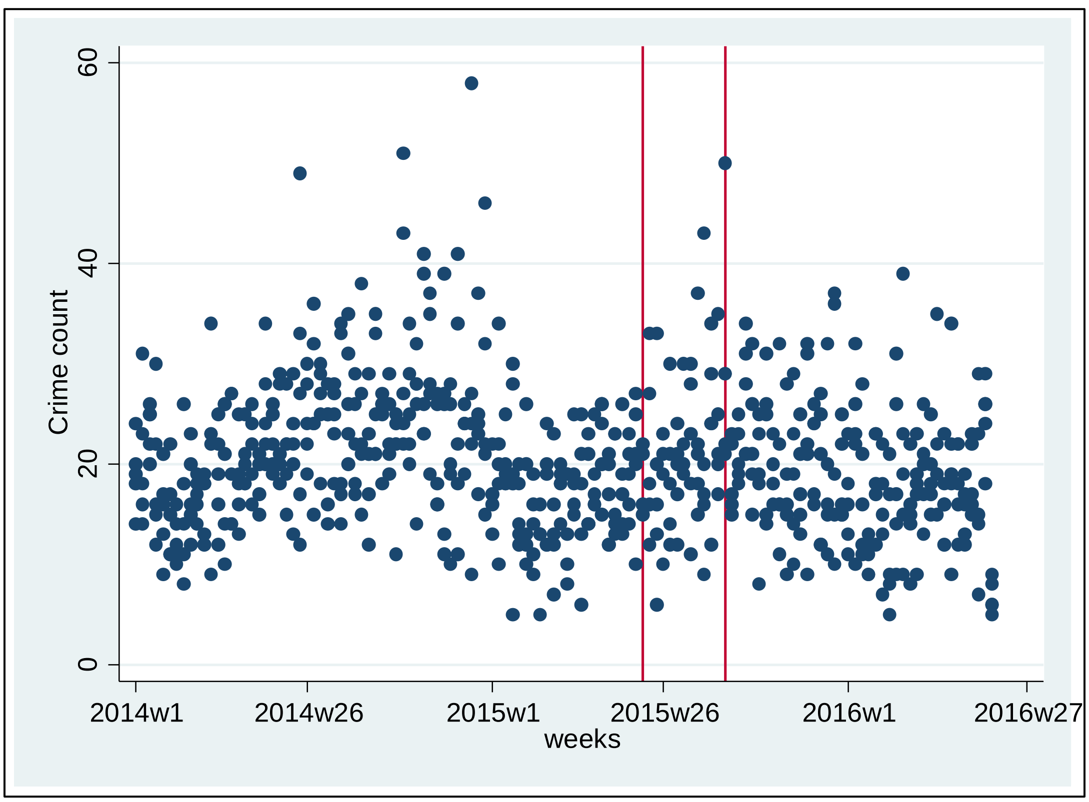
The timing of this significant discrepancy between these two ways to model property crime counts over time merits mention. The significant divergence begins just a very few weeks before the actual experiment begins. There are at least two possibilities for the timing of this discrepancy. One possibility is that there was something 'naturally' occurring in Philadelphia, starting in late April or early May of 2015, which was leading to lower property crime counts at the district level. This could be the threat to internal validity of local history which is 'local' in that it is specific to just these few weeks. Another possibility is that, as a result of informal communication and formal meetings held at headquarters, and visits out to particular districts, district leadership, shift supervisors, and line officers may have become more aware of the upcoming experiment. The divergence starts shrinking toward the end of the property experiment. Immediately following the conclusion of the property experiment on August 31, the PPD went into even more intensive preparation for Pope Francis's visit to Philadelphia on September 26^{th} and 27^{th} of 2015.

But there is one way we can learn more about this plausible threat to internal validity of local (in time) history. If something was 'naturally' occurring throughout PPD districts at just around the time prior to initiating the property experiment, then the significant divergence between the locally weighted robust prediction and the curvilinear global prediction should be equally evident across all four experimental conditions. But if something was happening that was connected to the documented significant and localized mission-level impact of the **marked** condition on crime counts in buffered mission areas, we would expect the divergence to be significant *especially or only* when examining districts assigned to this condition.

Districts in marked condition

Figure 10 shows the distribution of weekly reported property crime counts for all the districts in the marked condition. During the time the experiment was active – between the two vertical reference lines – there were two weeks when one district had a reported property crime counts above 40. There had been no weekly counts above 40 for several preceding weeks in these districts assigned to the marked condition.

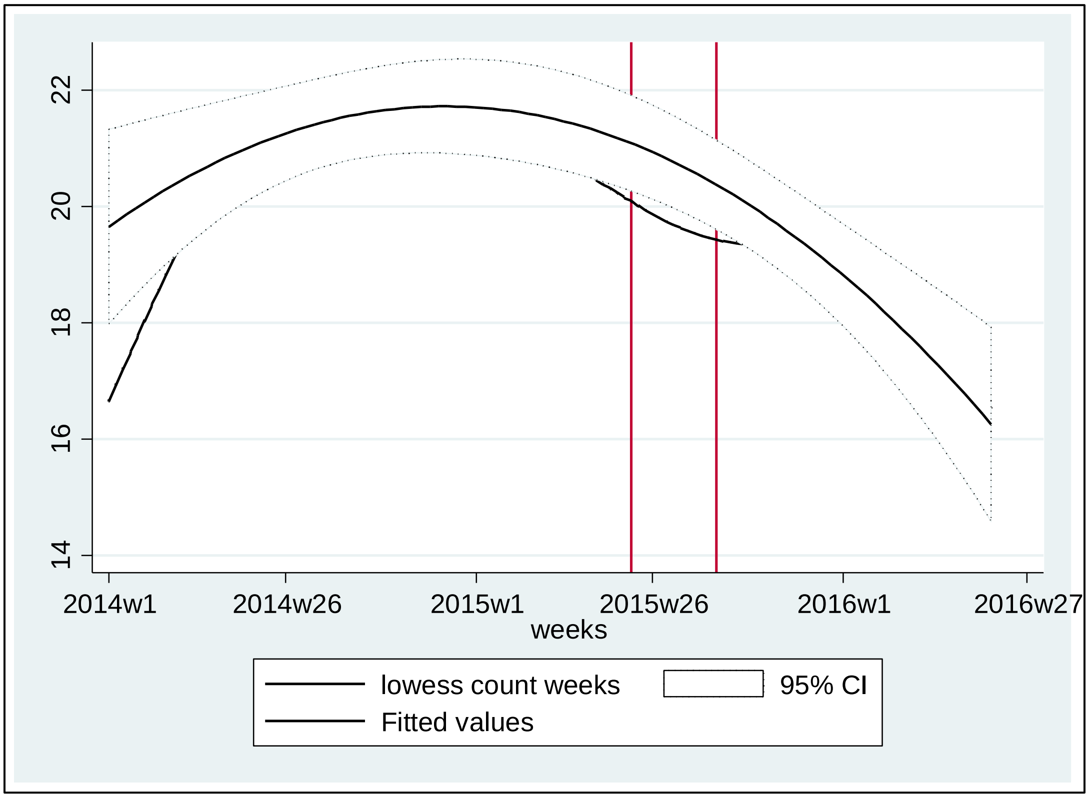
Figure 10 Marked condition only: Weekly district-level reported property crime counts



Note. Reported district-level weekly property crime counts, five marked condition districts only. Vertical lines reference beginning and ending weeks of the property experiment. Source: Philadelphia Police Department.

Figure 11 contrasts the curvilinear smoothing – the global relationship – with the locally weighted robust smoothing – the temporally proximate or local relationship. The locally weighted predictions for weekly property crime counts diverged significantly from the curvilinear predictions during the time the experiment was operative. The LOWESS smoothed function for marked condition districts started moving toward the bottom portion of the conference interval a few weeks before the experiment started (2015w22) but actually dipped below and outside of the confidence interval during the time the experiment was operative. Soon after the property experiment terminated, the locally weighted trend line for these districts moved back within the confidence interval based on the global relationship. *In short*, from slightly before the property experiment began, to slightly after it ended, the locally weighted relationship between time and district-wide property counts diverged significantly below the expectations based on the global relationship. Stated differently, weekly property crime counts were coming down significantly faster for these districts during the experiment than would have been expected given the overall temporal patterning.

Figure 11 Marked condition only: Global and local smoothed relationships between time and property crime counts

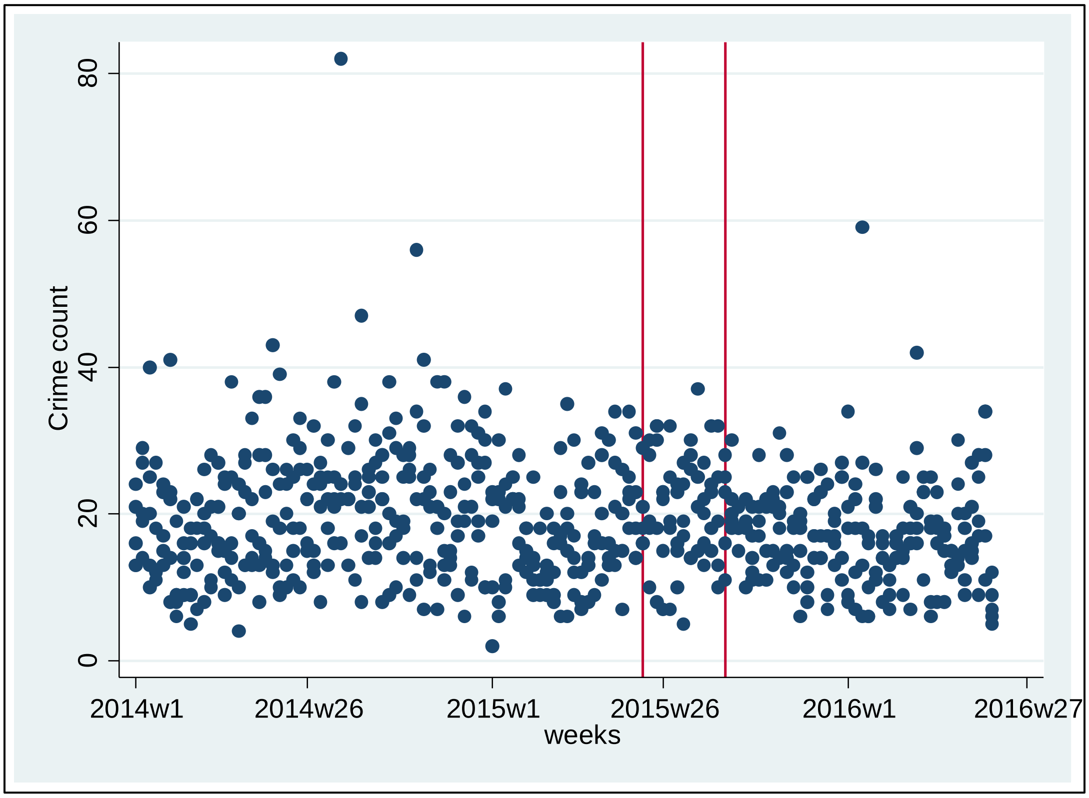


Note. Estimated relationships between weekly district property crime counts and time. Only districts in marked condition modeled. Vertical lines reference beginning and ending weeks of the property experiment. Smoothed line in the middle of the confidence interval shows global curvilinear relationship between time and crime counts. Dashed lines show upper and lower bounds of the 95 percent confidence interval of the curvilinear relationship. LOWESS line shows locally weighted relationship. Data source: Philadelphia Police Department.

Districts in unmarked condition

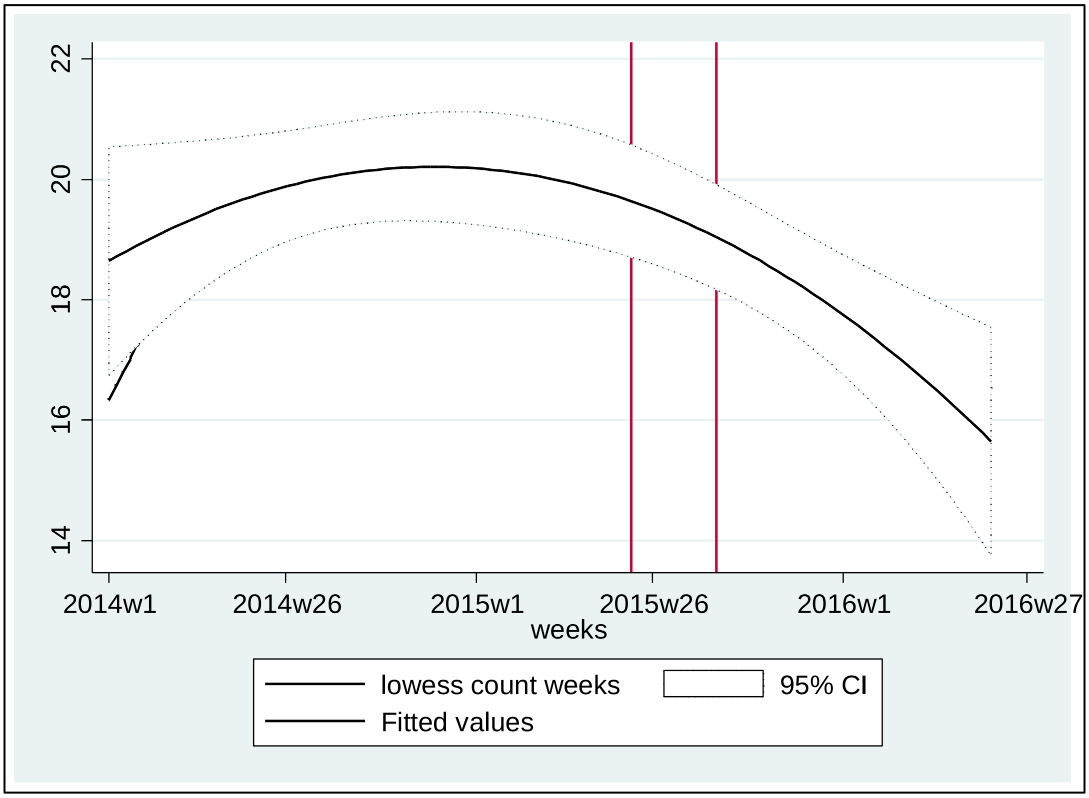
Figure 12 shows weekly property crime counts for all the districts in the unmarked condition. Figure 13 contrasts predicted crime counts based on the global vs. local relationship with time. Except for a very slight discrepancy at the beginning of the series in very early 2014, at no other point was there significant divergence between these two predictions. In other words, the local relationship between time and weekly property crime counts in these districts generally followed what would be expected given the global relationship between time and property crime counts at the district level.

Figure 12 Unmarked condition only: Weekly district-level reported property crime counts



Note. Reported district-level weekly property crime counts, five ***un****marked condition districts only. Vertical lines reference beginning and ending weeks of the property experiment. Source: Philadelphia Police Department.*

Figure 13 Unmarked condition only: Global and local smoothed relationships between time and property crime counts

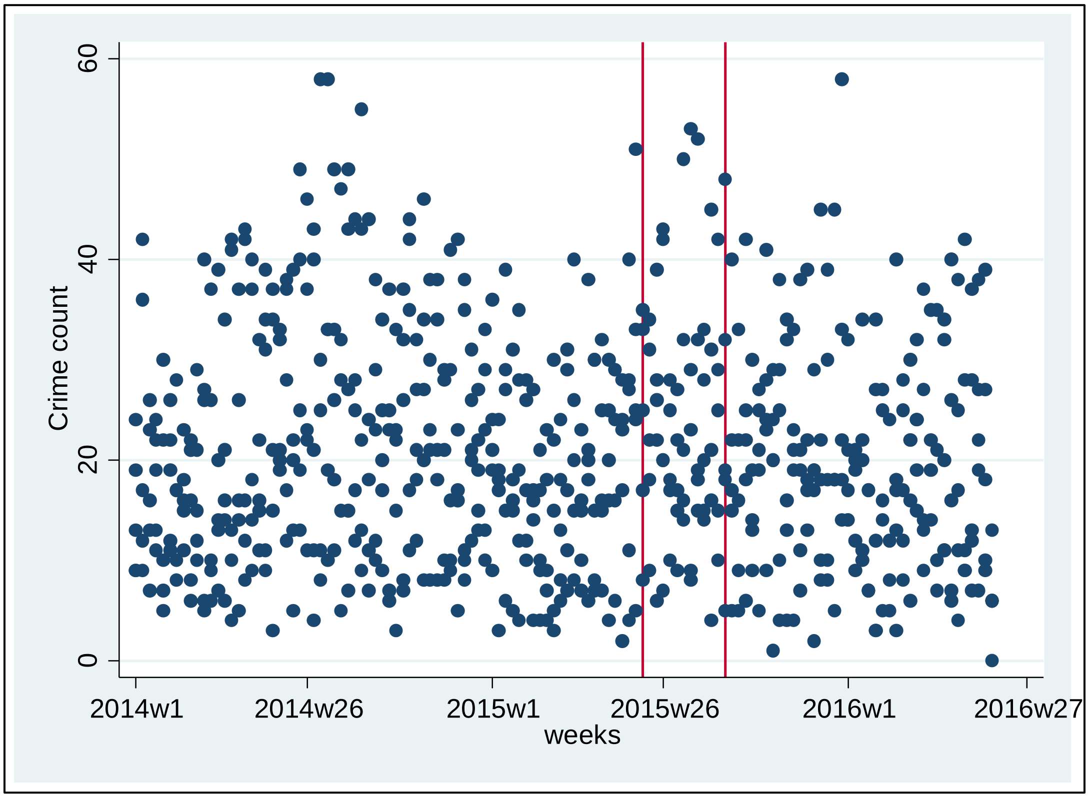


*Note. Estimated relationships between weekly district property crime counts and time. Only districts in **un**marked condition modeled. Vertical lines reference beginning and ending weeks of the property experiment. Smoothed line in the middle of the confidence interval shows global curvilinear relationship between time and crime counts. Dashed lines show upper and lower bounds of the 95 percent confidence interval of the curvilinear relationship. LOWESS line shows locally weighted relationship. Data source: Philadelphia Police Department.*

Districts in awareness condition

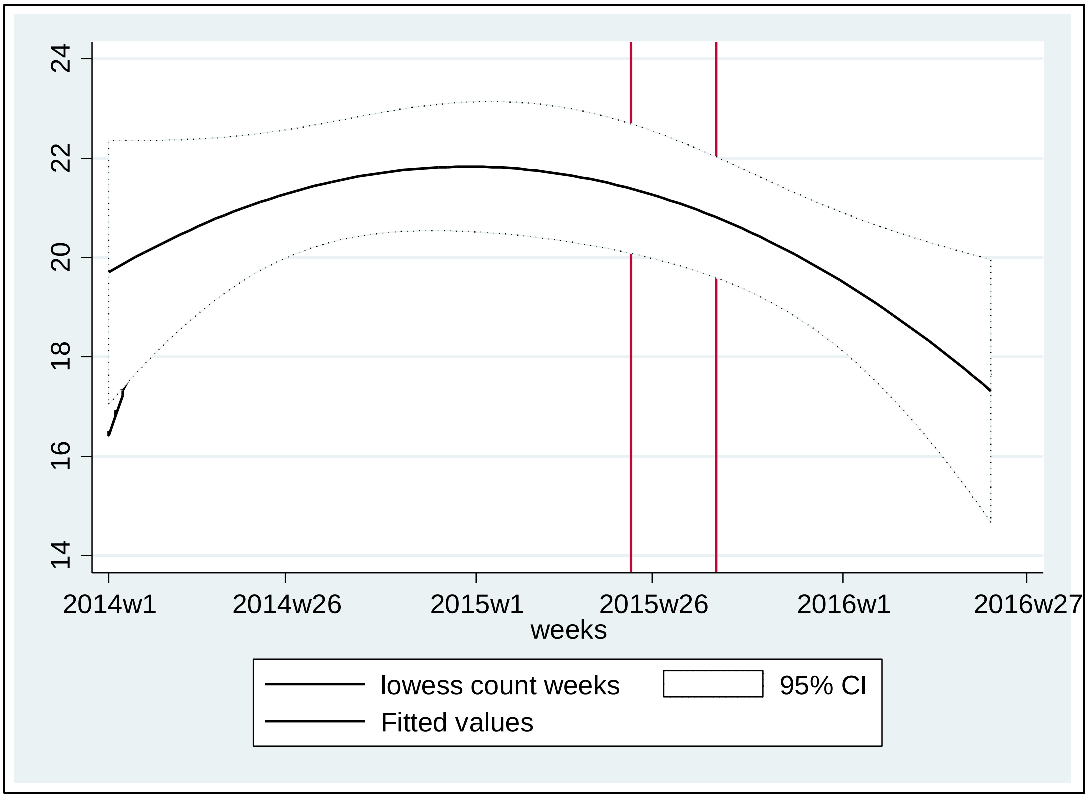
Figure 14 shows weekly property crime counts for all the districts in the awareness condition. Figure 15 contrasts predicted crime counts based on the global curvilinear fit versus the locally weighted smoothed trendline for these districts. Except for a very slight discrepancy at the very beginning of the series, the global modeled relationship between time and crime counts and the local model relationship did not differ significantly. Subsequent to this initial discrepancy during the first couple of weeks of the series, the robust locally weighted predictions were always within the confidence interval based on the curvilinear predictions.

Figure 14 Awareness condition only: Weekly district-level reported property crime counts



Note. Reported district-level weekly property crime counts, five awareness condition districts only. Vertical lines reference beginning and ending weeks of the property experiment. Source: Philadelphia Police Department.

Figure 15 Awareness condition only: Global and local smoothed relationships between time and property crime counts



Note. Estimated relationships between weekly district property crime counts and time. Only districts in awareness condition modeled. Vertical lines reference beginning and ending weeks of the property experiment. Smoothed line in the middle of the confidence interval shows global curvilinear relationship between time and crime counts. Dashed lines show upper and lower bounds of the 95 percent confidence interval of the curvilinear relationship. LOWESS line shows locally weighted relationship. Data source: Philadelphia Police Department.

Districts in control condition

Figure 16 shows weekly property crime counts for all the districts in the control condition. District 7 is not included here. These two figures are based only on the five districts randomly assigned to the control condition. Figure 17, based also just on those five districts, again shows the global vs. local over time district property crime count predictions based on the two methodologies for the control districts. Again, except for the very beginning of the series, there are no significant discrepancies between the two prediction methods. The locally weighted relationship between time and property counts roughly aligns with the globally weighted relationship throughout the series.

Figure 16 Control condition only: Weekly district-level reported property crime counts

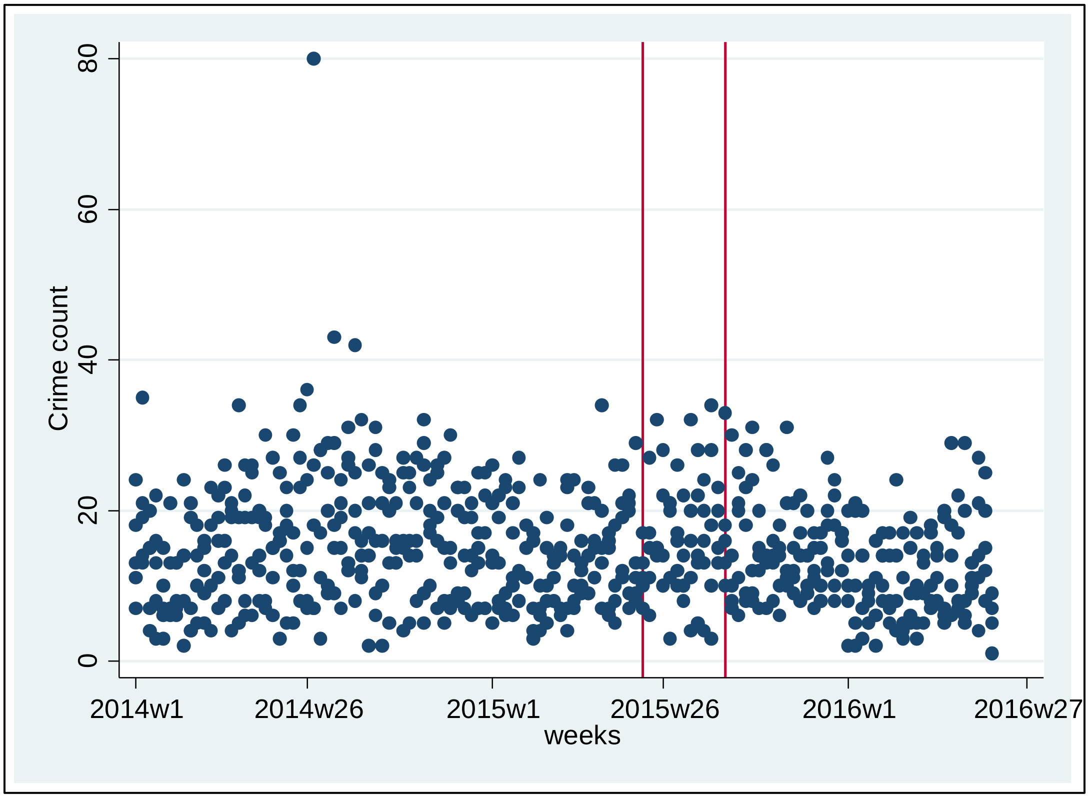
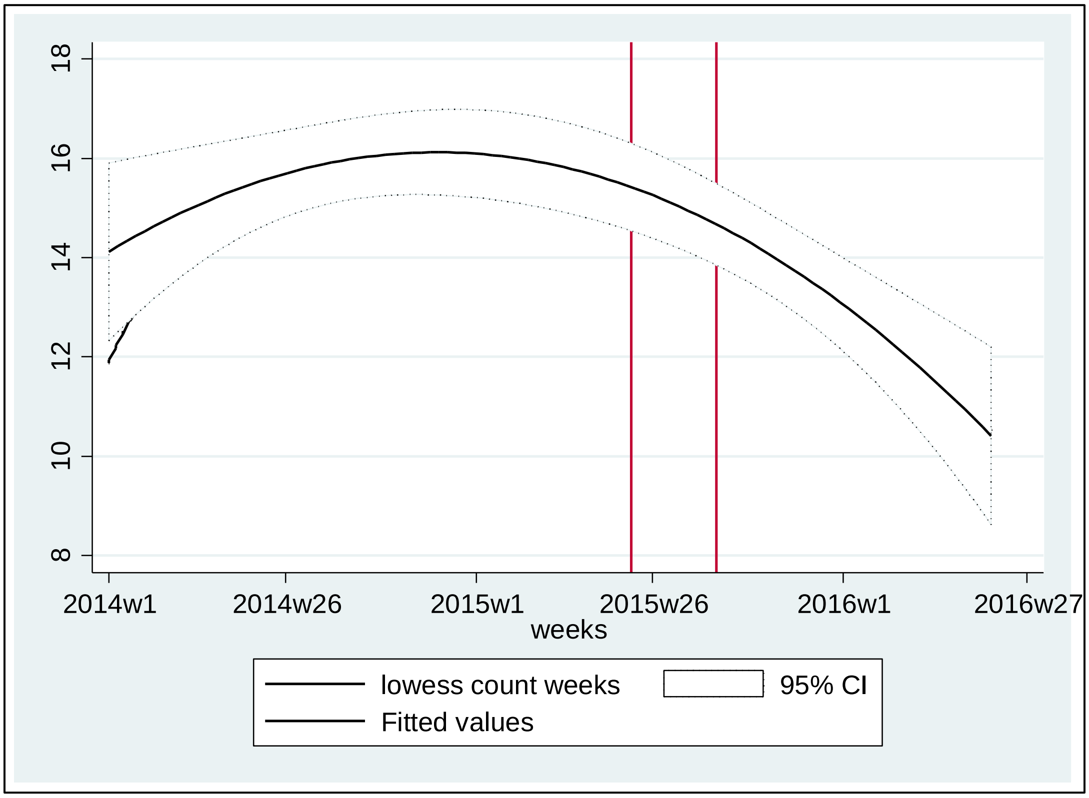


Figure 17 Control condition only: Global and local smoothed relationships between time and property crime counts



Note. Estimated relationships between weekly district property crime counts and time. Only districts in control condition modeled. Vertical lines reference beginning and ending weeks of the property experiment. Smoothed line in the middle of the confidence interval shows global curvilinear relationship between time and crime counts. Dashed lines show upper and lower bounds of the 95 percent confidence interval of the curvilinear relationship. LOWESS line shows locally weighted relationship. Data source: Philadelphia Police Department.

Summary on global versus local predictions about time and crime across conditions

Only in the marked condition districts do we see significant discrepancies during the time of the property experiment between the global relationship, between time and property crime counts at the district level, and the local relationship. It is only for this condition that, during the time the property experiment was active, the locally weighted predictions fall outside the confidence interval based on the global relationship between time and district level weekly reported property crime counts. In no other conditions during the time when the experiment was operational do the local predictions, heavily weighting data points nearby in time, fall below the confidence interval based on the global relationship between time and crime.

There is a broader implication. This descriptive graphical examination suggests that in addition to the impacts of the marked condition on property crime counts within the buffered mission grid areas, there also may have been a district-wide impact on weekly property crime counts during the experiment in the marked condition districts.

The next questions to consider include the following:

- During the time of the experiment, how did the expected weekly district level property crime counts in the three experimental treatment conditions contrast with those in the control condition?
- If expected property crime counts are lower in the marked condition as compared to the control condition, how sizable is that discrepancy and is it statistically significant?
- Further, if crime counts are lower in the marked as compared to the control condition, how is the size of that disparity affected when temporal trends within the experiment period, and serial autocorrelation of errors, get introduced?

RESULTS: DISTRICT-LEVEL TREATMENT SPECIFIC FINDINGS

Did weekly property crime counts at the district level vary as a function of treatment assigned? To learn whether this is the case, cross-sectional panel design generalized estimating equation models (*xtgee* command in Stata) analyzed two sets of weekly property crime counts: 1) during the entire experiment; and 2) the property crime phase minus the first and last week of the phase. The rationale for this second examination will be explained later.

Predictors and outcome variables

The outcome variable for this part of the study was the weekly reported property crime counts at the district level during the time of the property experiment which were analyzed only in raw counts. The Winsorized variable was identical to the raw variable for this subset of weeks. With regard to predictor and outcome variables, descriptive statistics for the variables used in the models of weekly district-level property crime counts appear in Table 15. For these weeks, the count outcome variable in the Winsorized version of the count were both equivalent.

The exposure variable used was the district total population estimated from the 2010 – 2014 American community survey data allocated to districts. The population figure used was in the thousands, so the resulting constant in in the models can be interpreted as a weekly district level property crime rate per thousand persons in the district.

Linear and curvilinear trends during the period were controlled with two variables: a centered count of weeks, and the square of that centered week count. The weekly count began in the first week of 2014 and ended the first week of May 2016. Because the linear time variable was centered before squaring these two temporal components prove orthogonal to one another.

Proportional contrast codes (Serlin & Levin, 1985) were used to compare specific treatment conditions to the control condition districts. There was a proportional contrast code for each of the three experimental conditions: awareness, marked, and unmarked. After some basic algebra (Serlin & Levin, 1985: 226-228) a district was assigned a value of +.75 if it was in the condition capture by that specific code, otherwise the district was assigned a value of -.25.[6] The weight associated with each contrast code means that each coefficient reflects the mean difference between the weekly district-level property crime rate in the condition assigned, and the same rate in the control condition. The constant reflects the overall mean of twenty districts in the three different experimental conditions and the control condition. This analysis also used post-estimation techniques to investigate mean differences between the three treatment conditions themselves.[7]

Model sequence

A series of models were run for the full timeframe of the property experiment, and again for the experiment minus a week at the front end of the experiment and minus a week at the back end. This latter model was conducted for a number of reasons. First, there were inevitably some startup glitches with the experiment in terms of operationalization. For a day or two at least, a number of districts struggled to log into the web-based predictive software program, or print the maps, or regularly access the requisite vehicles, or understand the focus of the intervention. This is not to be critical of the Philadelphia Police Department. These are arguably typical of inevitable teething problems to be expected when a headquarters initiative is implemented for the first time in the field. At the end of the experimental phase, the police department were gearing up for the visit of the Pope, and we began, in the week before the property phase concluded, to see officers who had served on the predictive policing experiment reassigned to attend briefings related to anticipated activity around the Pope's visit. The visit caused significant upheaval to the city, so there were inevitably numerous briefings and trainings that many

[6] The only interpretive implication of using proportional contrast codes as compared to dummy codes with the control district as the reference string is that with the proportional codes the constant now reflects the average for *all twenty* districts, not just the control districts.
[7] Stata's lincom tests the null hypothesis that for a pair of coefficients the difference between them is zero.

officers were required to attend in the run-up to the visit. Both the initial teething problems and the upheaval of the Pope's visit affected implementation of the experimental conditions. As a result, we wanted to look at any possible impacts district-wide absent these unusual conditions.

The model series included: just the proportional contrast codes; those codes plus centered linear time and centered quadratic time; the codes plus time plus autoregressive errors ranging from a one-week lag up to and including a four-week leg. The tables show the results for two models in the series: the model with just the proportional contrast codes (Model A); and the model with the contrast codes, the two time controls, and a four lags autoregressive error structure (Model B). The autoregressive models were rerun without the linear and quadratic time variables (results not shown). Results were closely comparable.

Following each model, the discrepancy between the marked condition and each of the other two conditions (awareness and unmarked) was tested for significance with the Stata lincom command.

Table 15 Descriptive statistics, weekly district-level property crime counts, during property experiment

During experiment							
Variable	Variable name	N	Min	Max	Mean	SD	Median
Weekly district level property crime count	count	260	3	53	20.446	9.315	19.5
District population in thousands (exposure)	pop_ap1k	260	33.949	127.532	76.321	30.496	73.808
Marked proportional contrast code	mark_EC3	260	-0.25	0.75	0	0.434	-0.25
Unmarked proportional contrast code	unma_EC3	260	-0.25	0.75	0	0.434	-0.25
Awareness proportional contrast code	awar_EC3	260	-0.25	0.75	0	0.434	-0.25
N of weeks (centered)	c_isweeks	260	11.5	23.5	17.5	3.749	17.5
Centered N of weeks, squared	c_wk_sq	260	121	529	303	128.067	289
During experiment, less first and last week							
Weekly district level property crime count	count	220	3	53	20.382	9.209	20
District population in thousands (exposure)	pop_ap1k	220	33.949	127.532	76.321	30.507	73.808
Marked proportional contrast code	mark_EC32	220	-0.25	0.75	0	0.434	-0.25
Unmarked proportional contrast code	unma_EC32	220	-0.25	0.75	0	0.434	-0.25
Awareness proportional contrast code	awar_EC32	220	-0.25	0.75	0	0.434	-0.25
N of weeks (centered)	c_isweeks	220	12.5	22.5	17.5	3.169	17.5
Centered N of weeks, squared	c_wk_sq	220	144	484	299	108.126	289
Note. Descriptive statistics for weeks the property experiment active (June 1 - August 31, 2015).							
Note. Numbers in bottom half of table drop off the first and last week of the property experiment							
Note. See text for details on proportional contrast codes							
Note. Centered variables do not have a mean score of zero because centering was done on entire period.							

Table 16 During property experiment: discrepancies with control condition districts when predicting weekly counts

	Model A		Model B	
	b	IRR	b	IRR
	se		se	
	t		t	
Predictor				
Contrast marked v. control (mark_EC3)	-0.0570	0.945	-0.0315	0.969
	0.147		0.140	
	-0.388		-0.225	
Contrast unmarked v. control (unma_EC3)	0.185	1.203	0.196	1.217
	0.173		0.188	
	1.071		1.048	
Contrast awareness v. control (awar_EC3)	0.266	1.305	0.285	1.330
	0.190		0.208	
	1.394		1.368	
Weeks since 1/1/2014 (c_isweeks)			-0.0513	0.950
			0.0616	
			-0.833	
Weeks squared (c_wk_sq)			0.00189	1.002
			0.00186	
			1.016	
Exposure ln(population in thousands) (pop_ap1k)	1			
Constant (average of all 20 districts)	-1.250	0.287	-0.925	0.397
Observations	260		260	
Number of districts	20		20	
df for χ^2	3		5	
χ^2	3.260		7.693	

Note. Predicting district level weekly property crime counts during the property crime experiment. Cross-sectional panel design generalized estimating equations with weeks nested within districts. Model A includes just proportional contrast codes. Model B includes those codes was linear and quadratic time and an autoregressive function with a lag of four weeks for the autoregressive errors.

Results using all the weeks of the property experiment appear in Table 16. Although none of the proportional contrast codes are significantly different from the control condition, there are noteworthy features of the results. First of all, the only condition where the expected weekly property crime count was *lower* than the expected property crime count in the control condition, was the marked condition. The expected property crime count was anywhere from five and a half percent lower than in the control districts, to three percent lower than expected property crime counts in the control districts, depending on the model examined. Note that in the unmarked condition expected weekly property crime counts were about 20 to 22 percent higher than in the control condition districts, and in the awareness districts expected property crime counts were anywhere from 31 percent to 33 percent higher compared to the control districts. The incident rate ratios for the different conditions contrasted with the control condition were closely comparable to those shown in the tables if no variables were included to control for the passage of time (results not shown). The estimated within-district serial error autocorrelations were .62 for a lag of one, .46 for a lag of two, .40 for a lag of three, and .39 for a lag of four. Earlier modeling efforts suggested that adding lags greater than four did not improve model fit while controlling for model complexity (results not shown). The post-estimation Z tests comparing the marked condition with the other two conditions were all nonsignificant (P >.05, one tailed).

In short, the results of this model do not definitively confirm that the marked districts were doing better than the control districts; nor do they definitively confirm that the marked districts were doing better than either the awareness districts or the unmarked districts. The results do *hint*, however, that things were going slightly worse in the unmarked and awareness districts as compared to the control districts, and reinforce the suspicion that things were going slightly better in the marked districts as compared to the control districts during this period. This is evidenced by the Incident Rate Ratio for the mark_EC3 variable showing a (non-significant) 3.1% lower level of expected property crime weekly counts in the marked districts compared to control during the experimental period of 12 weeks across the five districts.

These models align with, in essence, ‘intention to treat’ models. Realistically, however, given some minor startup hiccups, and the close out of the experiment in anticipation of the papal visit, it is also of interest to look at how the middle ten weeks of the property experiment fared. Attention turns to that next.

Table 17 Property experiment less first and last weeks: discrepancies with control condition districts when predicting weekly counts

	Model A		Model B	
	b	IRR	b	IRR
	se		se	
	t		t	
Predictors				
Contrast marked v. control (mark_EC32)	-0.0848	0.919	-0.0840	0.919
	0.166		0.183	
	-0.512		-0.460	
Contrast unmarked v. control (unma_EC32)	0.179	1.196	0.176	1.192
	0.170		0.181	
	1.051		0.972	
Contrast awareness v. control (awar_EC32)	0.249	1.283	0.252	1.287
	0.184		0.198	
	1.358		1.274	
Weeks since 1/1/2014 (c_isweeks)			-0.130	0.878
			0.0755	
			-1.720	
Weeks squared (c_wk_sq)			0.00395	1.004
			0.00212	
			1.865	
Exposure ln(population in thousands (pop_ap1k)	1			
Constant (average across all districts)	-1.254	0.285	-0.162	0.850
Observations	220		220	
Number of districts	20		20	
df for χ^2	3		5	
χ^2	3.292		7.496	

Note. Predicting district level weekly property crime counts during the property crime experiment. Cross-sectional panel design generalized estimating equations with weeks nested within districts. Model A includes just proportional contrast codes. Model B includes those codes was linear and *quadratic time and an autoregressive function with a lag of four weeks for the autoregressive errors. Period examined excludes the first and last week of the property experiment.*

Table 17 displays the results of the models. As in the prior table, the same two models are shown: the model with just the contrast codes, and the model with the contrast codes, with the two time variables, and with an autoregressive error structure with a lag of 4. Although none of the experimental conditions differs significantly from the control condition, there are points worthy of interest. Most importantly, the expected weekly property crime counts were 8 percent lower in the marked condition districts as compared to the control condition districts during the weeks when the property experiment was most fully functional (the IRR value of 0.919 for the variable mark_EC32). This contrasts with the other two conditions. In the unmarked districts the expected property crime counts were about 19 percent higher compared to the control districts; in the awareness condition districts the expected counts were about 28 percent higher compared to the control districts.

Although the contrast coefficient for the marked condition is negative, and the contrast coefficient for the other two treatment conditions are positive, there was no significant difference between the marked coefficient and the awareness coefficient, or between the marked coefficient and the unmarked coefficient. The estimated within-district serial autocorrelations were .58 for a lag of one, .43 for a lag of two, .41 for a lag of three, and .43 for a lag of four.

Summary

The graphical results contrasting the global relationship between time and weekly district level property crime counts with the local relationship suggested that in the marked car condition the local prediction for district property crime counts deviated significantly below the global prediction during the weeks when the property crime experiment was operative. This suggested that there may have been a district level dynamic at work in the marked condition that might have been affecting the entire district property crime levels during the experiment.

Statistical models contrasting each treatment condition with the control condition during the experiment, and during the experiment focusing just on the weeks when it was most fully operational, tended toward supporting this scenario but did not confirm it at the required levels of statistical significance. During the full property experiment, expected weekly property crime counts for the entire district were between three and six percent lower in the marked condition districts as compared to the control condition districts. If the first and last weeks of the property experiment are dropped out of the analysis that looks like the expected weekly property crime counts in the marked condition were about eight percent lower compared to the control districts during the same timeframe. This contrasts with higher than expected crime counts in awareness versus control and in unmarked versus control districts.

RESULTS: DAILY REPORT LOGS

Property crime experiment phase (officer forms)

This section reports descriptive statistics for the officer feedback forms (n = 747) from both the marked as well as the unmarked car districts.

Resource allocation

The first set of descriptive statistics seeks to answer the research question regarding the extent to which resources were allocated, and correctly allocated, to the prescribed areas. For the districts that ran the project using unmarked cars, the correct car was used 97.52% [96.72% if including missing] of the time as compared to 98.41% [97.38% if including missing] of the time in the districts that used marked cars. 97.52% [96.72% if including missing] of the time, the unmarked districts allocated two officers to the project, while marked districts allocated two officers 72.05% [69.03% if including missing] of the time. In the unmarked districts, the officers reported being out of service for an extended period of time in 20.87% [18.31% if including missing] of cases. The most frequently occurring reasons for these departures were making arrests and assisting other officers. In the marked districts, the officers reported being out of service for an extended period of the time in 25.07% [23.10% if including missing] of cases. The most frequent reasons for these departures were making arrests and handling hospital cases. Altogether, marked cars report spending more than 60% of their shift within the predicted grid areas 89.92% [88.98% if including missing] of the time, while the unmarked cars report spending more than 60% of their shifts within the predicted grid areas 91.41% [90.17% if including missing] (Table 18).

Table 18 How much additional patrol time do you think your officers were able to provide?

	Condition	
	Marked	Unmarked
Most of the time	63.93	68.42
Much of the time	25.99	22.99
Some of the time	7.43	6.65
Just a little of the time	2.65	1.94

Accuracy

A second set of descriptive statistics seeks to assess the accuracy of the predicted areas according to the officers' perceptions. Busyness was assessed using a scale ranging from 1-10 with 1 indicating the least busy and 10 indicating the busiest. The officers within the marked districts reported a 3 or lower 74.40% [73.13% if including missing] of the time where officers in unmarked districts reported a 3 or lower 92.29% [91.53% if including missing] of the time. On the other end of the spectrum, only 4.81% [% if including missing] of officers from the marked districts reported a busyness score of 6 or higher. Similarly, only 2.20% [2.19% if including missing] of officers from the unmarked districts reported similar levels of busyness (Table 19).

Table 19 How busy were the predicted areas?

	Condition	
	Marked	Unmarked
1 (Really quiet)	40.53	46.01
2	23.2	29.48
3	10.67	16.8
4 (A few incidents)	14.13	3.86
5	6.67	1.65
6	2.4	0.55
7 (Busy on-and-off)	1.87	1.1
8	0.27	0.55
9	0.27	0
10 (Busy all the time)	0	0

When asked whether or not the officers agreed on where the grid cells were located, 53.62% [52.49% if including missing] of officers in marked districts agreed somewhat. In unmarked districts, 61.80% [60.11% if including missing] of officers agreed to some extent that the grids were appropriately located (Table 20).

Table 20 How accurate were the predicted areas?

	Condition	
	Marked	Unmarked
Agree Strongly	15.55	5.06
Agree Somewhat	38.07	56.74
Disagree Somewhat	26.54	24.44
Disagree Strongly	19.84	13.76

Shift content

The final research question was concerned with determining what actually occurred within the grid cells during the shifts. Concerning the major activity within the grid cells, 60.39% [57.22% if including missing] of marked officers reported responding to radio calls as their most frequent activity. 31.58% [29.92% if including missing] of marked officers reported that community policing was their most frequent activity, making it the second most cited activity within the grid areas. Concerning unmarked officers, 73.61% [68.58% if including missing] reported that responding to radio calls was their most frequent activity within the grid cells and 12.32% [11.48% if including missing] reported that community policing was their most frequent activity (Table 21).

Table 21 What was the major activity in the predicted area during the shift?

	Condition	
	Marked	Unmarked
Violent	0.83	0
Property	4.99	8.8
Radio	60.39	73.61
Drug/gang	1.39	2.64
Community	31.58	12.32
Radio & Community	0.28	0
Property & Radio	0	0.29
Radio & Drug/gang	0.28	1.47
Violent & Community	0.28	0.59
Drug/gang & Community	0	0.29

Routine patrols alone were cited as the most often employed overall main strategy by both marked (87.30%, 86.61% if including missing) and unmarked (82.97%, 82.51% if including missing) units (Table 22).

Table 22 What was the major strategy in the predicted area during the shift?

	Condition	
	Marked	Unmarked
Focus on offenders	2.38	7.97
Interact with public	8.2	4.95
Routine Patrol	87.3	82.97
Offender & Interact	0	0.27
Offender & Patrol	0.26	0.82
Public & Patrol	1.85	2.47
All of above	0	0.55

56.84% [55.65% if including missing] of marked officers reported getting out of their vehicle more than three times during a shift compared to only 39.56% [38.79% if including missing] of unmarked officers (Table 23).

Table 23 How many times did officers exit vehicle?

	Condition	
	Marked	Unmarked
Frequently (6+)	16.09	8.08
A handful of times (3-5)	40.75	31.48
Once or twice (1-2)	30.03	39.83
Not at all (0)	13.14	20.61

Burned

Unmarked officers were asked about how often they were identified as police officers. 72.44% [69.68% if including missing] of officers reported being identified as police at least once during their shift. The most frequent response indicates that 38.35% [36.89% if including missing] of officers were identified only once or twice to the responding officer's knowledge.

Property crime experiment phase (supervisor forms)

This section consists of the supervisor feedback forms from both marked and unmarked districts (N=701).

Resource allocation

The first set of descriptive statistics seeks to answer the research question concerning dosage. Specifically, this set of descriptive statistics sheds light on the degree that resources were correctly allocated to the prescribed areas. For districts that were running the project with unmarked cars, a dedicated car was assigned to the grids 84.92% [78.19% if including missing] of the time, compared to 96.44% [94.20% if including missing] of the time in the marked car districts. 66.17% of the time, the unmarked districts gave maps to only 1 or 2 officers, while marked districts gave no more than 2 officers maps 88.36% of the time. When asked if they had any additional units patrolling within the grid areas, supervisors in unmarked districts reported additional units present 60.52% [59.49% if including missing] of the time. Most frequently (57.22% of responses), the additional units within the target areas were car patrols. Additional unit types were rarely patrolling within the grids, with only 1.70% of supervisors responding that they had foot beats, 1.70% of supervisors had bicycle patrols, .57% of supervisors had specialized units, and .85% of supervisors had a type of additional units not specified in the survey patrolling within the target area. Supervisors surveyed in marked districts indicated that additional units were patrolling within the grid areas 59.05% of the time [57.68% if including missing]. Similar to the unmarked districts, the most frequent additional unit present within the predicted grid areas were car patrols (53.04%). Another parallel to the unmarked districts, additional units within the predicted areas aside from car patrols were exceedingly rare in the marked districts as well. Only 3.48% of supervisors indicated that they had foot beats, .87% of supervisors had bicycle patrols, .58% of supervisors had specialized units, and 1.74% of supervisors had a type of additional units not specified in the survey patrolling within the target areas.

Concerning the availability of the assigned car, in unmarked districts, supervisors reported their officers being out of service for an extended period of time in only 14.09% of shifts surveyed [11.61% if including missing]. The most frequently occurring reasons for these departures were making arrests, and going to court. For those who were out of service, 14.63% of the time they were able to replace the car. In the marked districts, supervisors reported their officers being out of service for an extended period of time in 22.93% [20.87% if including missing] of shifts surveyed. Supervisors reported the reasons being arrests, court, live stops, and lunch most frequently, along with a wide variety of other radio calls. For those who were out of service, 14.49% of the time they were able to replace the car. A final question asked how much patrol time officers were able to dedicate to the grids. In unmarked districts, 67.91% [67.14% if including missing] of supervisors reported their officers were able to provide at least occasional patrolling, however, only 28.08% [27.76% if including missing] reported that they were able to provide a lot of patrol time or full saturation of the predicted areas. In marked districts, 82.70% [81.74% if including missing] of supervisors reported their officers were able to provide at least occasional patrolling, with 61.59% [60.87% if including missing] reporting they were able to provide a lot of patrol time or full area saturation (Table 24).

Table 24 How much additional patrol time do you think your officers were able to provide?

	Condition	
	Marked	Unmarked
Saturation	13.2	1.72
A lot	48.39	26.36
Occasional/infrequent	21.11	39.83
A few drive throughs	14.37	22.35
None	2.93	9.74

Based on the supervisors' surveys during the property phase of the experiment, it would appear that the experimental districts allocated a fair amount of resources to the predicted areas during the experiment, however, the marked districts were able to allocate more resources than the unmarked districts. Concerning patrol time provided to the predicted target areas, marked districts were able to allocate a greater degree of time more frequently than the unmarked districts were able to. This is interesting as the unmarked districts were given a lesser degree of discretion regarding shift activity compared to their marked colleagues. One explanation for the exceedingly low frequencies where unmarked cars were able to saturate the target areas may be due to the wording in the survey. Supervisors may have interpreted the term as having multiple cars within the areas to achieve saturation. Concerning conditional differences further, the marked districts were more likely to provide maps of the targeted patrol areas to one or two officers compared to the unmarked districts. However, while the unmarked districts were less likely to provide maps to one or two officers, they were more likely to provide them to a greater number of officers compared to marked districts resulting in shifts in unmarked districts with greater numbers of officers aware of target areas. Also, unmarked districts were less likely to have their dedicated patrol cars go out of service for extended periods of time than the marked districts and were able to replace out of service cars at roughly the same rate as marked districts.

Accuracy

A second set of descriptive statistics is concerned with the supervisors' perceptions of accuracy regarding grid placement. On a 1-10 scale assessing busyness, with 1 being the least busy and ten being the busiest, 86.38% [84.43% if including missing] of supervisors in unmarked districts gave a 3 or lower. For supervisors of marked districts, 76.40% [75.07% if including missing] gave a 3 or lower. Only 2.06% [2.03% if including missing] of marked and 1.74% [1.70% if including missing] of unmarked district supervisors reported a score of 7 ('busy on and off') or higher (Table 25).

Table 25 How busy were the predicted areas?

	Condition	
	Marked	Unmarked
1 (Really quiet)	35.69	42.03
2	27.73	28.7
3	12.98	15.65
4 (A few incidents)	12.68	4.93
5	6.49	3.48
6	2.36	3.48
7 (Busy on-and-off)	1.77	1.74
8	0	0
9	0.29	0
10 (Busy all the time)	0	0

Regarding the supervisors' perceptions of predicted area accuracy, 50.29% [48.72% if including missing] of supervisors in unmarked districts agreed to some extent or agreed strongly. 63.23%

[62.31% if including missing] of supervisors in marked districts agreed to some extent or agreed strongly (Table 26).

Table 26 Based on your experience, do you think the predicted areas were in correct places?

	Condition	
	Marked	Unmarked
Agree Strongly	10.88	4.68
Agree Somewhat	52.35	45.61
Disagree Somewhat	24.12	33.33
Disagree Strongly	12.65	16.37

Considering the supervisors' surveys concerning accuracy, they reported higher levels of shift busyness in the marked districts than the unmarked ones. However, the busyness differences between condition types can be explained by the greater frequencies of marked districts reporting scores of 4 and 5 than the unmarked districts. This slight increase in terms of officer busyness in marked districts may shed light on why marked district supervisors were more likely to perceive the predicted grid locations as being accurately located within their districts than unmarked district supervisors. After the halfway point in the busyness scale, however, the differences between marked and unmarked districts melt away and become insignificant.

Shift content

A third and final set of descriptive statistics seeks to determine what activity occurred within the targeted grid areas. When asked what the major police officer activity within the grids had been, 77.94% [75.07% if including missing] of unmarked supervisors reported their officers were responding to radio calls, with the second highest category being responding to property crime at 10.59% [10.20% if including missing]. 64.72% [61.16% if including missing] of marked supervisors reported responding to radio calls as their officer's most frequent activity, with community policing in second at 24.54% [23.19% if including missing] (Table 27).

Table 27 What was the major activity in the predicted area during the shift?

	Condition	
	Marked	Unmarked
Violent	1.84	0
Property	6.13	10.59
Radio	64.72	77.94
Drug/gang	1.23	0.29
Community	24.54	10.29
Radio & Community	0.61	0
Property & Radio	0	0.88
Radio & Drug/gang	0.92	0

In the unmarked districts, supervisors reported out of the ordinary events occurring in the predicted grids only 2.87% [2.83% if including missing] of the time. The reasons reported were events such as homicide, an explosion, and assisting individuals on narcotics. In the marked districts, supervisors reported out of the ordinary events occurring in the predicted grids only 3.53% [3.48% if including missing] of the time. The reasons reported were events such as officers in court, shootings, and a VUFA arrest. In unmarked districts, supervisors reported out of the ordinary events happening in the rest of their district 11.11% [10.76% if including missing] of the time. The reasons reported were events such as bank robbery, bomb threats, and a body found in a wooded area. In marked districts, supervisors reported out of the ordinary events happening in the rest of their district 19.34% [18.55% if including missing] of the time. The reasons reported were events such as burglaries, homicides, and priority calls in other areas.

This set of descriptive statistics helps to shed light on the area of shift content which has been typically overlooked in traditional policing studies (i.e. Sherman & Weisburd, 1995). Regarding the shift content during the experiment, supervisors in both conditions cited responding to radio calls as the most common major activity conducted by the officers assigned to the predicted areas with unmarked district officers being more likely to respond to radio calls than their marked counterparts. Concerning out of the ordinary incidents occurring within the districts, both conditions cited similar occurrences which provides context to the types of incidents that occur within the city.

Violent crime experiment phase (officer forms)

This section reports on the officer feedback forms (n = 413) from both the marked as well as the unmarked districts.

Resource allocation

The first set of descriptive statistics seeks to answer the research question regarding how much resources were allocated, and correctly allocated, to the prescribed areas. For the districts that ran the project using unmarked cars, the correct car was used 98.32% [97.34% if including missing] of the time as compared to 96.43% [96.43% if including missing] of the time in the districts that were assigned to use marked cars. 98.65% [97.01% if including missing] of the time, the unmarked districts allocated two officers to the project, while marked districts allocated two officers 77.48% [76.79% if including missing] of the time. In the unmarked districts, the officers reported being out of service for an extended period of time in 18.86% [17.61% if including missing] of cases. The most frequently occurring reasons for these departures were making arrests. In the marked districts, the officers reported being out of service for an extended period of the time in 21.36% [19.64% if including missing] of cases. The most frequent reasons for these departures were making arrests and handling hospital cases. Altogether, marked cars report spending more than 60% of their shift within the predicted grid areas 77.06% [75.00% if including missing] of the time, while the unmarked cars report spending more than 60% of their shifts within the predicted grid areas 91.58% [90.37% if including missing] of the time (Table 28).

Table 28 How much additional patrol time do you think your officers were able to provide?

	Condition	
	Marked	Unmarked
Most of the time	42.2	64.98
Much of the time	34.86	26.6
Some of the time	19.27	6.4
Just a little of the time	3.67	2.02

Accuracy

A second set of descriptive statistics seeks to assess the accuracy of the predicted areas according to the officers' perceptions. Busyness was assessed using a scale ranging from 1-10 with 1 indicating the least busy and 10 indicating the busiest. The officers within the marked districts reported a 3 or lower 84.69% [83.93% if including missing] of the time where officers in unmarked districts reported a 3 or lower 94.28% [93.02% if including missing] of the time. No officer indicated a busyness greater than 6 in either the marked or the unmarked groups (Table 29).

Table 29 How busy were the predicted areas?

	Condition	
	Marked	Unmarked
1 (Really quiet)	36.94	55.56
2	25.23	25.93
3	22.52	12.79
4 (A few incidents)	8.11	4.04
5	6.31	1.35
6	0.9	0.34
7 (Busy on-and-off)	0	0
8	0	0
9	0	0
10 (Busy all the time)	0	0

When asked whether the officers agreed on where the grid cells were located, 30.91% [30.36% if including missing] of officers in marked districts agreed at least somewhat. In unmarked districts, 52.52% [51.83% if including missing] of officers agreed at least somewhat that the grids were appropriately located (Table 30).

Table 30 How accurate were the predicted areas?

	Condition	
	Marked	Unmarked
Agree Strongly	1.82	7.07
Agree Somewhat	29.09	45.45
Disagree Somewhat	50.00	25.59
Disagree Strongly	19.09	21.89

Shift content

The final research question was concerned with determining what actually occurred within the grid cells during the shifts. Concerning the major activity within the grid cells, 58.72% [57.14% if including missing] of marked officers reported responding to radio calls as their most frequent activity. 19.27% [18.75% if including missing] of marked officers reported that community policing was their most frequent activity, making it the second most cited activity within the grid areas. Concerning unmarked officers, 47.37% [44.85% if including missing] reported that responding to radio calls was their most frequent activity within the grid cells and 33.33% [31.56% if including missing] reported that community policing was their most frequent activity (Table 31).

Table 31 What was the major activity in the predicted area during the shift?

	Condition	
	Marked	Unmarked
Violent	0.92	2.81
Property	10.09	0.35
Radio	58.72	47.37
Drug/gang	7.34	14.04
Community	19.27	33.33
Radio & Community	0	0.7
Property & Radio	0.92	0
Violent & Community	2.75	0.35
Violent & Drug/gang	0	0.7
Violent & Radio	0	0.35

Routine patrols alone were cited as the most often employed overall main strategy by both marked (78.38%, 77.68% if including missing) and unmarked (80.00%, 79.73% if including missing) units (Table 32).

Table 32 What was the major strategy in the predicted area during the shift?

	Condition	
	Marked	Unmarked
Focus on offenders	14.41	16.33
Interact with public	4.5	2.33
Routine patrol	78.38	80
Offender & public	0	0.33
Offender & patrol	0	0.33
Public & patrol	0	0.33
All of the above	2.7	0.33

62.17% [61.60% if including missing] of marked officers reported getting out of their vehicle to handle stops and/or street activity more than three times during a shift compared to only 37.29% [36.54% if including missing] of unmarked officers (Table 33).

Table 33 How many times did officers exit vehicle?

	Condition	
	Marked	Unmarked
Frequently (6+)	26.13	6.44
A handful of times (3-5)	36.04	30.85
Once or twice (1-2)	29.73	31.86
Not at all (0)	8.11	30.85

Burned

Unmarked officers were asked about how often they were identified as police officers. 72.12% [70.43% if including missing] of officers reported being identified as police once or more. The most frequent response indicates that 39.80% [38.87% if including missing] of officers were identified more than six times.

Violent crime experiment phase (supervisor forms)

The section reports on the supervisor feedback forms from both marked and unmarked districts (N=323).

Resource allocation

The first set of descriptive statistics seeks to answer the research question concerning dosage. Specifically, this set of descriptive statistics sheds light on the degree that resources were correctly allocated to the prescribed areas. For districts that were running the project with unmarked cars, a dedicated car was assigned to the grids 93.10% [87.10% if including missing] of the time, compared to 89.13% [78.10% if including missing] of the time in the marked car districts. 96.51% of the time [89.40% if including missing], the unmarked districts gave maps to only 1 or 2 officers, while marked districts gave no more than 2 officers maps 96.88% [88.57% if including missing] of the time. When asked if they had any additional units patrolling within the grid areas, supervisors in unmarked districts reported additional units present 62.96% [62.67% if including missing] of the time. Most frequently (58.06% of responses), the additional units within the target areas were car patrols. Additional unit types were rarely patrolling within the grids, with only 1.84% of supervisors responding that they had foot beats, 0.46 % of supervisors had bicycle patrols, 5.07% of supervisors had specialized units, and 1.84% of supervisors had a type of additional unit not specified in the survey patrolling within the target area. Supervisors surveyed in marked districts indicated that additional units were patrolling within the grid areas 66.34% of the time [63.81% if including missing]. Similar to the unmarked districts, the most frequent additional unit present within the predicted grid areas were car patrols (56.19%). Another parallel to the unmarked districts, additional units aside from car patrols were exceedingly rare. Only 6.67% of supervisors indicated that they had foot beats, 1.90% of supervisors had bicycle patrols, 13.33% of supervisors had specialized units, and no supervisor reported a type of additional unit not specified in the survey patrolling within the target areas.

Concerning the availability of the assigned car, in unmarked districts, supervisors reported their officers being out of service for an extended period of time in only 17.68% of shifts surveyed [16.13% if including missing]. The most frequently occurring reasons for these departures were making arrests, and responding to incidents. For those who were out of service, 17.65% of the time they were able to replace the car. In the marked districts, supervisors reported their officers being out of service for an extended period of time in 17.95% [13.33% if including missing] of shifts surveyed. Supervisors reported similar reasons as the unmarked districts court. For those who were out of service, 14.29% of the time they were able to replace the car. A final question asked how much patrol time officers were able to dedicate to the grids. In unmarked districts, 77.21% [75.50% if including missing] of supervisors reported their officers were able to provide at least occasional patrolling, however, only 55.81% [55.30% if including missing] reported that they were able to provide a lot of patrol time or full saturation of the predicted areas. In marked districts, 100% [99.05% if including missing] of supervisors reported their officers were able to provide at least occasional patrolling, with 86.53% [85.72% if including missing] reporting they were able to provide a lot of patrol time or full area saturation (Table 34).

Table 34 How much additional patrol time do you think your officers were able to provide?

	Condition	
	Marked	Unmarked
Saturation	2.88	15.81
A lot	83.65	40
Occasional/infrequent	13.46	21.4
A few drive throughs	0	18.6
None	0	4.19

Accuracy

A second set of descriptive statistics is concerned with the supervisors' perceptions of accuracy regarding grid placement. On a 1-10 scale assessing busyness, with 1 being the least busy and ten being the busiest, 91.55% [89.86% if including missing] of supervisors in unmarked districts gave a 3 or lower. For supervisors of marked districts, 85.72% [80.00% if including missing] gave a 3 or lower. No supervisors in either the marked or unmarked districts gave a busyness score above 5 (Table 35).

Table 35 How busy were the predicted areas?

	Condition	
	Marked	Unmarked
1 (Really quiet)	21.43	47.42
2	42.86	25.82
3	21.43	18.31
4 (A few incidents)	10.2	7.98
5	4.08	0.47
6	0	0
7 (Busy on-and-off)	0	0
8	0	0
9	0	0
10 (Busy all the time)	0	0

Regarding the supervisors' perceptions of predicted area accuracy, 51.16% [50.69% if including missing] of supervisors in unmarked districts agreed to some extent or agreed strongly. 51.43% [no missing responses] of supervisors in marked districts agreed to some extent or agreed strongly (Table 36).

Table 36 Based on your experience, do you think the predicted areas were in correct places?

	Condition	
	Marked	Unmarked
Agree Strongly	0.95	3.72
Agree Somewhat	50.48	47.44
Disagree Somewhat	38.1	25.12
Disagree Strongly	10.48	23.72

Shift content

A third and final set of descriptive statistics seeks to determine what activity occurred within the targeted grid areas. When asked what the major police officer activity within the grids had been, 68.75% [60.83% if including missing] of unmarked supervisors reported their officers were responding to radio calls, with the second highest category being responding to drug and gang crime at 12.50% [11.06% if including missing]. 71.00% [67.62% if including missing] of marked supervisors reported responding to radio calls as their officer's most frequent activity, with community policing in second at 23.00% [21.90% if including missing] (Table 37).

Table 37 What was the major activity in the predicted area during the shift?

	Condition	
	Marked	Unmarked
Violent	0	5.73
Property	2	0.52
Radio	71	68.75
Drug/gang	2	12.5
Community	23	8.33
Radio & community	1	0
Radio & drug/gang	1	3.13
Violent & drug/gang	0	0.52
Drug/gang & community	0	0.52

In the unmarked districts, supervisors reported out of the ordinary events occurring in the predicted grids only 8.37% [8.29% if including missing] of the time. The reasons reported were events such as homicide and assisting patients on narcotics. In the marked districts, supervisors reported out of the ordinary events occurring in the predicted grids only 0.96% [0.95% if including missing] of the time. The reasons reported were events such as shootings and VUFA arrests. In unmarked districts, supervisors reported out of the ordinary events happening in the rest of their district 19.72% [19.35% if including missing] of the time. In marked districts, supervisors reported out of the ordinary events happening in the rest of their district 14.71% [14.29% if including missing] of the time.

RESULTS: FIELDWORK OBSERVATIONS

The multifaceted qualitative data that was captured by field researchers throughout the experiment helped to produce a more holistic view of the implementation challenges faced by the department, echoing Greene's (2014) call to explore the process concerns and contextual dynamics of innovation in policing. In particular, we were interested in uncovering how institutional-level inputs such as resource allocation and organizational culture coupled with perceptions of innovation effectiveness ultimately affect the implementation of this new patrol technology and strategy.

The following chapter describes the results of the fieldwork and the qualitative analysis and will explore the institutional-level themes of competing resource allocation for innovations and the effects of organizational culture on implementation and employee buy-in. In addition, the perceived ability and effectiveness of the predictive policing strategy and technology by front-line employees is a key component of the implementation process. These findings provide an invaluable glimpse into the organizational responses and implementation struggles faced by key personnel in a major police department when introducing a new operational paradigm dictated by predictive policing analytics.

Field observation forms

Field observers completed a one-page field observation form during each ride-along, noting where the officers spent the majority of their time for 15-minute blocks during the experimental treatment phase.

Table 38 Field codes for location and activity

Assigned code	Value
1	At station
2	In car outside grid areas
3	In car inside grid areas
4	In grids speaking to people
5	On break
6	Dealing with incident in grids
7	Dealing with incident outside of grids
8	Other
-99	Missing

As explained elsewhere in this report, there is a discrepancy between property crime phase and violent crime phase observation frequency; however, we were still able to elicit satisfactory measures of field observations with regard to where officers were spending their time, and how they were spending that time. The results are shown in the next two graphs.

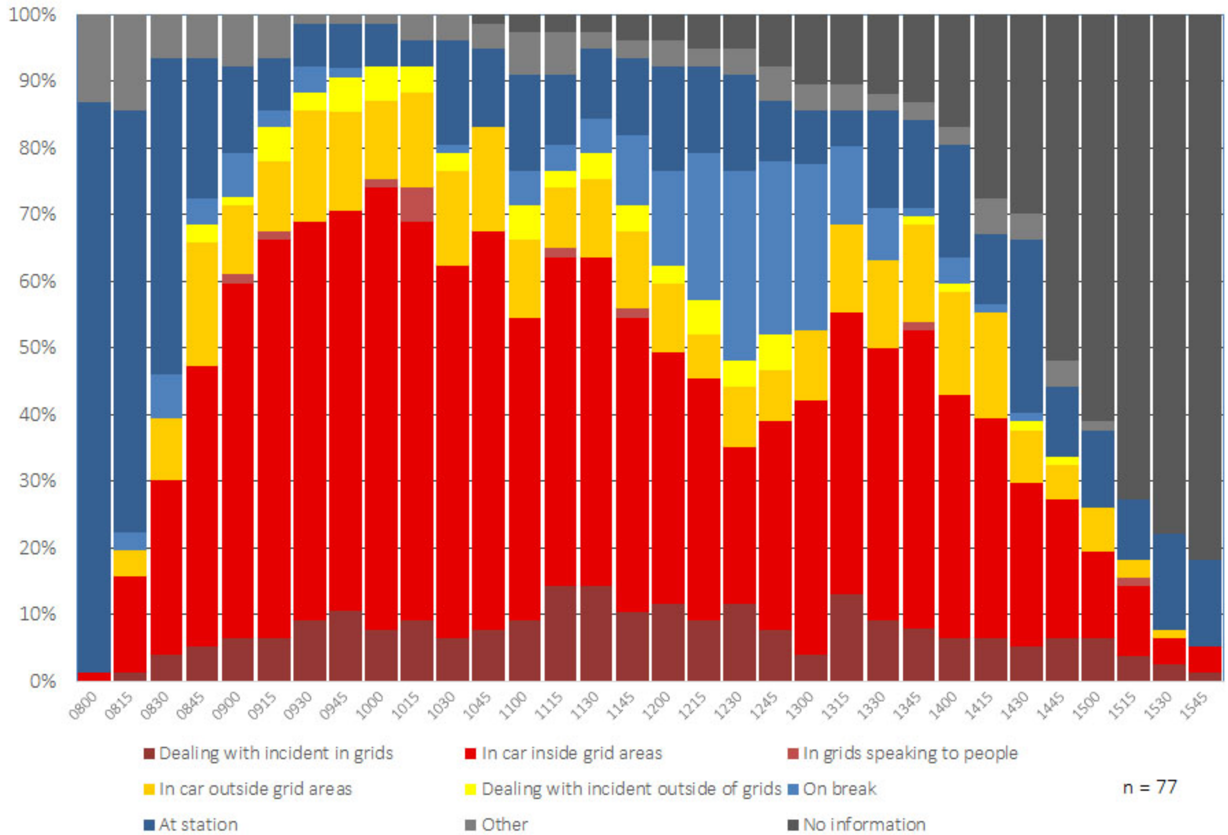


Figure 18 Dedicated car time from field observations (property)

As can be seen from Figure 18, the vast majority of officer time was spent inside grid areas (or very nearby) in the police car. The other category shown in red indicates time spent dealing with an incident. Categories indicated in blue show time at the station or on break, and a common time for refreshments for the officers is clear at about noon.

By comparison, in the evening shift for the violent crime phase, the officers spent a little longer at the station at the start of the shift but remained in the assigned grids later into the shift (Figure 19).

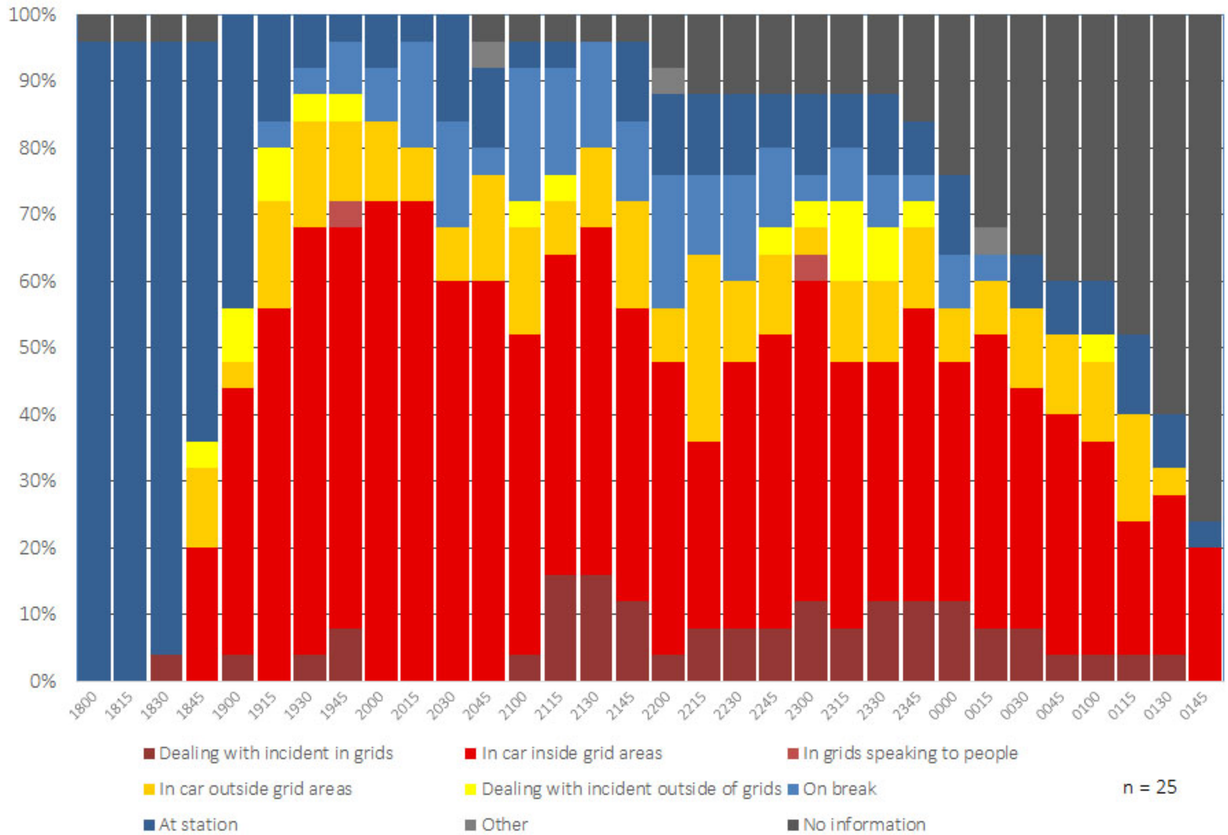


Figure 19 Dedicated car time from field observations (violent)

Institutional enabling mechanisms

The institutional resources dedicated to an innovation are a key determinant in the ultimate success or failure of the implementation process. Highly bureaucratic institutions depend on the buy-in and support of management and effective communication between individuals directly responsible for day to day operations. As of May 2016, the Philadelphia Police Department staffing levels had reached a 22-year low with a shortage of an estimated 400 sworn officers. This lead to the department exceeding its overtime budget by nearly $20 million (Allyn, 2016). Thus, the allocation of resources (i.e. personnel and material) and the organizational culture (i.e. skepticism towards outsiders and miscommunication) are important driving forces behind the implementation of new technologies and strategy in the department.

Methodological note: Please be aware that in the remainder of this chapter, direct quotes from officers are in bold font while field notes written by researchers are in italics.

Personnel allocation

Several officers and supervisors expressed concern with the implementation of the predictive policing experiment due to a lack of staffing, particularly during the property crime portion. An organization's inability to maintain sufficient numbers of front-line personnel during implementation can have an especially deleterious effect on the buy-in and effectiveness of an innovation (Clayton, 1997; Klein, Conn, & Sorra, 2001). One of the main issues identified by police personnel regarding the experiment was the fact that staffing levels were inadequate due to vacation time and higher call volumes. This increased their skepticism of the project itself, as it detracted from their ability to do their job adequately. As one supervisor pointed out;

"They are doing this during the time of year when manpower is already low…If I don't have an officer to respond to a call for a suspicious person…then that potential suspect is never apprehended and questioned" (Field Note #75).

In addition, some members of the department displayed outright hostility towards to the project due to the lack of resources; **"We hate the prediction project. I mean, if we had what we needed (staff and cars) it wouldn't be so bad"** (Field Note #45). Observers also noted throughout the property crime portion of the experiment, the department was forced to take '5-squad' officers (the department's specialized units of veteran officers) in order to remedy work force deficiencies; *The officers I was on the ride-along with were actually 5 squad officers who were working the shift temporarily to address the staffing issue* (Field Note #45). Understandably, the staffing commitment required for the experiment was concerning to a police department that was already struggling with staffing issues. Due to these staffing deficiencies, officers and supervisors felt that the experiment was detracting from their patrol duties and impeding the call response and crime-fighting orientation that police agencies value (Manning, 2010). Even officers who were receptive to the project noted the possibility of personnel shortages affecting the fidelity of the experiment itself, as one observer noted;

"The officers expressed that sometimes it is very difficult to maintain grid-patrol integrity as there is a lack of manpower (due to vacation, court appearances, overall hiring) and therefore they are forced to take calls in order to perform their duties" (Field Note #42).

The backup of call logs in the police districts were a hindrance to the ability of officers assigned to patrol the predicted grids throughout their shift, as they had been trained to respond to calls in a timely and efficient manner. Officers' reactions to their inability to answer calls due to the experimental conditions ranged from sarcasm to anger. During one ride-along, an officer expressly pointed out the backed-up call log to the observer; **"Look at all the jobs I have left [10] – So I'm *really* gonna patrol that grid area"** (sarcastic tone noted in Field Note #2). On another occasion, the observer noted; *The officers got more and more angry about the project with each passing call that they weren't allowed to attend. Each additional call would send them into a string of curses about the project and their time being wasted* (Field Note #82). It is evident that a lack of necessary institutional resources devoted to personnel can lead to issues with the buy-in of the front-line employees and the successful implementation and functioning of the innovation itself.

Reallocation of scarce resources

One of the biggest concerns in any organization is the amount of capital available to maintain the smooth functioning of their operation. Crucially, any time an organization is planning to incorporate an innovation in their daily operations, the financial investment on the part of the company is a critical determinant in the success of the implementation (Clayton, 1997; Klein et al., 2001). In particular, highly bureaucratic organizations such as a major police department often operate with tight, limited budgets and this can limit their ability to bring in expensive new technologies (Manning, 2010). The material resources devoted to the Philadelphia Predictive Policing Experiment were a frequent concern of the front-line personnel throughout the course of the project. A common sentiment of supervisors reflected the need for balancing the introduction of new patrol technology and strategy with the basic needs of the department. "**We need working computers and updated computer programs…or even just clean mops! Not this stuff** [referring to the predictive software]" (Field Note #6). Fiscally restricted organizations such as large, urban police departments which set out budgets at the beginning of each fiscal year are painfully aware of the amount of money that can be devoted to basic operations and upgrades, and the cost of new technology and policies. Even when the police department can secure grant money to test innovations (as with the predictive policing experiment), there are some initial discussions of monetary and resource allocation that don't eventually come to fruition, creating further tension during the implementation phase. As one supervisor noted to a researcher in explaining an issue with finding a car and officers to dedicate to patrolling the predicted grids for the shift; **“In the meeting with [a senior officer], they promised us cars and overtime. We still haven’t seen any of that”** (Field Note #75).

When front-line employees believe that the organization has not maintained its promises of monetary or resource compensation that was previously discussed, it often results in hostility towards the innovation itself. In situations like this, misinformation about expenditure on the innovation itself often perpetuate as well, demonstrated by an officer complaining about the project; **“I can’t believe we spent 4 million dollars on this HunchLab bull shit.”** *The other officer seemed surprised and questioned him, to which he responded*, “**I’m telling you, it’s a fact”** (Field Note #22). Even though the department did not spend even a fraction of $4 million dollars on the predictive software or the experiment as a whole (the experiment was conducted at no cost to the police department) this type of misinformation is transmitted through the ranks of front-line officers who are consistently confronted with the budgetary realities of having older vehicles and computers, or even a lack of clean mops. When front-line personnel within an organization are presented with technological or operational innovations but feel that basic resource allocations are not being met as is, it severely challenges the implementation of the innovation. For some officers, it erodes their confidence that the organization has fully committed to it; **“I just don’t see investment from the PD on this. It doesn’t seem like they are taking it seriously”** (Field Note #75). In addition, the lack of follow-through on resource allocation promises further diminishes employee confidence in the organization and the innovation, hampering successful implementation and may ultimately result in the perpetuation of misinformation regarding the new technology and operations.

Departmental skepticism

As noted in the previous section, supervisors in particular had a number of personnel and fiscal concerns with the project that may be indicative of a lack of overall support for the innovation itself. Police departments maintain a bureaucratic organizational structure which relies heavily on the delegation of control and command from the top-down. This type of structure emphasizes the role of supervisors and higher-ups in driving the organizational culture and daily operations of officers (Mastrofski & Willis, 2010). Due to the existing power structure of the department, contempt towards an innovation can be translated directly to the front-line officers as one researcher observed; *Officer 1 and Officer 2 stopped to talk with other officers who were at the station. They started talking about the project and the supervisor joined in on the conversation. They all expressed the same concern that the project was a waste of time* (Field

Note #6). The disdain for the experiment is thus being expressed at numerous levels of the organization. It is an interesting discussion point of had the supervisor stepped in to comment about the possible utility or benefits of the experiment, their higher standing may have influenced patrol officer attitudes or behaviors regarding the project. Patrol officers in the department are keenly aware of their position in the organizational hierarchy, as one officer pointed out **"…. the people who make decisions in the department are the higher-ups, not the people on the street…they don't ever really ask us [for input]"** (Field Note #90). Thus, due to their unique position within the organization, supervisors and other 'higher-ups' yield an immense amount of authority in determining the actions of front-line officers, as well as, their positions and attitudes towards an innovation. The institutional support fostered by managers and supervisors is crucial to the ultimate effectiveness of any innovation implementation. Unfortunately, supervisors in the department were resistant to the experiment as a whole and often the software itself;

I would walk into districts and be introduced to supervisors who would ask me 'Do you really think you can predict something using a computer?' That was primarily the largest issue that supervisors—and officers as well—had with this project: they had a very hard time accepting that computer software could be successful (Field Note #104).

Other staff were franker about their opinion; I was greeted by the lieutenant who was not bashful about expressing a general hatred for the prediction project by everyone in the district (Field Note, 102). This district-wide resentment to the project as a whole could potentially undermine the successful implementation of the experiment, as officers are less likely to dedicate the necessary attention and resources to ensure its success. Even with support from supervisors and other institutional elites, it is difficult to implement new practices based on technology into departments as it is;

"To be honest, most police officers are so hesitant to change, you're going to see that a lot. The last generation of police officer was furious about having to use these things [Mobile data terminals]. That was like ten years ago. I think this is going to be the new MDT – people are gonna complain, but they'll get used to it. That's just how it works, I guess" (Field Note #7).

Institution-wide resistance to change and a lack of support from certain supervisors in the department is a huge challenge facing successful innovation implementation in highly bureaucratic organizations.

Miscommunication

A related factor influencing the successful implementation of predictive policing software and the fidelity of the experiment itself was department-wide miscommunication. Although there were several memorandums sent around the department to explain the innovation, the experiment, and related procedures, there was still a great deal of confusion among supervisors and patrol officers regarding the experiment. One of the most evident and common miscommunication issues involved districts not being ready for the field researchers to ride-along with the patrol officers. There were over a half-dozen examples throughout the field notes of researchers waiting for confirmation, cars, or officers with whom to perform the ride-alongs (Field Notes #1; 2; 7; 9; 26; 91; 97). Although several of these instances were concentrated solely at the district level, there were times where the confusion stemmed from a lack of communication department-wide, as schedules for ride-alongs or other communications related to the predictive policing experiment were not disseminated to all of the participating districts (Field Notes #31; 37; 45). These may be administrative issues common to many large police departments, where (to be frank) the experimental needs of academic researchers are hardly a priority. Throughout the course of the project, there was also confusion from the patrol officers related to their expected patrol actions while inside the grids. It became clear that much of this confusion stemmed from either ineffective or insufficient communication from headquarters. As one field researcher noted; *Officers discussed the serious lack of communication between supervisors and officers about the purpose and clear directions about how to patrol in regards to HunchLab* (Field Note #48). Furthermore, researchers observed supervisors giving limited or incorrect information to the officers regarding the specific areas to be patrolled:

Sgt. drafted up a memo for the PTM car, which describes where the mission areas are located. The officers do not actually get a printed copy of the maps (Field Note #57).

The sergeant who assigned them wrote a couple of them down incorrectly, so [we] returned to the station and ironed it out (Field Note #67).

In addition, the lack of communication between officers and supervisors extended to their actual responsibilities while patrolling inside the grids. Officers described their confusion about the specific duties to be undertaken while in the grids, and the need for greater clarification (Field Note #9; 19). At times, the misunderstanding arose from departmental errors, as officers believed that they were only supposed to spend 25% of their time within the grids as that number was highlighted on the paperwork they were given at the start of the shift (Field Note #81; 86). However, even supervisors in the department appeared to be unclear on the expected activities of officers in the grids; **"All I was told to do was to answer priorities in the grids [by the sergeant]….[and was] following the last orders received"** (Field Note #100). Due to this type of confusion, researchers involved in the later portions of the experiment were encouraged to show officers a video that outlined their expected actions inside of the grid, which appeared to clarify some of these issues (Field Note #100; 101).

It is evident that attempts at introducing innovative technology and operational practices into a large, metropolitan police department face myriad challenges resulting from the institutional enabling mechanisms present within the organization. Staffing and material resource shortfalls are two large obstacles that were consistently identified by members of the department as barriers to successful implementation of the predictive policing experiment. When members of the department were confronted with an innovation that required a substantial monetary and staffing investment, there was considerable hesitation and resistance due to an apparent lack of institutional-wide devotion of resources. Furthermore, the successful implementation of the predictive policing experiment was threatened by occasional lapses of managerial support and subsequent miscommunication throughout the organization. Without strong reinforcement at the supervisory level, it reduced the buy-in of the front-line officers who were already dealing with confusion from lack of successful communication about the project. Thus, when instituting innovative technology and new policies into a police department, it is key that organizational resources and members are fully committed to ensure successful implementation.

Innovation utility and acceptance

The other key element driving successful innovation implementation is its functional utility and acceptance by front-line employees. As was discussed previously, members of the department felt that the predictive policing experiment was already consuming limited resources which would negatively impact departmental attitudes towards the experiment from the outset. Thus, if the innovation additionally appeared to lack information accuracy, or conflicted with officer's perceptions of police work, it is unlikely that implementation or adoption would ultimately prove successful. These initial issues could obscure the potential benefits of predictive policing strategies for front-line employees and make them less likely to fully implement the innovation as intended.

Information quality and patrol strategy

One of the most important aspects of technological innovations in policing and their ultimate acceptance by front-line officers is the timeliness of the information provided. If an officer is given information by the predictive policing software that isn't temporally oriented to the patterns that they see in their work, it ultimately won't prove useful to them. This in turn will make officers less likely to buy-in to the utility of the software and severely inhibit successful implementation. Although many officers found the grids were accurate, they ultimately were concerned with the time-ordering of the experiment itself and the way the grids were temporally constructed. Regarding the timeliness of the experiment and the patrols, a number of officers believed that daytime patrols were not going to be effective for property crime. **"Property crime happens at night. People call the next day when they go to their car and see that it had been broken into. If we want to prevent property crime, we need to do these patrols overnight when the crimes are happening"** (Field Note #74). On a related note, officers felt that day-time shifts lacked utility in catching criminals in the act because they operated on different schedules than the patrol shifts. **"The people we want to stop are still sleeping. They don't get up until 11:30 or 12:00"** (Field Note #57).

Another issue arose with the timeliness of the software itself, especially in regards to grid creation for violent crime. Several officers expressed that they believed there to be a temporal lag with regards to shootings and violent crime. In particular, it appeared to some officers that the software wasn't adjusting accordingly when it came to more recent violent crime in the areas. Sometimes, the grids remained unchanged day after day; *They [the officers] were disconcerted by the fact roughly the grids in this part of the district showed up night after night:* "*Three nights in a row, right D?"* (Field Note #96). In other cases, the grids 'missed' crime; *a homicide had recently occurred* **"…off of homicide by 1 block"** *but the grid boundaries had not adjusted in the days since.* **"There is something like a shooting, but it [the program] does not put you there the next day."** (Field Note #96). Other officers argued that the grids were too sensitive temporally; **"One area had been a grid and then two days later it wasn't and THEN there was a homicide there. This grid wasn't a grid yesterday and there was a robbery last night, and now today it's one"** (Field Note #85). When officers are unable to rely on the timeliness of the information given by the innovative technology, they are much less likely to trust its utility.

Another concern of department members was the physical map output that the predictive software provided. Officers and supervisors pointed out that the grid maps were confusing, and often lacking in important details such as street names (Field Note #4; 57; 102). This led to confusion about the grid boundaries, which could only be rectified by manually looking up the areas on a map to orient the officers. The readability of the map output was an immediate barrier for officers and supervisors, especially considering their hesitancy about the predictive policing experiment in the first place. As one officer pointed out to the researcher; *In short, they wanted to push one button…that would generate large grids on a page, with clear street names, and the grids oriented to the street network to make them easier to interpret* (Field Note #96). In addition, including more information on the maps could potentially have increased officer buy-in and more seamless implementation of the project; **"It would be nice to see some more info attached to these grids. That could really help us out here. Like residential burglary? What kinda burglary are we talking here?....That little bit of info could really help direct our focus."** (Field Note #7).

In addition to the physical grid output themselves, officers and supervisors questioned some of the grid locations that the predictive policing software created. In other cases, the officers pointed out; *"…the grid locations were in awkward areas that did not match the street network…One grid area, for instance, included two large churches and a small business district. The churches took up a majority of the grid area, which meant we would in essence need to circle the churches to patrol the grid*" (Field Note #29). There were also instances where grids were created for the patrol shift that had little to no functionality; *Two of the grids were centered on railroad tracks, allowing only one small block on either side for patrolling. The supervisor asked if there was some way to patch the software to ignore dead space like railroad tracks, and focus on actual street blocks* (Field Note #18). On occasion, the grids were seemingly created in error as they could not be patrolled at all; *HunchLab produced three grids that were impossible to patrol. The first two were a park and a private field that was property of the Northeast Airport. The third grid was in the middle of a cloverleaf highway* (Field Note #77). Therefore, officers struggled at times to find the utility in a program that wasn't geographically or temporally accurate enough to assist in their actual patrol functions. Not only would this cause frustration on the part of officers and decrease their innovation acceptance, but would also severely undermine successful implementation of the project. It is crucial that any new technological innovation intended for use by front-line officers of a department focus on timely, readable, easily interpretable information and output that has enough information to be directly actionable for daily patrol.

Computer/algorithm vs. experiential 'street' knowledge

Due to the craft orientation of policing, individuals within a police department are often quite skeptical of innovations that come from outside of this culture. The craft focus of the policing community tends to emphasize the importance of experiential knowledge and skill, while devaluing or resisting more scientific forms of knowledge (Thacher, 2008; Willis & Mastrofski, 2014, 2016). As expected, there was a good deal of incredulity on the part of the members of the police department regarding the predictive policing software itself. The use of something as abstract to officers as a computer algorithm dictating the patrol tactics of a department by predicting daily high-crime areas was unbelievable to many of the officers, as one explained **"…no computer algorithm can interpret real life"** (Field Note #25). The lack of a 'human element' to the software seemed to bother many of the officers. This was a common refrain—essentially that without the experiential knowledge gleaned from years patrolling the streets and working in the field, the information was never going to be correct. Several officers took umbrage with the grid locations provided by the program, explaining **"You [the researcher] need to come and ask street officers where the boxes should be, they know where crime is"** (Field Note #1). One supervisor even went so far as to invite the research team and the software developer to **"…come to the [99th] district, you can see our crime maps, to see where crime really happens"** (Field Note #18). Understandably, members of the department found it implausible and unrealistic that any software or computer algorithm could replicate the first-hand knowledge and experience of a seasoned officer. In addition, the experimental conditions in combination with the software eliminated the autonomy and control that officers and supervisors typically maintain regarding daily patrol operations.

Even though the software was not intended to replace the essential craft of policing but used as a tool to improve patrol functions, some were vehemently opposed. When asked about any improvements that could be made to the software, one officer eloquently explained; **"…. instead of slapping squares on a map and hoping they fucking stick…Tell them to have real people who have been out in this [patrolling the community], not behind a desk all day"** (Field Note #82). This quote highlights perhaps the biggest issue that department members had with the software itself—the absence of accumulated experiential knowledge. As mentioned previously, police culture is often resistant to scientific knowledge, especially when it comes from outside the ranks of police. Thus, predictive policing strategies and software that lack any human element, first-hand experiential knowledge, and is a product from outside of their professional or peer network is likely to be met with a good deal of resistance.

Usefulness

If the front-line employees of the organization don't find the innovation itself useful or productive, they are less likely to accept it. Thus, decreasing implementation fidelity and ultimately, the successful integration of the innovation into their operations (Klein et al., 2001). Despite some of the issues surrounding the technology and the patrol strategies, some officers and supervisors expressed support for the project and its possible utility in daily operations. These officers typically endorsed the perspective that change can be a positive benefit for the department as a whole, and the role that technology could play in these changes:

The officer also mentioned how she was supportive of any types of changes in policy trying out new things (Field Note #70).

[The] officer mentioned that he likes the idea of incorporating more technology into policing (Field Note #13).

A crucial element of 'good police work' and the craft orientation of officers is the intimate knowledge of their patrol environment. One of the consistent themes of the officers who embraced the utility of predictive policing and the project was the opportunity to be present in places that may not be a typical patrol area. *Another officer commented that she liked being in the predictive grids because the grids would occasionally take the officer down streets that she had never patrolled before. This officer has been in the district for quite a few years and she was surprised there were so many streets that she has never patrolled* (Field Note #107). Since these were predictive grids, it made it possible for some of these officers to be present for the possible commission of a crime or as a form of deterrence. *Officer 1 emphasized that the usefulness of the grids in predictions, as "3/4 of police work is right place, right time" and therefore the grids could provide officers with better chance to be in the "right place"* (Field Note #67).

As the officers were required to spend most of their time patrolling these smaller grid areas, they were more diligent about investigating new areas: *They expressed that the grids encompass areas that very rarely see patrols due to this fact, and that there were certain alleyways/areas that they had never seen before the institution of the grid patrols* (Field Note #49). By patrolling alleyways and other areas within these grids that crime was predicted to occur, there is a much greater chance that they could prevent crime before it happens, or even catch someone in the commission of a crime. Furthermore, by patrolling in these smaller areas, **"…it gave the chance for communities to see an increased police presence"** (Field Note #70).

Another significant component of the 'craft' of policing is the interaction between officers and citizens as this greatly enhances their ability to instruct, compel, and understand the individuals within the community while helping to establish relationships that may provide useful information in the future (Bayley & Bittner, 1984; Willis & Mastrofski, 2016). Over the course of the experiment, researchers observed the lengths to which officers would interact with the citizens of the community and the value they placed in these interactions. Many of these interactions were friendly and intended to build better relations between the community and the police. As one researcher noted; *Officers consistently interacted with individuals on the street- often saying hi to passerby or even stopping to chat with some individuals. They both stressed the importance of community relations and being cordial with citizens* (Field Note #56). As the experiment dictated that officers focus their patrols on smaller areas, they were more likely to interact with the citizens and get to know them better, which could potentially improve this aspect of their craft. Certain officers were aware of this potential benefit, and pointed this out to one of the researchers; *The officers felt that the experiment allowed them to establish a presence within the grids, which allowed them to interact with citizens much more than they typically would* (Field Note #14). The potential benefit of increased citizen interactions is important for the craft aspect of policing as citizens are often the ones who provide the most helpful information to officers. As one officer pointed out; **"I think it's productive. We're gathering intelligence, we are speaking to people"** (Field Note #50). Throughout the course of the experiment, officers emphasized this to the research team;

"It's always great to talk to people, I find. They tell you some important stuff. I have a bunch of people who, yeah they're criminals and they know if I find them doing something illegal, I'm gonna lock them up, but they tell me what I need to know. Like, 'Hey, did you hear about that shooting?' 'No, I didn't, why don't you tell me about it.' And that's how you get some good information." (Field Note #7).

Ultimately, for any innovation to be successful it must provide tangible benefits to the front-line employees that will be utilizing it on a regular basis. Although at times, the predictive policing software and associated patrol strategies proved to be temporally and spatially inaccurate patrol officers recognized its potential utility and benefits. By patrolling in these micro-grid zones dictated by the predictive policing experiment, the officers are interacting with the community much more, gathering intelligence, and are able to focus their 'craft' on smaller areas. In addition, it allows officers to patrol in areas that may have been neglected or unknown previously as being at high risk for crime incidents, increases their effectiveness at preventing crime or responding quickly by being in the 'right place' at the 'right time'.

Summary

In summation, the qualitative data presented highlights the challenges inherent in innovation implementation within a major metropolitan police department. The use of field researchers to observe the actions and attitudes of front-line officers yielded a large, rich dataset that provides a more nuanced understanding of the process concerns when implementing predictive policing software and tactics. Focusing on the challenges faced during the Philadelphia Predictive Policing Experiment rather than the success or failure of the innovation itself is an understudied aspect within the policing innovation literature. However, this line of inquiry is important to understanding exactly how policing organizations deal with change and should prove informative when applying new strategies and technology in the future. Beyond the institutional inputs which remain crucial to implementation, there is a clear need to address the buy-in and attitudes of front-line employees regarding the innovation. Specifically, this research underscores the significance of unifying the craft-orientation of police with scientific and technological innovation and the strategic or organizational changes required for successful operationalization. Moving forward, it is critical to the evolution of policing that scholars and practitioners work in conjunction to ensure that any technological or strategic innovation addresses these issues to ensure successful implementation.

Certain themes emerged from the qualitative work. For traditional police departments that equate responding to calls for service and seeking out suspicious persons (and supporting other officers in these activities), anything that takes resources from these endeavors is viewed with suspicion and often considered to have a negative impact on the craft of policing. Couple this with the existing lack of personnel in the police department, and it was clear that the predictive policing experiment –and predictive policing in general—was not positively supported.

Furthermore, skepticism and miscommunication hampered the implementation of predictive policing. The former has been long recognized as a stalwart police response to many initiatives, and it may be that this skepticism affected the communication that was deemed necessary to a strong implementation of the project. Skepticism might hamper a mid-level commander's willingness to faithfully pass on necessary details, or might inadvertently affect an officer's enthusiasm to pay attention to received instructions (a challenge for officers at every rank).

Conflict between the software predictions and the front-line officers' perception of appropriate crime locations hampered implementation because sometimes officers were concerned that the software was predicting poorly. This challenge provides an interesting conundrum, because few officers provided any evidence to support their opinion of the crime hot spots. So it may be that the software predicted the correct locations, but because they did not mesh with the officer's experience or opinion, the predictions were discarded. This gets to the core of the problems with comparing actuarial data from software with clinical (officer experience) opinions. The latter are rarely evaluated (for a rare exception see Ratcliffe & McCullagh, 2001). The comparison is thus between evidence-based data and untested experience.

Finally, the research supports other studies that have concluded that officers are "much more likely to use IT to guide and assist them with traditional enforcement-oriented activities" (Koper, Lum, Willis, Woods, & Hibdon, 2015: 4) than to use technology to engage with new practices.

CONCLUSIONS

Discussion of findings

This study set out to compare different types of police responses used in predicted crime areas. The analysis was structured to test four hypotheses. We do not find evidence to support hypothesis 1: *greater awareness among general duties patrol officers of the predicted crime areas was not sufficient to deter crime.* We did, however, find some evidence to support hypothesis 2: *a dedicated uniform patrol attendance in predictive areas will increase visible police presence sufficiently in the local area to deter crime.* Although floor effects inhibited our ability to detect a statistically significant effect, we found that the marked patrol car condition resulted in a 31 percent reduction in property crime if examining counts of crime, or a 36 percent reduction in the number of cells experiencing at least one crime. We did not find a similar effect for the unmarked condition. Hypothesis 3, therefore, was not supported. *Dedicated plain-clothes units performing surveillance and unmarked patrol did not increase interdiction and offender incapacitation sufficiently to reduce crime.* Finally, hypothesis 4 predicted that each condition would *cause temporal displacement and increase crime in the predicted areas in the hours following an intervention.* In the eight hours after the property treatment, the marked car districts were actually associated with a *reduced* crime compared to the control areas, such that expected crime counts that were 41.6 percent lower and expected crime occurrences that were 48.1 percent lower.

Even though we have some supporting evidence for crime reduction, the quantitative findings have to be viewed in the context of the implementation and the qualitative research. For traditional police departments that equate responding to calls for service and seeking out suspicious persons (and supporting other officers in these activities), anything that takes resources from these endeavors is viewed with suspicion and often considered to have a negative impact on the craft of policing. It is therefore not surprising that an untested innovation like predicted policing would run into some resistance and skepticism.

The software did a good job of predicting crime areas; however, it is also clear that crime at the micro-level is difficult to predict for a narrow eight-hour window. Given the absence of crime on some days, conflict between the software predictions and the front-line officers' perception of appropriate crime locations hampered implementation because sometimes officers were concerned that the software was predicting poorly. The predicted areas did not mesh with the officer's experience or opinion. We saw evidence of skepticism and disdain for the project as a result. And as stated above, we would concur that officers are "much more likely to use IT to guide and assist them with traditional enforcement-oriented activities" (Koper et al., 2015: 4) than to use technology to engage with new practices.

All this means that implementation of predictive policing is not a software issue. It is a personnel issue that has a technology component. Without addressing the context of the police department's operational ethos and how officers perceive their role, no technology implementation will be an overwhelming success if it challenges the status quo. Considerable effort would appear to be needed if officers are to accept and embrace a change to long-established practices based on relatively untested technology and practices. That being said, the Philadelphia Police Department did make considerable inroads to help the implementation and there is evidence from the field research of a reasonable implementation of the marked and unmarked car interventions. We can therefore make the case that the quantitative results are reflective of a realistic implementation of the project.

Limitations

Predicting crime and the tactics used in predicted crime areas are arguably two distinct considerations. As such, this experiment did not aim to explicitly evaluate the first consideration, the effectiveness of HunchLab as a crime prediction tool. Consequently, the experiment is constrained by the effectiveness of the software. It is possible that different results might have been observed using different software, or using HunchLab with different parameters. Furthermore, it is almost certainly the case that the software will perform more effectively when released from the constraints of the demands arising from running a randomized and controlled experiment.

The intervention is evaluated here 'as intended.' We present 'intention to treat' analyses. Inevitably, operational constraints prevented the police department from achieving a complete implementation of the experimental treatments. For example, on the first day, it took some districts a couple of hours to get up and running because they were unable to log into the software's web platform. Sometimes in the car districts, officers made arrests and the police district was unable to spare personnel to replace the operational crew for the remainder of the shift. And sometimes it was the case that a vehicle broke down and a replacement was not available. We examine the data as per treatment intent, though the reader is cautioned that inevitably operational policing constraints will interfere with planned deployments. If this is seen as a constraint, we would argue that it is a constraint that reflects the reality of policing environments the world over.

A further consideration relates to police crime recording practices. Property crime suffers from inaccuracies in crime reporting since victims are sometimes not present during the crime event. While techniques such as aoristic analysis (Ratcliffe, 2000; Ratcliffe & McCullagh, 1998) can help, limitations with data reporting precluded our using this approach. In lieu of this, we were limited to using the time that the crime was reported to police.

Discussions with police personnel during field observations did raise the possibility that some of the unmarked police cars got 'burned.' That is, the community, including the potential offenders within that community, were well aware that the unmarked cars were police vehicles. In theory, this could change the nature of the intervention in those districts, at least in part – depending on how many members of the community recognized the vehicles as belonging to police. The extent of this inadvertent experimental condition change is difficult to estimate, though given the experimental constraint of driving around a small number of streets for extended periods of time, it certainly cannot be ruled out. Nevertheless, given the notably different property crime impacts for the marked as compared to the unmarked condition, we do not think that unmarked cars, broadly, proved as visible to the community as the marked patrol cars.

Perhaps one of the biggest limitations arises from crime infrequency at extremely small spatial and temporal scales. The floor effects, and the attendant impaired levels of statistical power are sketched in more detail in the online appendix. This seems to be an inevitable adverse side effect of the geographic fine tuning accompanying current predictive policing efforts.

Costs

Our original discussions with Philadelphia police did raise the possibility of overtime payments to fuel the additional cars for the marked and unmarked car condition. Regrettably, these funds never materialized. As a result, vehicles for the marked and unmarked car conditions were pulled from general duties. No additional costs were therefore incurred by the police department. Instead, any experimental findings represent an internal reallocation of resources within each district.

Generalizability

As Weisburd and colleagues note, "police and scholars must ask not just 'what the strategy is' that will be effective but also 'how we gain enough dosage' for such strategies to have large-scale crime prevention impacts across the city" (Weisburd et al., 2015: 384). This experiment was designed in close collaboration with the Philadelphia Police Department leadership team so that the intervention conditions reflected realistic treatments that a police department could sustain indefinitely. The awareness condition required simply passing on information about the predicted grids, while the two vehicle treatments necessitated only a realignment of existing resources. In terms of the marked cars, these were usually reassigned cars that would be used for patrol. The unmarked cars were usually not available to the patrol officers; however, it did require reassignment of uniform personnel from the available patrol force to use the unmarked vehicle. While external validity is ultimately an empirical question (Ralph B. Taylor, 1994), our ad hoc discussions with police personnel from other departments suggests that our experimental conditions reflect the likely manner by which other police departments would implement a long-term predictive policing strategy.

Interpretation

If predictive policing is, in essence, an application of micro-level hot spots policing, then there is a need to consider explicitly the strategic translation of a place-specific micro-level policing tactic up to a district-wide or city-level implementation (L. W. Sherman et al., 2014). Reliably predicting crime to small grids comprising little more than a single street block remains a challenge for the nascent discipline of predictive crime analysis; further, the vagaries of stochastic and dynamic micro-locations, can play havoc with even the best predictions. As a result, we chose to expand the analytical areas to more closely reflect the realities on the ground. Even then, because of the small areas and limited periods of predictions, we experienced floor effects, especially in terms of the violent crime phase. As such, we caution against interpreting the results purely in terms of patterns of statistical significance. The odds ratio changes with regard to the property crime phase and the marked car units are substantial, with the caveat that the numbers remain relatively small.

Overall, the quantitative results suggest that a dedicated marked police car can reduce property crime, as long as participants and the community manage their expectations in terms of the volume of crime prevented. Small micro-grids will never generate a massive volume of crime in a short time frame, therefore the raw numbers will remain small.

That said, this is the first robust empirical evidence derived from a randomized experiment of a tactic tied to a predictive policing implementation demonstrating some crime reduction. At present, there is little available evidence regarding how best to use shift-based patrol police officers beyond hot spots policing. This experiment provides an additional avenue for patrol commanders.

This remains the result of just one experiment. Not only are further experiments needed that have the highest standards of scientific validity; such investigations also should test deployments that are both realistic and operationally viable. If the current enthusiasm for predictive policing is to develop into a mainstream strategy, there is no point testing resource intensive law enforcement strategies that police departments cannot sustain on an ongoing basis.

DISSEMINATION OF RESEARCH FINDINGS

Project results have been distributed in the following ways.

Journal articles

Ratcliffe JH, Taylor RB, Perenzin Askey A, Thomas K, Grasso J, Bethel K (under review) *The Philadelphia Predictive Policing Experiment*. *Criminology*.

Further articles are in preparation.

Conference presentations

Past and anticipated conference presentations are listed here.

- Ratcliffe (2015) ‘3PE: The Philadelphia Predictive Policing Experiment’, plenary presentation outlining the experimental design to the 9th Evidence-Based Policing Conference, University of Cambridge, 7th June 2015.
- Ratcliffe (2015) ‘3PE: The Philadelphia Predictive Policing Experiment: Preliminary results’, plenary presentation to the 11th Evidence-Based Policing Conference, University of Cambridge, 10th July 2017.
- Ratcliffe (2017) ‘Results from the Philadelphia Predictive Policing Experiment’, paper presented in panel, ‘The Philadelphia Predictive Policing Experiment’ at the annual meeting of the American Society of Criminology, Philadelphia, PA, 16th November 2017.
- Grasso, J and Fisher, R (2017) ‘The officer experience of a predictive policing experiment’, paper presented in panel, ‘The Philadelphia Predictive Policing Experiment’ at the annual meeting of the American Society of Criminology, Philadelphia, PA, 16th November 2017.

- Thomas, K (2017) ‘Organizational benefits and challenges of an RCT’, paper presented in panel, ‘The Philadelphia Predictive Policing Experiment’ at the annual meeting of the American Society of Criminology, Philadelphia, PA, 16th November 2017.
- Taylor, R.B. (2017) Discussant on above panel, ‘The Philadelphia Predictive Policing Experiment’, the annual meeting of the American Society of Criminology, Philadelphia, PA, 16th November 2017.
- Ratcliffe (2018) ‘The Philadelphia Predictive Policing Experiment: A summary of the main findings’, paper to be presented at the annual meeting of the Western Society of Criminology, Long Beach, CA, 3rd February 2018.
- Ratcliffe (2018) ‘The Philadelphia Predictive Policing Experiment’, paper to be presented at the annual meeting of the Academy of Criminal Justice Sciences, New Orleans, LA, 17th February 2018.
- Ratcliffe (2018) ‘The Philadelphia Predictive Policing Experiment’, paper to be presented at the 2nd annual meeting of the American Society of Evidence-Based Policing, Philadelphia, PA, 21st May 2018.

Web sites

A web site to advertise the project and disseminate results in available at Temple University.

http://www.cla.temple.edu/cj/center-for-security-and-crime-science/the-philadelphia-predictive-policing-experiment/

There is also a shortlink to the same site: **http://bit.ly/CSCS_3PE**

The website contains a brief summary of the experiment as well as a number of short 2-3 page pdf documents explaining various parts of the experiment. These have been included as appendices to this document. In time we will include links to any peer-reviewed journal articles.

REFERENCES

Allyn, B. (2016). *18). As quest for new police recruits continues, Philly department scraps college-credit requirement. WHYY Newsworks*.

Azavea. (2014). *HunchLab: Under the Hood*. Retrieved from Philadelphia, PA:

Bayley, D. H., & Bittner, E. (1984). Learning the skills of policing. *Law and contemporary problems, 47*(4), 35-59.

Berk, R. A. (2005). Randomized experiments as the bronze standard. *Journal of Experimental Criminology, 1*, 417-433.

Bhati, A. S. (2007). Estimating the number of crimes averted by incapacitation: An information theoretic approach. *Journal of Quantitative Criminology, 23*(4), 355-375.

Boruch, R., May, H., Turner, H., Lavenberg, J., Petrosino, A., de Moya, D., . . . Foley, E. (2004). Estimating the Effects of Interventions That are Deployed in Many Places: Place-Randomized Trials. *American Behavioral Scientist, 47*(5), 608-633. doi:10.1177/0002764203259291

Bowers, K. J., & Johnson, S. D. (2003). Measuring the geographical displacement and diffusion of benefit effects of crime prevention activity. *Journal of Quantitative Criminology, 19*(3), 275-301.

Bowers, K. J., & Johnson, S. D. (2004). Who commits near repeats? A test of the boost explanation. *Western Criminology Review, 5*(3), 12-24.

Box, J. F. (1980). R. A. Fisher and the Design of Experiments, 1922-1926. *The American Statistician, 34*(1), 1-7. doi:10.2307/2682986

Braga, A. A. (2005). Hot spots policing and crime prevention: A systematic review of randomized controlled trials. *Journal of Experimental Criminology, 1*(3), 317-342.

Braga, A. A., & Bond, B. J. (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology, 46*(3), 577-607.

Braga, A. A., Papachristos, A., & Hureau, D. (2012). *Hot spots policing effects on crime*. Retrieved from Oslo, Norway:

Braga, A. A., Weisburd, D. L., Waring, E. J., Mazerolle, L. G., Spelman, W., & Gajewski, F. (1999). Problem-oriented policing in violent crime places: A randomized control experiment. *Criminology, 37*(3), 541-580.

Braga, A. A., Welsh, B. C., Papachristos, A. V., Schnell, C., & Grossman, L. (2014). The Growth of randomized experiments in policing: The Vital few and the salience of mentoring. *Journal of Experimental Criminology, 10*, 1-28.

Brantingham, P. L., & Brantingham, P. J. (1981-2). Mobility, notoriety, and crime: A study in the crime patterns of urban nodal points. *Journal of Environmental Systems, 11*(1), 89-99.

Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk Terrain Modeling: Brokering criminological theory and GIS methods for crime forecasting. *Justice Quarterly, 28*(2), 381.

Chainey, S., & Ratcliffe, J. H. (2005). *GIS and Crime Mapping*. London: John Wiley and Sons.

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal, 21*(1-2), 4-28.

Chilvers, M., & Weatherburn, D. (2001a). Do targeted arrests reduce crime? *Contemporary Issues in Crime and Justice (NSW Bureau of Crime Statistics and Research), Paper 63*, 16.

Chilvers, M., & Weatherburn, D. (2001b). Operation and Crime Review panels: Their impact on break and enter. *Crime and Justice Statistics: Bureau Brief (NSW Bureau of Crime Statistics and Research)*.

Clayton, P. C. (1997). *Implementation of organizational innovation: studies of academic and research libraries*: Academic Press.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*, 829-836.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second Edn. ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44*, 588-608.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation*. Chcago: Rand-McNally.

Cornish, D., & Clarke, R. (1986). *The Reasoning Criminal: Rational Choice Perspectives on Offending*. New York: Springer-Verlag.

Detry, M. A., & Lewis, R. J. (2014). The Intention-to-treat principle: How to assess the true effect of choosing a medical treatment. *JAMA, 312*(1), 85-86.

Eck, J. E., Lee, Y., O, S., & Martinez, N. (2017). Compared to what? Estimating the relative concentration of crime at places using systematic and other reviews. *Crime Science, 6*(8), 1-17.

Farrell, G., Chenery, S., & Pease, K. (1998). *Consolidating police crackdowns: findings from an anti-burglary project* (113 (Police Research Series)). Retrieved from London: file://C:%5CDocuments%20and%20Settings%5Cjhr%5CMy%20Documents%5CVirtual%20library%5CVirtual%206%5CFarell%20et%20al%20(1998)%20Consolidating%20Police%20Crackdowns.pdf

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavioral Research Methods, 41*(4), 1149-1160.

Felson, M. (1987). Routine activities and crime prevention in the developing metropolis. *Criminology, 25*(4), 911-932.

Fisher, R. A. (1935). *Design of Experiments*. Edinburgh: Oliver and Boyd.

Gill, C. E., & Weisburd, D. (2013). Increasing equivalence in small-sample, place-based experiments: Taking advantage of block randomization methods. In B. C. Welsh, A. A. Braga, & G. J. N. Bruinsma (Eds.), *Experimental Criminology: Prospects for Advancing Science and Public Policy* (pp. 141-162). Cambridge: Cambridge University Press.

Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term forecasting of crime. *International Journal of Forecasting, 19*(4), 594.

Gorr, W. L., & Lee, Y. (in press). Early warning system for temporary crime hot spots. *Journal of Quantitative Criminology*.

Greene, J. R. (2014). New directions in policing: Balancing prediction and meaning in police research. *Justice quarterly, 31*(2), 193-228.

Groff, E. R., Ratcliffe, J. H., Haberman, C., Sorg, E., Joyce, N., & Taylor, R. B. (2015). Does what police do at hot spots matter? The Philadelphia Policing Tactics Experiment. *Criminology, 51*(1), 23-53.

Haberman, C. P., & Ratcliffe, J. H. (2012). The predictive policing challenges of near repeat armed street robberies. *Policing: A Journal of Policy and Practice, 6*(2), 151-166.

Hamilton, L. C. (1992). *Regression with graphics*. Pacific Grove, Calif.: Brooks/Cole.

Herbert, S. (1997). *Policing space: Territoriality and the Los Angeles Police Department*. Minneapolis: University of Minnesota Press.

Herbert, S. (1998). Police subculture revisited. *Criminology, 36*(2), 343-369.

Homans, G. (1950). *The Human group*. New York: Harcourt, Brace.

Ignatans, D., & Pease, K. (in press). Taking crime seriously: Playing the weighting game. *Policing: A Journal of Policy and Practice*.

Johnson, S. D., Bowers, K. J., Birks, D., & Pease, K. (2009). Predictive mapping of crime by ProMap: Accuracy, units of analysis and the environmental backcloth. In D. Weisburd, W. Bernasco, & G. J. N. Bruinsma (Eds.), *Putting Crime in its Place: Units of Analysis in Geographic Criminology* (pp. 171-198). New York: Springer-Verlag.

Jones, P. R. (1990). Community corrections in Kansas: Extending community-based corrections or widening the net? . *Journal of Reseach in Crime and Delinquency, 27*(1), 79-101.

Kennedy, L. W., Caplan, J. M., & Piza, E. (2011). Risk clusters, hotspots, and spatial intelligence: Risk terrain modeling as an algorithm for police resource allocation strategies. *Journal of Quantitative Criminology, 27*(3), 362.

Kent, D. R., Donaldson, S. I., Wyrick, P. A., & Smith, P. J. (2000). Evaluating criminal justice programs designed to reduce crime by targeting repeat gang offenders. *Evaluation and Program Planning, 23*(1), 115-124.

Klein, K. J., Conn, A. B., & Sorra, J. S. (2001). Implementing computerized technology: an organizational analysis. *Journal of Applied Psychology*.

Klinger, D. A. (1997). Negotiating order in patrol work: An Ecological theory of police response to deviance. *Criminology, 35*(2), 277-306.

Koper, C. S. (1995). Just enough police presence: Reducing crime and disorderly behavior by optimizing patrol time in crime hot spots. *Justice Quarterly, 12*(4), 649-672.

Koper, C. S., Lum, C., Willis, J. J., Woods, D. J., & Hibdon, J. (2015). *Realizing the potential of technology in policing*. Retrieved from Washington DC:

Lachin, J. S. (1988). Statistical properties of randomization in clinical trials. *Controlled Clinical Trials, 9*, 289-311.

Lachin, J. S., Matts, J. P., & Wei, L. J. (1988). Randomization in clinical trials: Conclusions and recommendations. *Controlled Clinical Trials, 9*, 365-374.

Manning, P. K. (2010). *Policing Contingencies*: University of Chicago Press.

Marshall, B., & Townsley, M. (2006). *Needles or needless: The Applicability of data mining to the academic criminological community*. Retrieved from London:

Martinez, N. N., Lee, Y., Eck, J. E., & O, S. (2017). Ravenous wolves revisited: A systematic review of offending concentration. *Crime Science, 6*(10), 1-16.

Mastrofski, S. D., & Willis, J. J. (2010). Police Organization Continuity and Change: Into the Twenty-first Century. *Crime and Justice, 39*(1), 55-144.

Matts, J. P., & Lachin, J. S. (1988). Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials, 9*, 327-344.

McCleary, R., McDowall, D., & Bartos, B. J. (2017). *Design and Analysis of Time Series Experiments*. Oxford: Oxford University Press.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association, 106*(493), 108.

Moreto, W. D., Piza, E. L., & Caplan, J. M. (2014). “A plague on both your houses?”: Risks, repeats and reconsiderations of urban residential burglary. *Justice Quarterly, 31*(6), 1102-1126.

National Research Council. (2004). *Fairness and Effectiveness in Policing: The Evidence* (Editors: Wesley Skogan and Kathleen Frydl). Retrieved from Washington, DC:

Nunn, S., Quinet, K., Rowe, K., & Christ, D. (2006). Interdiction Day: Covert surveillance operations, drugs, and serious crime in an inner-city neighborhood. *Police Quarterly, 9*(1), 73-99.

Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Retrieved from Washington DC:

Piquero, A. R., & Blumstein, A. (2007). Does incapacitation reduce crime? *Journal of Quantitative Criminology, 23*(4), 267-285.

Pratt, T. C., & Cullen, F. T. (2005). Assessing macro-level predictors and theories of crime: A meta-analysis. *Crime and Justice, 32*, 373-450.

Ratcliffe, J. H. (2000). Aoristic analysis: The spatial interpretation of unspecific temporal events. *International Journal of Geographical Information Science, 14*(7), 669-679.

Ratcliffe, J. H. (2008). Intelligence-led policing. In R. Wortley & L. Mazerolle (Eds.), *Environmental Criminology and Crime Analysis* (pp. 263-282). Cullompton, Devon: Willan Publishing.

Ratcliffe, J. H. (2014a). Toward an index for harm-focused policing. *Policing, doi:10.1093/police/pau032*, 1-23.

Ratcliffe, J. H. (2014b). What is the future... of predictive policing? *Translational Criminology, 2014* (Spring), 4-5.

Ratcliffe, J. H. (2015a). *Harm-focused policing*. Retrieved from Washington DC:

Ratcliffe, J. H. (2015b). Towards an index for harm-focused policing. *Policing: A Journal of Policy and Practice, 9*(2), 164-182.

Ratcliffe, J. H. (2016). *Intelligence-Led Policing (2nd edn)* (Second ed.). Abingdon, Oxon.: Routledge.

Ratcliffe, J. H., & McCullagh, M. J. (1998). Aoristic crime analysis. *International Journal of Geographical Information Science, 12*(7), 751-764.

Ratcliffe, J. H., & McCullagh, M. J. (2001). Chasing ghosts? Police perception of high crime areas. *British Journal of Criminology, 41*(2), 330-341.

Ratcliffe, J. H., & Rengert, G. F. (2008). Near repeat patterns in Philadelphia shootings. *Security Journal, 21*(1-2), 58-76.

Ratcliffe, J. H., Taniguchi, T., Groff, E. R., & Wood, J. D. (2011). The Philadelphia Foot Patrol Experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology, 49*(3), 795-831.

Rummens, A., Hardyns, W., & Pauwels, L. (in press). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Applied Geography*.

Serlin, R. C., & Levin, J. R. (1985). Teaching How to Derive Directly Interpretable Coding Schemes for Multiple Regression Analysis. *Journal of Educational Statistics, 10*(3), 223-238. doi:10.2307/1164794

Sherman, L., Neyroud, P. W., & Neyroud, E. (2016). The Cambridge crime harm index: Measuring total harm from crime based on sentencing guidelines. *Policing: A Journal of Policy and Practice, 10*(3), 171-183.

Sherman, L. W. (2007). The power few: Experimental criminology and the reduction of harm. *Journal of Experimental Criminology, 3*(4), 299-321.

Sherman, L. W., Gartin, P., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology, 27*(1), 27-55.

Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice Quarterly, 12*(4), 625-648.

Sherman, L. W., Williams, S., Ariel, B., Strang, L. R., Wain, N., Slothower, M., & Norton, A. (2014). An integrated theory of hot spots patrol strategy: Implementing prevention by scaling up and feeding back. *Journal of Contemporary Criminal Justice, 30*(2), 95-112.

Short, M. B., D'Orsogna, M. R., Brantingham, P. J., & Tita, G. E. (2009). Measuring and modeling repeat and near-repeat burglary effects. *Journal of Quantitative Criminology, 25*, 325–339.

Spelman, W. (2000). What recent studies do (and don't) tell us about imprisonment and crime. In M. Tonry (Ed.), *From Crime and Justice: A Review of Research* (Vol. 27, pp. 419-494). Chicago: University of Chicago Press.

Stata Corporation. (2015). lowess (Stata Help file, v. 14).

Strauss, A. L., & Corbin, J. M. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks: Sage.

Taniguchi, T. (2010). *Policing a negotiated world: An Empirical assessment of the ecological theory of policing. Unpublished doctoral dissertation.* (Ph.D.), Temple University, Philadelphia, PA.

Taylor, B., Koper, C. S., & Woods, D. J. (2011). A randomized controlled trial of different policing strategies at hot spots of violent crime. *Journal of Experimental Criminology, 7*(2), 149-181.

Taylor, R. B. (1994). *Research Methods in Criminal Justice*. New York: McGraw-Hill.

Taylor, R. B. (2015). *Community Criminology: Fundamentals of Spatial and Temporal Scaling, Ecological Indicators, and Selectivity Bias*. New York: New York University Press.

Taylor, R. B., Ratcliffe, J. H., & Perenzin, A. (2015). Can we predict long-term community crime problems? The estimation of ecological continuity to model risk heterogeneity. *Journal of Research in Crime and Delinquency, 52*(5), 635-657. doi:10.1177/0022427815586062

Thacher, D. (2008). Research for the front lines. *Policing & Society, 18*(1), 46-59.

Townsley, M., Homel, R., & Chaseling, J. (2003). Infectious Burglaries. A Test of the Near Repeat Hypothesis. *Br J Criminol, 43*(3), 615-633. doi:10.1093/bjc/43.3.615

Townsley, M., & Pease, K. (2002). How efficiently can we target prolific offenders? *International Journal of Police Science and Management, 4*(4), 323-331.

Twisk, J. W. R. (2013). *Applied Longitudinal Data Analysis for Epidemiology* (Second ed.). Cambridge (UK): Cambridge University Press.

Weisburd, D., & Braga, A. A. (2006). Hot spots policing as a model for police innovation. In D. Weisburd & A. A. Braga (Eds.), *Police Innovation: Contrasting Perspectives* (pp. 225-244). New York: Cambridge University Press.

Weisburd, D., & Eck, J. (2004). What can police do to reduce crime, disorder, and fear? *The Annals of the American Academy of Political and Social Science, 593*(1), 43-65.

Weisburd, D., & Green, L. (1995). Measuring immediate spatial displacement: methodological issues and problems. In J. E. Eck & D. Weisburd (Eds.), *Crime and Place* (Vol. Volume 4, pp. 349-361). Monsey, NY: Criminal Justice Press.

Weisburd, D., Groff, E. R., Jones, G., Cave, B., Amendola, K. L., Yang, S.-M., & Emison, R. F. (2015). The Dallas patrol management experiment: can AVL technologies be used to harness unallocated patrol time for crime prevention? *Journal of Experimental Criminology, 11*(3), 367-391.

Weisburd, D., & Telep, C. (in press). Hot Spots Policing: What we know and what we need to know. *Journal of Contemporary Criminal Justice*.

Willis, J. J., & Mastrofski, S. D. (2014). Pulling together: Integrating craft and science. *Policing, 8*(4), 321-329.

Willis, J. J., & Mastrofski, S. D. (2016). Improving policing by integrating craft and science: what can patrol officers teach us about good police work? *Policing and Society*, 1-18.

Wolfgang, M. E., Figlio, R. M., & Sellin, T. (1972). *Delinquency in a Birth Cohort*. Chicago: University of Chicago Press.

Wyant, B. R., Taylor, R. B., Ratcliffe, J. H., & Wood, J. (2012). Deterrence, firearm arrests, and subsequent shootings: A micro-level spatio-temporal analysis. *Justice Quarterly, 29*(4), 524-545.

APPENDICES

Public reports currently available from http://bit.ly/CSCS_3PE

Center for Security and Crime Science

The Philadelphia Predictive Policing Experiment

SUMMARY OF THE EXPERIMENTAL DESIGN

Jerry H. Ratcliffe and Ralph B. Taylor

Overview

The Philadelphia Predictive Policing Experiment represents a two-year collaboration between Temple University's Center for Security and Crime Science, housed in the Department of Criminal Justice at Temple, and the Philadelphia Police Department. This National Institute of Justice funded research project was the first place-based, randomized experiment to study the impact of *different* patrol strategies on violent and property crime in predicted criminal activity areas. The experiment hopes to learn whether different but operationally-realistic police responses to crime forecasts estimated by a predictive policing software program can reduce crime.

What is predictive policing?

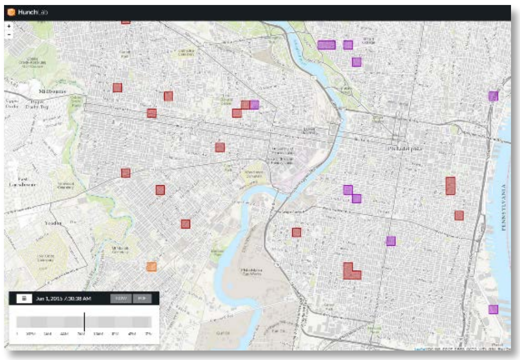
Predictive policing is an emerging tactic relying in part on software forecasting the likely locations of criminal events. Predictive policing, while sometimes applied to offenders, is also frequently applied to high crime places. In this context, it involves "the use of historical data to create a spatiotemporal forecast of areas of criminality or crime hot spots that will be the basis for police resource allocation decisions with the expectation that having officers at the proposed place and time will deter or detect criminal activity"[1].

Experimental design

The research team from Temple University and the Research and Analysis Unit of the Philadelphia Police Department randomly assigned 20 Philadelphia Police Department (PPD) districts into one of four experimental conditions. Block randomization, for both the property and violent random assignments, assured that each of the four groups were comparable on the relevant crime, and on key demographic features. Five districts acted as controls, with a business-as-usual patrol strategy ('control' districts). In five districts, officers were made aware of the predicted high crime activity area at roll call and asked to concentrate there when able ('awareness' districts). Five districts received the awareness model treatment and also dedicated a patrol car to the predicted crime areas ('marked car' districts). Finally, five districts received the awareness model treatment as well as dedicating an unmarked unit to the predicted crime areas ('unmarked car' districts). Officers in marked and unmarked cars were from the local district

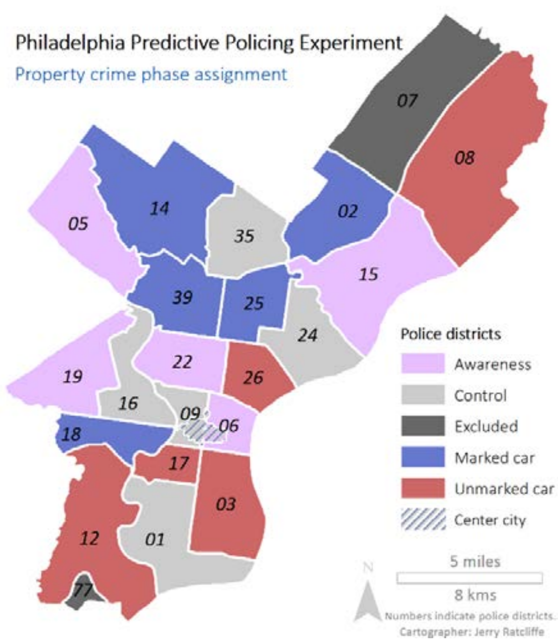
[1] Ratcliffe, JH (2014) "What is the future… of predictive policing?" *Translational Criminology*, 2014 (Spring) pg. 4

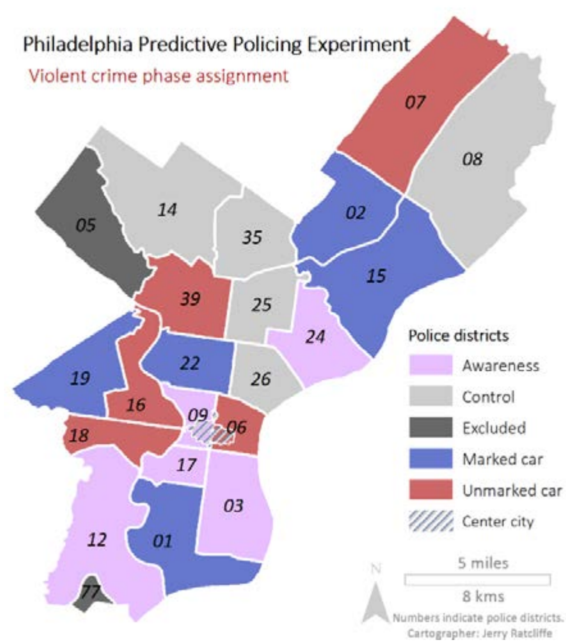
station, but were exempt from answering radio calls outside of their areas. They were encouraged to respond to related radio calls (property or violent crime) *inside* their predicted grid areas.



The predictive policing software used was the HunchLab program designed by Azavea. Hunchlab is a web-based predictive policing system that accesses real-time Philadelphia Police data to produce crime forecasts for the city. It incorporates statistical modeling that considers aoristic temporal analysis, seasonality, risk terrain modeling, near repeats, and collective efficacy. Officers at police district buildings in the marked and unmarked conditions could log in and print out maps for forthcoming 8-hour shifts. Azavea adapted the software at our request to generate three predicted 500 feet square grids per district per shift. They also included a slight randomization component to reduce the possibility that the same grid cells were predicted every day.

The software forecasted property crime areas from 8am to 4pm every day across Philadelphia from June 1st to August 29th, 2015. Property crime comprised residential and commercial burglary, motor vehicle theft, and theft from vehicles. After the three month property crime phase, the experiment paused to in recognition of Philadelphia Police Department preparations for the visit of Pope Francis and the World Meeting of Families in Philadelphia from 22-27 September. Subsequently, the violent crime phase ran from November 1st 2015 to January 31st, 2016. Predicted violent crime areas were projected every day from 6pm until 2am of the next day. Violent crime comprised shootings, robberies, aggravated assaults, and homicides.



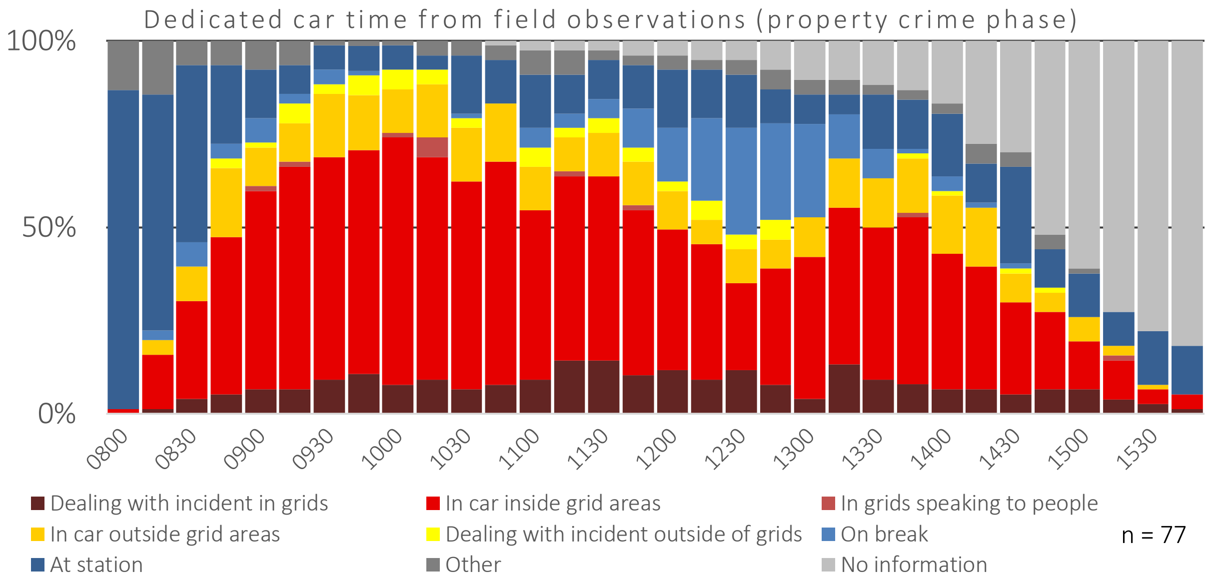


Implementation and fidelity

Prior to the experiment, researchers and senior officers involved in the experimental design provided briefings to command staff, researchers and staff from the PPD research and analysis section created short videos explaining the software and to explain what was expected of field personnel, and in the early stages of the experimental work they visited districts to help with software and implementation issues.

Every day, supervisors were asked to complete a one-page survey sheet asking them about their view regarding the accuracy of the predicted areas and what level of policing occurred in the predicted areas. Officers in marked and unmarked cars also completed a one-page summary form after every shift that asked how busy the grid areas were, and how often they were in the grid areas. These forms were gathered by the crime analysts in each district, and forwarded to police headquarters and the researchers. A senior analyst at police headquarters monitored compliance with the forms. The senior analyst also provided a centralized source to field personnel for questions related to the experiment.

Trained researchers accompanied marked and unmarked cars on over 100 ride-alongs during the experiment. These researchers noted police activities, amount of time spent in grid areas, and gathered information from the patrol officers regarding their views about predictive policing. Observers recorded field notes as soon as practical after each ride-along, and completed a time sheet detailing the main police activity and car location relative to the grids assigned for 15 minute blocks throughout the shift (see example below).



Further information

For additional and current information, please visit the project website at **bit.ly/CSCS_3PE**

Suggested citation: Ratcliffe, JH & Taylor, RB (2017) *The Philadelphia Predictive Policing Experiment: Summary of the Experimental Design*. December 2017. Online at bit.ly/CSCS_3PE. 3 pages.

TEMPLE UNIVERSITY

Center for Security and Crime Science

THE PHILADELPHIA PREDICTIVE POLICING EXPERIMENT

EFFECTIVENESS OF THE PREDICTION MODELS

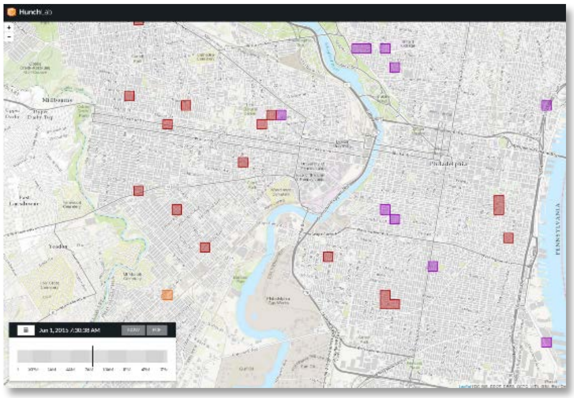
Jerry H. Ratcliffe, Ralph B. Taylor and Amber P. Askey

Overview

The Philadelphia Predictive Policing Experiment represents a two-year collaboration between Temple University's Center for Security and Crime Science, housed in the Department of Criminal Justice at Temple, and the Philadelphia Police Department. This National Institute of Justice-funded research project was the first place-based, randomized experiment to study the impact of *different* patrol strategies on violent and property crime in predicted criminal activity areas. The experiment aimed to learn whether different but operationally-realistic police responses to crime forecasts estimated by a predictive policing software program can reduce crime.

Software performance

The predictive policing software used was the HunchLab program designed by Azavea. HunchLab is a web-based predictive policing system that accesses real-time Philadelphia Police data to produce crime forecasts for the city. It incorporates statistical modeling that considers seasonality, risk terrain modeling, near repeats, and collective efficacy. Officers at police district buildings in the marked and unmarked conditions could log in and print out maps for forthcoming 8-hour shifts. Azavea adapted the software at the request of the Philadelphia Police Department and researchers from Temple University to generate three predicted 500 feet square grids per district per shift. They also included a slight randomization component to reduce the possibility that the same grid cells were predicted every day. This allowed the project to create the necessary experimental conditions by limiting the output of the software to generate three predicted grids (500 feet by 500 feet) in every district of the city each day. **It is important to note therefore that the experiment artificially reduced the efficiency of the software**, because it forced the software to choose grids in low crime districts, and limited the number of grids it could assign in high crime districts.



During the experiment, researchers accompanied officers on over 100 dedicated car assignment patrols (either marked or unmarked cars). It was clear from these observations that it was impossible and unrealistic for assigned officers to exclusively patrol the 500 feet square predicted grid areas. In many cases the grids identified one-way streets, necessitating driving through the surrounding block or two to return to the grid assignment. We therefore examine the efficacy of the software considering not just the predicted grids alone, but also and more realistically the surrounding grids as well. We do this for the grids nearest to each predicted grid (in technical terms, *first order queen contiguity*), and for grids within two cells of each predicted grid (*second order queen contiguity*). The table below shows the results of the software prediction for control districts, as the control districts were not affected by focused police activity resulting from the experiment. The control districts are a good indication of the efficacy of the software program.

The unit of analysis is the district day – that is, an eight-hour shift in one district in one day. For each experimental condition (such as the control sites shown here) there were five districts across a 90 day experimental phase. This equals 450 district days (5 districts x 90 days = 450) and 460 days for the violent crime phase (the violent crime phase ran for two additional days, so 5 districts x 92 days = 460).

In the table you can see that in the control sites for property crime, 274 district days (60.9%) had at least one crime in the entire district during the shift. For property crime in these 274 district days, the software predicted nearly 14 percent of the crime available to predict in these districts, and did so by identifying just 6.8 percent of the district area. When extended up to two grid cells away from the predicted grids, it predicted 30.3 percent of the property crime while highlighting only 15.3 percent of the district. Please remember that these predictions were not optimized – the software was deliberately hobbled to enable the experimental design. The results are therefore conservative. Also note that the results differed by district and crime type, as shown below. **In general, however, it appears that the software was able to predict twice as much crime as we would expect if crime were spread uniformly across the districts, even when artificially constrained by our experiment to be less effective than designed.**

Phase / district	**District days**	**Crime days in district**	**Total crimes**	**Within one grid (% total)**	**% area of district**	**Within two grids (% total)**	**% area of district**
Property	**450**	**274 (60.1%)**	**489**	**68 (13.9%)**	**6.8%**	**148 (30.3%)**	**15.3%**
24	90	61 (67.7%)	127	11 (8.6%)	5.1%	31 (24.4%)	11.4%
16	90	46 (51.1%)	64	12 (18.7%)	6.1%	22 (34.4%)	12.9%
35	90	74 (82.2%)	157	22 (14%)	5.7%	34 (21.6%)	14.2%
9	90	59 (65.5%)	92	15 (16.3%)	12.9%	38 (41.3%)	28.6%
1	90	34 (37.7%)	49	8 (16.3%)	2.6%	23 (46.9%)	4.9%
Violent	**460**	**137 (29.8%)**	**176**	**24 (13.6%)**	**5.1%**	**43 (24.4%)**	**12.1%**
25	92	49 (53.2%)	68	7 (10.3%)	5.8%	18 (26.5%)	13.9%
35	92	32 (24.4%)	40	4 (10%)	5.9%	6 (15%)	14.4%
26	92	19 (20.6%)	20	4 (20%)	7.4%	10 (50%)	16.2%
14	92	27 (29.3%)	37	8 (21.6%)	2.6%	8 (21.6%)	6.5%
8	92	10 (10.8%)	11	1 (9.1%)	1.4%	1 (9.1%)	3.6%

Further information

For additional and current information, please visit the project website at **bit.ly/CSCS_3PE**

Suggested citation: Ratcliffe, JH, Taylor, RB and Askey, AP (2017) *The Philadelphia Predictive Policing Experiment: Effectiveness of the Prediction Models*. December 2017. Online at bit.ly/CSCS_3PE. 2 pages.

Center for Security and Crime Science

The Philadelphia Predictive Policing Experiment

IMPACTS OF POLICE CARS ASSIGNED TO HIGH CRIME GRIDS

Jerry H. Ratcliffe, Ralph B. Taylor, Amber P. Askey, John Grasso and Ryan Fisher

Project background

The Philadelphia Predictive Policing Experiment was a two-year collaboration between Temple University's Center for Security and Crime Science, housed in the Department of Criminal Justice, and the Philadelphia Police Department. This NIJ-funded research project was the first place-based, randomized experiment to study the impact of different patrol strategies on violent and property crime in predicted criminal activity areas. The experiment aimed to learn whether different but operationally-realistic police responses to crime forecasts, estimated by a predictive policing software program, can reduce crime.

The experimental design

The experiment was designed to test two theoretically-relevant operational questions about police patrol. If police are able to dedicate a car to predicted crime areas, would it be better to use a marked car or an unmarked car? The visible police car would emphasize deterrence and prevention. The plain-clothes car would allow officers to conduct surveillance and approach crime undetected in more of an intelligence-led or apprehension mode. The experiment also examined if it was sufficient to just tell officers on roll call where the predicted grids were for the day without having a car dedicated to the task. These three interventions were compared to control areas which did policing as usual.

A property crime phase ran for 90 days from June 1, 2015 through August 25, 2015. A break was scheduled and then the violent crime phase ran for 92 days from November, 1, 2015 through January 31, 2016. Designed to target the time of the day with most crime problems, in each district three property crime grid predictions (500 feet by 500 feet) were active from 8am to 4pm, and three violent crime predictions were active in each district for an eight hour period from 6pm to 2am. After removing the airport and the lowest crime district, the remaining 20 districts were randomly assigned to one of four experimental conditions:

Awareness districts informed officers of the predicted target areas for that shift. Officers were asked to focus on those areas when they were able, but no cars were dedicated to the grids.

Marked car districts built upon the awareness model by dedicating a single marked vehicle to patrol of the predicted crime areas for the entirety of the shift.

Unmarked car districts were similar to the implementation for the marked districts except for the use of plain-clothes officers and unmarked police vehicles instead of uniformed resources.

Control districts were districts where police personnel did not have access to the crime prediction software, so they maintained a standard patrol strategy.

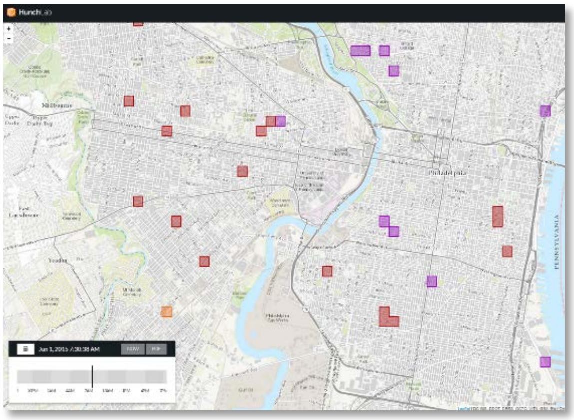
In the car districts, officers were instructed to remain in the grids as much as possible, but were not instructed how long, at what frequency, or in what order they should be patrolled.

Results

When examining both predicted high-crime grid cells and the grids cells immediately surrounding them, the marked car patrols resulted in a 31% reduction in property crime counts, or a 36% reduction in the number of cells experiencing at least one crime. While this sounds substantial, the specific numbers are small. This translates to a reduction in three crimes over 3 months for an average city district patrolling around three grids. To extrapolate, if each of the 21 geographic districts dedicated a marked car to three grids for an 8 hour shift each day, we estimate a reduction in 256 Part I property crimes per year.

There were signs of a temporal diffusion of benefits. In the eight hours following the property crime patrols, the marked car districts were associated with a reduction in crime compared to the control areas, with property crime counts that were 41% lower and expected crime occurrences that were 48% lower (but again, actual numbers were low). Unfortunately, the relative rarity of property crime on an hour-by-hour basis in such limited geographic areas hindered the ability to make confident inferences about any crime reductions regarding the experimental conditions. In other words, while the percentages were substantial, the results were not statistically significant due to floor effects.

There were no crime reduction benefits associated with the violent phase of the experiment, nor were there any benefits with the property crime awareness or unmarked car interventions.



The experiment used Azavea's HunchLab predictive policing software. The software predicted twice as much crime as would be expected if crime were uniformly distributed, even though the software was constrained from making its best predictions by the experimental design. The experiment did reveal the technical challenges in predicting crime in a limited number of small 500' x 500' grids. See additional report.

Further information

For additional and current information, visit the project website at **bit.ly/CSCS_3PE**

Suggested citation: Ratcliffe, JH, Taylor, RB, Askey, AP, Grasso, J and Fisher, R (2017) *The Philadelphia Predictive Policing Experiment: Impacts of police cars assigned to high crime grids.* December 2017. Online at bit.ly/CSCS_3PE.